Characterisation of Genes Involved in Oogenesis in *Drosophila melanogaster*

Ian Bryce MacIver

Thesis Presented for the Degree of Doctor of Philosophy
The University of Edinburgh
February 1997

Declaration

I declare that this thesis was composed by myself and that the work described is my own, unless otherwise stated.

Bryce MacIver
February 1997

Abstract

Oogenesis is the process by which an oocyte (egg) develops to a point where it is capable of fertilisation. This process is complex, requiring many fundamental cellular and molecular interactions. The study of oogenesis is greatly facilitated in the fruitfly *Drosophila melanogaster* due to this organism's accessible morphology and the advanced understanding of its genetics and molecular biology.

A reverse genetics approach has been adopted in this laboratory to investigate genes involved in oogenesis. This approach uses a P-element enhancer trap reporter system to provide clues to the temporal and spatial expression of genes. Molecular techniques are then used to clone genes situated near the enhancer trap in the genome with *in situ* hybridisation used to match cloned genes to the original reporter expression pattern. Occasionally an expression is observed from the *in situ* hybridisation that bears no relation to the reporter gene but is very interesting nonetheless. The work in this thesis stems from such an observation.

We isolated a cDNA with an *in situ* hybridisation expression pattern that showed gene expression very early in oogenesis with rapid localisation of the transcript to the developing oocyte. Egg chambers at mid-stages of oogenesis exhibited localisation of the transcript to the anterior region of the developing oocyte while late stages of oogenesis showed strong expression in the nurse cells and loss of localisation in the oocyte. A number of localised transcripts are known to have crucial roles in axis determination in the developing oocyte and embryo.

Work presented in this thesis shows that this gene encodes a class V unconventional myosin. Examples of this class have been described in mouse and yeast and there is evidence that they are involved in vesicle trafficking. Additional cDNAs were isolated and sequenced although no full length cDNAs were found. A PCR and genomic DNA cloning strategy was used to obtain most of the remaining coding sequence. Hybridisation to polytene chromosomes from the salivary glands of third instar larvae located the gene to the 43BC region on the right arm of the second chromosome. Results mapping the class V myosin to one of three genes in this region will be presented.

A region from the carboxyl terminus of the gene was expressed in *Escherichia coli* with the recombinant protein used to raise antibodies in rabbits. Preliminary results show that these antibodies detect protein from ovary tissue.

A cDNA encoding a putative ion-dependent inorganic phosphate co-transporter was isolated and sequenced in full. A PCR approach designed to amplify from the class V myosin ATP binding site identified sequence from a previously uncharacterised gene showing similarity to an ATPase required for meiotic spindle formation in *Caenorhabditis elegans*.

The implications of these findings and the future directions this work can take is discussed.

Acknowledgements

I would like to offer my sincere thanks to my supervisor, Professor Mary Bownes, for guidance throughout this project. Many thanks also to past and present members of the group; Debbie, Diane, Simone, Debiao, Wu min, Dot, Stephen, Lucy, Roger, Janis, Craig, Neil and Colin. Special thanks to Kathleen for showing me how to manage and dissect *Drosophila*. Thanks to Alison and P'ng for their work on the mouse myosin. Thanks to Claudia for her kind friendship and support. Very special thanks and best wishes to Angela, my ever helpful colleague on this project.

I would like to thank Professor Adrian Bird and Vicki Clark for use of the automated sequencing facility and also to thank Dr Andrew Coulson for assistance with running the coiled-coil programs.

I would especially like to thank Professor Richard Ambler and Professor Sir Ken Murray of the Darwin Trust of Edinburgh for allowing me this opportunity to come to Edinburgh.

I dedicate this thesis to my parents for their never-ending love and support throughout my long career in tertiary education.

Abbreviations

ATP	Adenosine-5'-triphosphate
bp	base pair
℃	degrees Centigrade
Ci	Curies
cDNA	Complementary deoxyribonucleic acid
cfu	colony forming units
cm	centimetre
DAB	3', 3'-Diaminobenzidine tetrahydrochloride
dATP	Deoxyadnenosine-5'-triphosphate
dCTP	Deoxycytosine-5'-triphosphate
dGTP	Deoxyguanosine-5'-triphosphate
dTTP	Deoxythymidine-5'-triphosphate
ddATP	2' (3'-di) Deoxyadnenosine-5'-triphosphate
ddCTP	2' (3'-di) Deoxycytosine-5'-triphosphate
ddGTP	2' (3'-di) Deoxyguanosine-5'-triphosphate
ddTTP	2' (3'-di) Deoxythymidine-5'-triphosphate
dNTP	deoxynucleotide-5'-triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Ethylenediamine-tetra-acetic acid
g	gram
g	gravitational force
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid
HRP	Horse radish peroxidase
IPTG	Isopropyl-β-D-thio-galactopyranoside
kDa	kiloDaltons
kb	kilobase pairs
Klenow	Large fragment of DNA Polymerase I
l	litre
M	moles per litre
MOPS	Morpholinopropanesulphonic acid
mRNA	messenger RNA
nt	nucleotide

OD	Optical density
OLB	Oligo labelling buffer
pfu	plaque forming units
PCR	Polymerase chain reaction
pH	-Log10 [hydrogen ion concentration]
RT-PCR	Reverse transcription polymerase chain reaction
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
rpm	revolutions per minute
SDS	Sodium dodecyl sulphate
Tris	Tris(hydroxymethyl)-amino-methane
Tris-HCl	Tris solution, pH adjusted with HCl acid
X-Gal	5-bromo-4-chloro-3-indolyl-β-galactopyranoside
TCA	Trichloroacetic acid
uv	ultra-violet
V	Volts
v/v	volume per volume
w/v	weight per volume
W	Watts
pmol	picomoles
ng	nanogram
nmol	nanomoles
µCi	microCuries
µg	microgram
µl	microlitres
µM	microMolar
µmol	micromole
mA	milliAmpere
mg	milligram
ml	millilitre
mm	millimetre
mM	milliMolar
msec	milliseconds
~	approximately

IUB symbols for amino acids

A	Alanine
R	Arginine
N	Asparagine
D	Aspartate
C	Cysteine
Q	Glutamine
E	Glutamate
G	Glycine
H	Histidine
I	Isoleucine
L	Leucine
K	Lysine
M	Methionine
F	Phenylalanine
P	Proline
S	Serine
T	Threonine
W	Tryptophan
Y	Tyrosine
V	Valine

IUB symbols for nucleic acid sequences

A	Adenosine
C	Cytosine
G	Guanosine
T	Thymidine
M	A or C
R	A of G
W	A or T
S	C or G
Y	C or T
K	G or T
V	A or C or G
H	A or C or T
D	A or G or T
B	C or G or T
N	A or C or G or T

Contents

Chapter 1

Introduction

1.1 Introduction

Intensive studies over many years have produced a detailed explanation of how an embryo develops from a single egg, one of the most fascinating occurrences in all of biology. The egg itself is a very complex cell containing a haploid genome, position information essential for the establishment of body pattern and, in non-mammals, nutrients which must support growth until the egg hatches. Oogenesis is the process by which an egg or oocyte is formed and thus the study of oogenesis is an integral part of understanding how life develops from a single cell.

The application of modern molecular techniques has revolutionised the study of many biological systems. These techniques are readily applicable to the field of developmental biology and the study of oogenesis. The fruit fly, *Drosophila melanogaster*, provides an ideal organism for the study of the developmental pathways in oogenesis because of the wealth of genetic information and the technical facilities which exist for this organism.

1.2 The Morphology of Oogenesis

The morphology of oogenesis is described in detail in Bownes and Dale (1982). An adult female *Drosophila* contains a pair of ovaries (figure 1.1A), each attached to a lateral oviduct which fuse into the common oviduct, which in turn leads to the uterus. Sperm is stored in the seminal receptacle and spermathecae, and is used to fertilise mature eggs in the uterus. A network of muscles fibres called the peritoneal sheath cover each ovary and assist with the progression of developing eggs down the ovarioles. Each ovary consists of 15-20 ovarioles which are joined at their ends by terminal filaments.

At the tip of each ovariole are 1-4 germ-line derived stem cells (figure 1.1B). A stem cell divides to give rise to a daughter stem cell and cystoblast. The cystoblast undergoes four further mitotic divisions, each with incomplete cytokinesis to produce a cluster of 16 interconnected cells or a cystocytes. One cystocyte, usually a posterior-most cell with four intercellular connections (called ring canals) will become the oocyte. The remaining 15 cells become nurse cells.

The cystocyte complex becomes surrounded by somatically derived prefollicular cells which later give rise to various follicle cell types. This collective group of cystocytes and follicle cells is termed an egg chamber. Egg chambers mature, enlarge and progress down the ovariole towards the oviduct. Thus, each ovariole contains a series of egg chambers at different stages of development.

There are 14 morphologically distinct stages of oogenesis as originally described by King (1970), see also figure 1.1. Stage 1 consists of the 16 cell cyst surrounded by a monolayer of approximately 80 follicle cells, and is located at the anterior most region of the ovariole known as the germarium. During stages 2 to 5 the follicle cells proliferate to approximately 1200 cells. One of the cystocytes becomes distinguishable as the oocyte and undergoes meiosis. The nurse cells become polyploid with up to 2096 genomic copies and they begin to synthesise the requirements of the oocyte. The nurse cells exhibit a graded ploidy with those closest to the oocyte having a greater number of genomic copies. The follicle cells also exhibit polyploidism, but to a lesser extent.

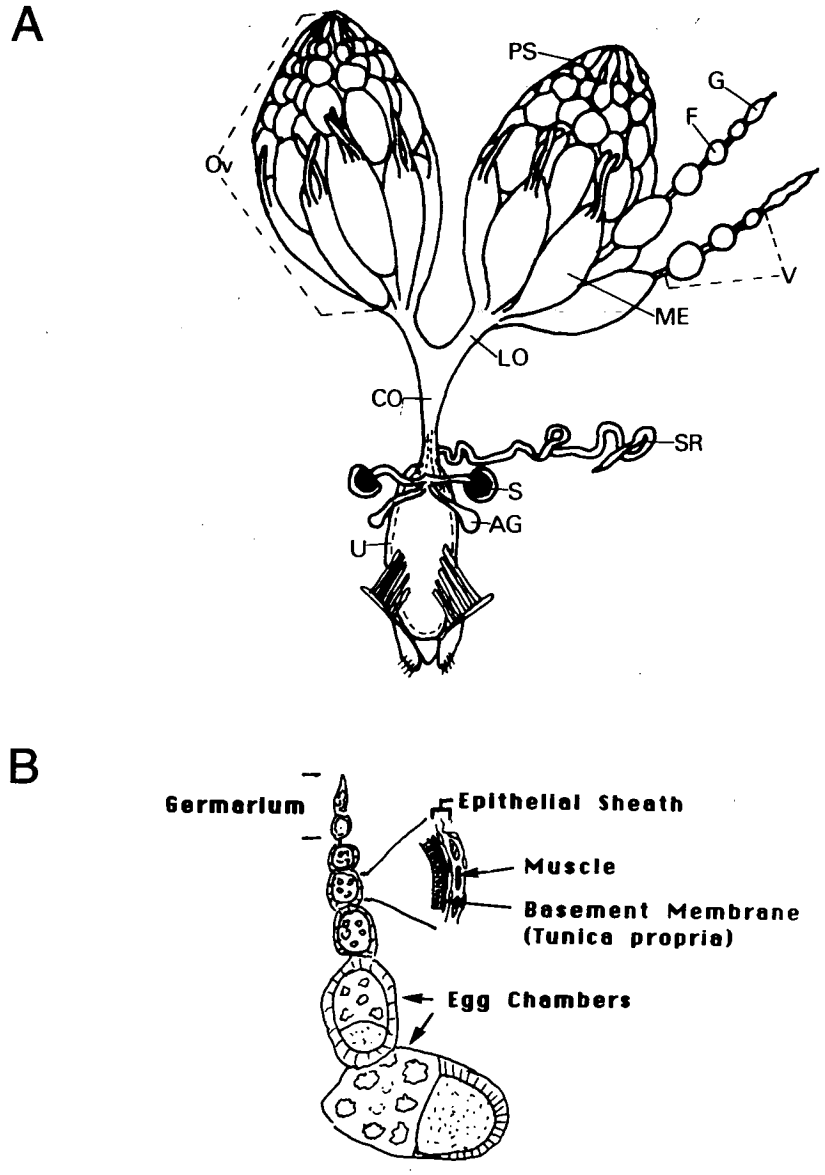


Figure 1.1: The structure of the adult female reproductive system of *Drosophila melanogaster*. A) Ovaries with 2 ovarioles separated from the ovary. Ov: ovary, LO: lateral oviduct, G: germarium, PS: peritoneal sheath, V: vitellarium, F: follicle, ME: mature egg, CO: common oviduct, SR: seminal receptacle, S: spermatheca, AG: accessory gland, U: uterus (from Bownes and Dale, 1982). B) Ovariole (from Spradling 1993a) .

During stages 6 to 8 the follicle cells become evenly distributed around the egg chamber with those around the oocyte becoming columnar in shape. Stages 9 to 11

During stages 6 to 8 the follicle cells become evenly distributed around the egg chamber with those around the oocyte becoming columnar in shape. Stages 9 to 11 see complex follicle cell migration and differentiation into several subtypes. A cluster of 6-10 cells, called border cells, migrate through the nurse cells to the anterior surface of the oocyte and are required for formation of the micropyle, the appendage which allows sperm to enter the oocyte for fertilisation. By stage 11 three distinct subtypes of follicle cell can be seen; anterior pole cells, posterior pole cells and columnar main body cells. About 80 squamous follicle cells remain surrounding the nurse cells.

Deposition of yolk protein begins from stage 8 and continues to stage 10. Yolk proteins are synthesised in the fat body and secreted into the haemolymph for transport to the oocyte. Yolk proteins are also secreted by the follicle cells. The oocyte takes up the yolk proteins by endocytosis with a dramatic increase in the size of the oocyte being observed. The inner protective layer of the egg, called the vitelline membrane is secreted by follicle cells surrounding the oocyte during stages 9 and 10. Deposition of yolk protein ceases once the vitelline membrane is complete. The egg shell, or chorion, is secreted by columnar follicle cells between stages 11 and 14. The follicle cells then begin to degenerate.

The nurse cells synthesise lipid droplets, a variety of organelles and RNA species, both ribosomal and messenger, which are required for early embryonic protein synthesis and pattern formation. Prior to stage 10, RNA synthesis is intense with the nurse cells increasing in volume. At stage 10B there is a rapid and massive movement of nearly all of the nurse cell cytoplasm into the oocyte. The nurse cells begin to degenerate after this stage.

Experiments measuring the incorporation of ^{3}H uridine show that the oocyte is transcriptionally inactive except for a brief period at stages 9 and 10. At stage 13 the nuclear envelope breaks down and the chromosomes move to the metaphase plate of meiosis and arrest.

1.3 The Molecular Basis of Oogenesis

1.3.1 Genetic Screens

Oogenesis is a highly complex process involving cell determination, intercellular and intracellular transport, cell to cell communication, and temporal and spatial genetic regulation (reviewed by Spradling [1993a] and Lasko [1994]).

A number of genetic screens have been carried out primarily in the 1970's and 1980's to identify mutations that affect female fertility (for review, see Lasko, 1994). A screen by Bakken (1973) identified 19 female sterile loci on the second chromosome and 17 loci on the third chromosome. These loci could be distinguished as belonging to one of three groups: maternal-effect lethals (group I), sterile, but producing a normal looking egg (group II), and defective in oogenesis (group III). This study was important in that it clearly demonstrated the requirement for maternal factors in embryonic development.

One question that has puzzled researchers is: how many genes are involved exclusively in oogenesis? Perrimon et al (1986) have investigated this question through the identification of female sterile mutants on the X-chromosome. Extrapolation of the results led them to suggest that only 75 genes in *Drosophila* are required exclusively for oogenesis. Schüpbach and Wieschaus (1989, 1991) have extensively analysed the second chromosome for female sterile and maternal effect mutations. Based on these results, these authors calculate that approximately 300 genes are required specifically for oogenesis. Analysis of zygotic lethals using mitotic recombination techniques have identified nearly 800 loci that are required maternally for oogenesis or early embryogenesis (see Lasko, 1994). In addition, many 'housekeeping' genes required throughout development will also function during oogenesis.

1.3.2 The Enhancer Trap

Drosophila has yielded many of its genes to researchers through traditional genetic approaches. This method involves identifying mutations which disrupt the process being investigated. The gene is mapped using genetic recombination, with greater accuracy obtained through deficiency mapping and by the study of chromosomal rearrangements that affect the gene's function. Cloning is facilitated by

chromosomal walking from a known region until a transcript is obtained. This process can be time and labour intensive and requires that the gene of interest has a selectable phenotype or is easily identifiable. Certain genes are difficult to detect using the classical method; those with a degree of redundancy; those with subtle functions, being expressed in a limited number of cells; those which function several times during development but which only exhibit one mutant phenotype - often prior to oogenesis (Freeman, 1991).

An alternative method was recently introduced which uses P-element mediated enhancer detection (O'Kane and Gehring, 1987). In this system a modified transposable P-element is randomly inserted into the genome. The modified P-element carries the *Escherichia coli lacZ* gene fused in frame to the second exon of the P-element transposase gene. Low levels of *lacZ* expression can occur from the constitutive but weak P-element promoter. However, when the promoter comes under the influence of nearby enhancer elements tissue specific expression of *lacZ* can occur. Thus, *lacZ* acts as a tissue specific reporter gene.

An enhancer trap line is generated by crossing a fly containing the modified P-element with a fly that carries a disabled P-element, one which produces active transposase but which is unable to transpose itself. Stable lines are generated by using a marker incorporated in the modified P-element, usually an eye colour gene. A number of assumptions were made about this system when it was introduced; that the promoter would function in various tissues and from a variety of enhancer elements and that there would no spurious enhancer elements giving "false" patterns. Also unknown was what number of specific expression patterns would be seen (Freeman, 1991).

Several studies have validated the enhancer trap approach. Fasano and Kerridge (1988) in a study on oogenesis obtained at least 184 lines with single insertions. Of these, 32% showed minor background β-galactosidase staining indicating no enhancer activity while 68% showed novel patterns. 13% were germ-line specific, 33% follicle cell specific while 20% were expressed in both the germ-line and the follicle cells and 2% showed staining specific to the germarium. Bier et al (1989) examined over 3700 lines by staining the embryos and obtained 64% with tissue specific β-galactosidase staining of which 49% were specific to either the nervous system or the gut. These results showed that the P-element mediated enhancer trap system was capable of producing specific staining patterns in the majority of cases.

Wilson et al (1989) used a modified P-element that also contained part of a bacterial plasmid cloning vector which facilitated the cloning of genomic DNA 3' to the P-element. In their study, 26 transposants (lines carrying a P-element) were examined with 4 showing an *in situ* hybridisation pattern from DNA cloned using the P-element-bacterial plasmid system. The *in situ* hybridisation patterns matched the β-galactosidase staining patterns. An extensive study of oogenesis by Grossniklaus et al (1989) involved the examination of over 600 lines. Approximately 47% of these lines showed β-galactosidase staining in one or more ovarian cell with most being temporally and spatially regulated. Germ-line expression showed staining mostly in the nurse cells but one line showed specific staining in the oocyte nucleus. The oocyte seemed more transcriptionally active than had been reported previously. Nurse cell expression was divided into two classes. The first class included genes active at an early stage of oogenesis, prior to stage 6 and accounted for 15% of nurse cell expression, while the second class included genes expressed from stage 6 on, but especially at stages 9 and 10, that is, just prior to the nurse cells dumping their cytoplasm into the oocyte. Several lines showed varying levels of expression among the nurse cells indicating that the nurse cells are not all transcriptionally equivalent. Subsets of the follicle cells showed specific staining patterns indicating that these cells are functionally different to their neighbours. Gradient patterns of expression were also observed in the follicle cells.

The P-element mediated enhancer trap system has proved extremely useful in the identification of a wide variety of genes and will provide the basis for considerable progress in our understanding of developmental systems. The P-element also provides a tool for further analysis of a gene by the generation of mutations either by insertion into the gene or by imprecise excision.

More recent developments in P-element have utilised the yeast Gal-4 transcriptional activator in one P-element and a Gal-4-Upstream-Activating-Sequence (UAS) reporter gene construct in a separate P-element (Freeman, 1991). Activation of the Gal-4 activator by nearby gene enhancers results in reporter gene expression from the second P-element to allow identification of tissue and temporal patterns of expression. However, this system is far more potent as the reporter gene can be replaced with other sequences, for example, a toxin to allow targeted cell killing, an antisense construct to nullify gene expression, or ectopic expression of the gene of

interest. A bank of Gal-4 lines allows gene expression in a variety of specific cells and at differing times.

1.3.3 To Make Ovaries, First Define Your Sex - Sex Determination

As oogenesis is a female specific process, the sex of the adult must be correctly established to enable female gametogenesis to occur. Sex determination is a complex process in itself and only a brief discussion of the molecular genetics will be given here as a lead into the molecular genetics of oogenesis.

1.3.3.1 Sex Determination in the Soma

Bridges (1916) observed that flies with an X:0 genotype could develop as males and although these flies were sterile, concluded that a Y chromosome was redundant in sex determination. Bridges was able to establish by 1921 that the ratio of X chromosomes to autosomes (X:A ratio) determined the sex of the fly (see Cline, 1993). Sex determination is currently best defined as an X:A ratio ≥ 1 will give rise to a female fly while an X:A ratio ≤ 0.5 produces a male fly. Flies with an intermediate ratio develop as intersexes with the most striking examples showing large patches of mosaic tissue derived from individual cells which have adopted either a male or a female state. Thus, the primary event for somatic sex determination in the fly is a system for determining the ratio of X chromosomes to the number of autosomes in a cell.

Sex-lethal (*Sxl*) is a gene central to sex determination and was identified as long ago as 1932 by Muller (see Cline, 1993). Muller realised that *Sex-lethal* was required for dosage compensation, that is, the level of transcription of the male X chromosome is, on average, twice that of a female X chromosome to compensate for the reduced dose of X linked genes in the male. However, at that time, the role *Sex-lethal* played in development was not realised.

Genes involved in the primary events of sex determination generally show sex-specific lethality because of dosage compensation effects whereas phenotypes showing altered sexual characteristics are usually the result of mutations in genes downstream of the primary signal. Those genes with a female determining role are X-linked and termed numerator elements while those that are male determining are autosomal and are termed denominators. These terms reflect the observation that

increased doses of such genes in one sex alter the X:A ratio towards the opposing sex.

1.3.3.2 Numerator and Denominator Elements

The first numerator identified was *sisterless-a* (*sis-a*), a gene which showed a female specific lethal phenotype. A recessive loss-of-function allele of *sisterless-a* was shown to be overcome by a gain-of-function *Sex-lethal* allele implying that *sisterless-a* was a zygotically acting positive regulator of *Sex-lethal*. The *sisterless-a* gene was cloned by Erickson and Cline (1993) and encoded a small 189 amino acid protein which had similarity to basic helix-loop-helix transcription factors. Analysis of the expression pattern showed that initial expression occurred in the embryo from nuclear cycle 8 and persists up to stages 12-13.

Genetic analysis identified a second numerator in the neurogenic *achaete-scute* complex (Cline, 1988). A duplication in this region was lethal to male flies also carrying a duplication of *Sex-lethal*. The numerator locus in the *achaete-scute* region was called *sisterless-b* (*sis-b*) and was mapped to the *sc-α* component. Of the 6 transcripts produced by the *sc-α* unit, the T4 transcript was shown to be mostly responsible for *sisterless-b* activity with a protein product that shows similarity to members of the *myc* family. A pattern of expression similar to *sisterless-a* was seen for *sisterless-b*. A genetic screen to identify further numerator elements has yielded the weakly acting *sisterless-c* locus (Cline, 1993).

The *runt* gene is a neurogenic locus which has also been identified as having a role in sex determination. This role was identified as an interaction between a *sisterless-b* deficiency and a null *runt* allele in a *trans*-heterozygous fly resulting in severely reduced female viability (Duffy and Gergen, 1991). This phenotype is suppressed by a dominant feminising *Sex-lethal* allele. Increased doses in triploid intersexes show a feminising effect and decreased doses masculinise the flies indicating that *runt* acts as a numerator element. The *runt* gene may code for a tyrosine kinase and its expression in the embryo is initially in a broad central domain suggesting that other elements may act as numerators at the ends of the embryo (Kania et al, 1990).

The *deadpan* (*dpn*) locus is the only known denominator element and is located on chromosome 2 (Bier et al, 1989). Studies showed that this locus had a stronger

effect on male viability than female and this effect was even more pronounced when combined with additional copies of *sisterless-b* (Younger-Shepherd et al, 1992). Mutations in *deadpan* can be rescued by a female-lethal *Sex-lethal* allele demonstrating that *deadpan* is a regulator of *Sex-lethal*. Expression is seen in the embryo at similar stages to that for *sisterless-a* and *sisterless-b* but subsequent expression becomes confined to neuroblasts showing *deadpan* as also having neurogenic functions.

1.3.3.3 Maternal Effect Genes

It was realised early in the study of *Drosophila* sex determination that maternal effect genes were involved in this process. The *daughterless* mutation showed lethality to female offspring with this effect being traced to a function in oogenesis by use of temperature sensitive mutations (Cline, 1976). These studies also demonstrated that *daughterless* had a zygotic role. The gene product is a protein with similarity to the Bicoid and Paired transcription factors (Cronmiller et al, 1988) and interacts with products of the *achaete-scute* complex. DNA binding activity has been demonstrated when the *daughterless* protein and *achaete-scute* T3 are used in combination but not when these factors are applied individually (Murre et al, 1989). The *Sex-lethal*^{M1} allele rescues *daughterless* mutants showing that there is an interaction between these genes (Cline, 1983). Null alleles show a zygotic requirement for *daughterless* in both sexes for somatic development while a role in the germ-line for egg chamber morphogenesis is also postulated (Cummings and Cronmiller, 1994).

The *hermaphrodite* (*her*) gene exhibits similar effects to those of *daughterless* on female progeny at elevated temperatures (Pultz et al, 1994). This gene shows distinct maternal and zygotic mutant phenotypes; the *herl(2)*^{mat} allele has a maternal phenotype only while zygotic alleles are not rescued by constitutive expression of *Sex-lethal* or *transformer*. The gene encodes a putative zinc finger protein implying a role as a transcription factor (Ryner and Swain, 1995).

Two neurogenic loci with involvement in sex determination are *groucho* and *extramacrochaetae*. The Groucho protein was implicated by identification in a yeast two-hybrid system as interacting with a number of basic helix-loop-helix proteins including Deadpan (Paroush et al, 1994). The *extramacrochaetae* gene encodes a protein with a helix-loop-helix domain but without the DNA-binding basic domain

suggesting that this protein binds to others to form heterodimers that are then unable to activate gene expression (Younger-Shephard et al, 1992).

Two alleles of the *fl(2)d* locus produce differing sex-specific effects; the *fl(2)d*^{1} allele is lethal to females and semi-lethal to males when homozygous whereas *fl(2)d*^{2} is lethal to both sexes when homozygous. The locus has been mapped to region 50A-F on chromosome 2 by deficiency mapping (Granadino et al, 1992) but the gene product is unknown. A role for female specific splicing of *Sex-lethal* is shown by the ability of *Sex-lethal*^{M1} to rescue *fl(2)d* mutants and the detection of male forms of Sex-lethal transcripts in XX;*fl(2)d* flies.

The *sans-fille* gene has roles in *Sex-lethal* expression in both the germ-line and the soma. Constitutively expressing alleles of *Sex-lethal* suppress the *sans-fille* mutation (in the soma) while germ-line clones of *sans-fille* lack sex-specific transcripts of *Sex-lethal*. The gene has been cloned and encodes a protein with similarity to the mammalian U1A snRNP protein suggesting a requirement for the maintenance of *Sex-lethal* expression rather than its initiation (Flickinger and Salz, 1994).

The *virilizer* gene was identified in a screen for female sterile loci on the second chromosome (Schüpbach and Wieschaus, 1989). This gene is required for dosage compensation and sex determination in females and has an unknown vital function in both sexes (Hilfiker et al, 1995). In genetic mosaics, XX cells with mutant *virilizer* differentiate with male attributes indicating that *virilizer* is a positive regulator of female specific *Sex-lethal* splicing.

1.3.3.4 Regulation of Sex-lethal

In response to the factors that have assessed the chromosome number, transcription of the *Sex-lethal* gene begins from the early promoter, P_E, in female embryos. This promoter results in the production of female specific *Sex-lethal* protein. At the blastoderm stage the maintenance or late promoter, P_M, becomes active in both sexes but the presence of *Sex-lethal* protein in the female results in the splicing out of the third exon. This exon contains a stop codon and results in truncated non-functional protein in the male (reviewed in Cline [1993[and MacDougall et al [1995]). The female form of *Sex-lethal* autoregulates its own splicing and also initiates a cascade that results in sex specific proteins from the *doublesex* gene. In

the female the *transformer* gene is spliced by *Sex-lethal* protein to produce active *transformer* protein, itself a splicing factor. The products of the *transformer* and *transformer-2* genes then control the splicing of the *doublesex* gene. Active *doublesex* protein is produced in both males and females but the male version is larger.

1.3.3.5 Sex determination in the Germ-line

The sex determination of the germ-line differs from that of the soma. This observation resulted from transplantation of pole cells from one sex into the other (Marsh and Wieschaus, 1978). Transplantation of pole cells showed that XX cells would give rise to eggs in a female whereas non functioning sperm were formed in males. A reciprocal transplantation of XY germ cells showed spermatogenic development in both males and females (Van Deusen, [1976], Steinmann-Zwicky et al [1989]). These results show that the female germ-line requires an inductive signal in addition to a cell autonomous signal.

Investigations into the role of the genetic cascade required for somatic sex determination, the genes *transformer*, *transformer-2*, *doublesex* and *intersex* show that these genes are not required in the germ-line for sex determination (Schüpbach [1982], Marsh and Wieschaus [1978]). However, *transformer-2* is required for the formation of fully functional sperm later in spermatogenesis (Belote and Baker, 1983). Correct functioning of this cascade in the soma is also required for the inductive signal that a female germ-line must receive for correct development (Nöthiger et al, 1989).

Sex-lethal is also essential for germ-line sex determination although this role appears to differ to that in the soma, based on mutants which affect one process but not the other (Salz et al, 1987)). Genes required for somatic induction of *Sex-lethal*; *sisterless-a*, *sisterless-b* and *daughterless* are not required for its activation in the germ-line (see Steinmann-Zwicky, 1992) Investigations into the expression of *Sex-lethal* in wild type ovaries has shown the protein to accumulate in the cytoplasm of stem cells cystoblasts and 2-cell cysts. The protein is not detectable in later stage cysts but reappears in the nuclei in the centre of the germarium (Bopp et al [1993], Oliver et al [1993]). The *Sex-lethal* protein is not detected in mutations of *ovo*, *ovarian tumor*, *fused* or *sans-fille* placing all of these genes

upstream of *Sex-lethal.* Analysis of *Sex-lethal* transcripts shows the presence of the default male form in mutant alleles of *sans-fille.*

Mutations in genes required for germ-line sex determination in the female demonstrate undifferentiated or spermatogenic phenotypes. To date, six such genes have been identified with this phenotype, *ovo*, *ovarian tumor*, *fused*, and *sans-fille* (also known as *liz* or *fs(1)A1621*) as mentioned above, as well as *fl(2)d* and *bag-of-marbles.*

The *ovarian tumor* gene has cloned and analysed at the molecular level. Mutations in *ovarian tumor* can be placed into three classes; quiescent - where no germ-line forms, oncogenic - where large numbers of undifferentiated germ cells are found, and differentiated - where oogenesis begins but becomes blocked at an early stage (Sass et al, 1995). These different states suggest multiple roles for *ovarian tumor* in the germ-line. Alternative splicing produces two isoforms; a 98kDa protein and a 104kDa protein (Steinhauer and Kalfayan, 1992). The 104kDa form can rescue all classes of mutant whereas the 98kDa form only rescues the least severe, differentiated class. Constitutive expression of *Sex-lethal* rescues the oncogenic class, not the quiescent or differentiated classes (Sass et al, 1995), although Rodesch et al (1995) claim that constitutive *Sex-lethal* rescues *ovarian tumor* mutants that lack the 104kDa protein, that is, all classes. The amino acid sequence does not predict a function but shows similarity to that deduced from the *bag-of-marbles* gene.

The *ovo* gene is genetically complex with both dominant and recessive mutations showing an essential requirement in the germ-line (Oliver et al [1987], Oliver et al [1990]). However, the *shaven-baby* and *lozenge-like* alleles show defects in the cuticle and in eye development respectively (Ménel-Ninio et al, 1991). The gene encodes a putative transcription factor of 131kDa containing four zinc fingers. Use of an *ovo-lacZ* reporter transgene in pseudofemales has demonstrated that *ovo* expression in the female germ-line is dependent on a female karotype in the germ-line (Oliver et al, 1994).

The *sans-fille* gene is likely to be involved in the maintenance of *Sex-lethal* splicing in the soma and this would also appear to be its role in the germ-line. The *fused* gene encodes a putative serine/threonine kinase and is a possible candidate in soma to germ-line communication (Oliver et al, 1993). Although alleles of the *bag-of-*

marbles gene have tumourous ovary phenotypes, it would appear to be involved in control of cystoblast and cystocyte division rather than germ-line sex determination, as will be described later.

1.3.3.6 Summary of Sex Determination

Sex determination in the soma is reasonably well understood. The primary control is *Sex-lethal* with activation of this gene being controlled at the level of transcription. The developing embryo is able to assess its chromosome content and respond by activating *Sex-lethal* accordingly with only females producing an active protein. The *Sex-lethal* signal is transduced at the level of splicing to control the female-specific splicing of the *doublesex* product.

Sex determination in the germ-line is less well understood but it is clear that two pathways are involved. First, the karotype of the germ-line must be XX for correct female development and second, a signal is required from the surrounding somatic tissue. The genes *Sex-lethal*, *ovo* and *ovarian tumor* have major roles in female germ-line development but many other genes have yet to be identified. Sex determination in both the soma and germ-line is summarised in figure 1.2.

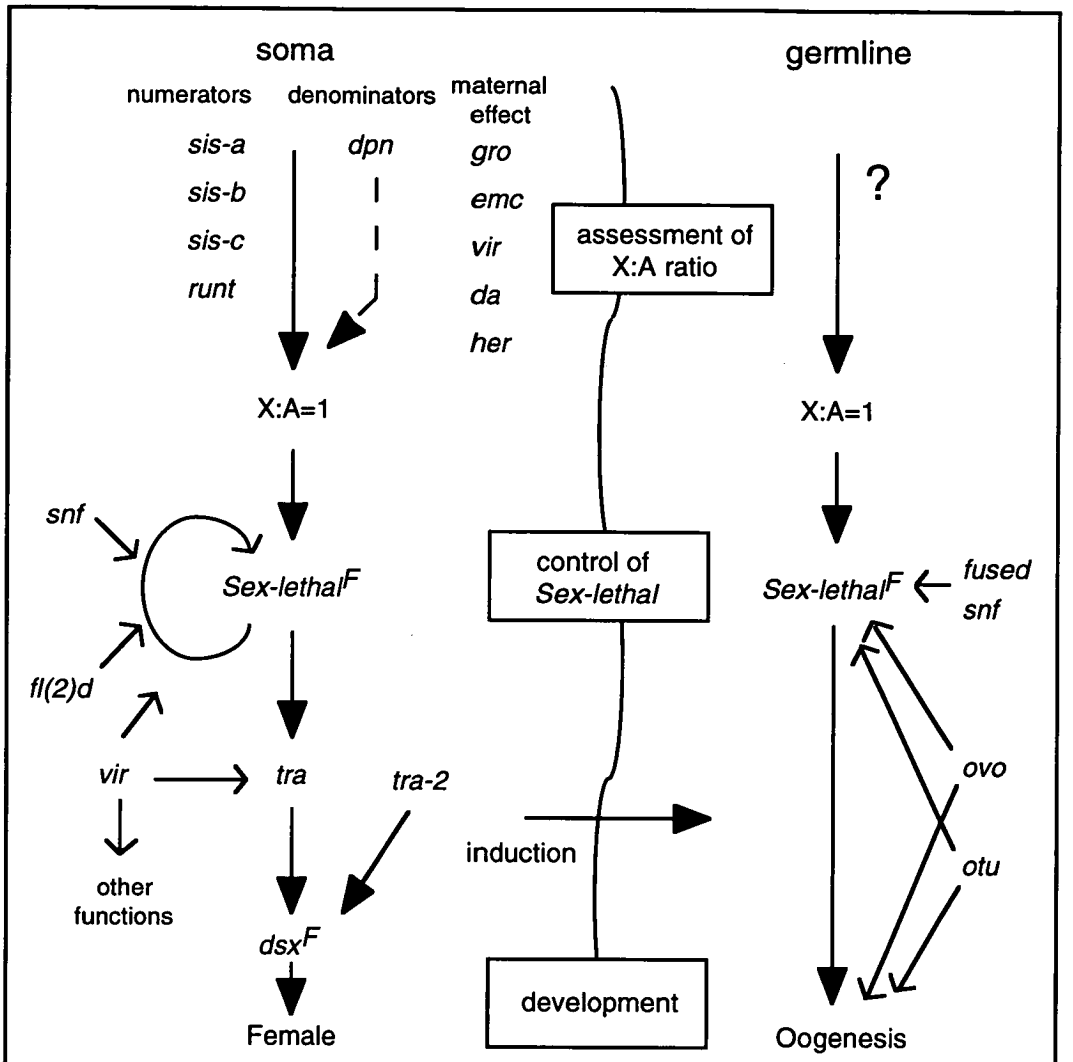


Figure 1.2 Summary of sex determination in the female (adapted from MacDougall et al, 1995). See text for details.

1.3.4 Determining the Oocyte

The earliest stages of oogenesis occur in the germarium. The germarium can be considered in 4 distinct cytological regions; region 1, region 2a, region 2b, and region 3 (Mahowald and Kambysellis [1980], Cooley and Therkauf [1994]). In region 1 the germ-line stem cells are dividing to produce cystoblasts. These cystoblasts undergo 4 incomplete divisions to produce an interconnected syncytium or cystocyte. The final division to produce the 16 cell cyst marks the beginning of germarial region 2a. Two cells within the 16 cell cyst contain 4 intercellular connections, the ring canals and either can become the oocyte. By region 2b the

oocyte is morphologically differentiated from the nurse cells. The pro-oocyte occupies the centre of the 16 cell cyst in region 2b but by region 3 it has taken up position at the posterior pole. Region 3 of the germarium is also called a stage 1 egg chamber. A stage 2 egg chamber is formed when follicle cells surround the 16 cell cyst and it buds from the germarium.

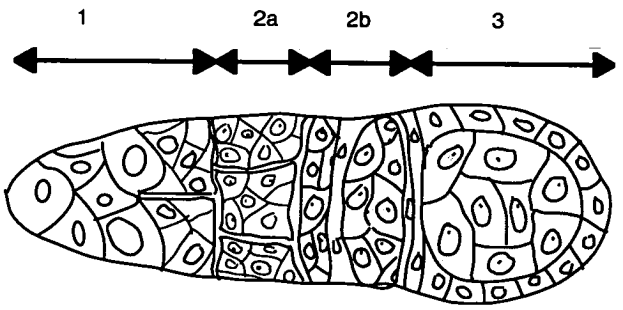


Figure 1.3 A schematic representation of the germarium adapted from Christerson and McKearin (1994). See text for complete description.

The fusome is a structure that is observed in region 1 and 2a of the germarium (Lin et al, 1994). It is composed of residual mitotic spindle components arising from incomplete mitosis during cystocyte formation. The fusome is seen as a distinct region of cystocyte cytoplasm extending through the ring canals into all cells in the cystocyte and contains vesicles but lacks ribosomes or mitochondria. Once the ring canals have matured the fusome disappears. The fusome is proposed to be involved in a number of roles in oocyte development; 1) physical blockage of cytokinesis, 2) controlling the pattern of cystocyte interconnections by orienting the mitotic spindles, 3) to synchronise cystocyte divisions, and 4) may determine the oocyte by creating polarity through unequal divisions.

The early stages of gametogenesis appear similar between males and females indicating some shared processes between the sexes. The *bag-of-marbles* gene was identified in a P-element screen that resulted in both male and female sterility with the ovaries exhibiting abnormal germaria and testes cells that failed to progress from the primary spermatocyte stage (McKearin and Spradling, 1990). The transcriptional unit was identified from a chromosomal walk and confirmed by its ability to rescue the mutant phenotype in transformed flies. The deduced protein sequence of 442 amino acids contains PEST sequences, potential targets for

proteolytic degradation (Rogers et al, 1986) and weak similarity to the *ovarian tumor* protein. The transcript is seen shortly after stem cell division and then in stage 10b egg chambers as well as the early embryo. The early stem cell expression suggests a role in control of cystoblast division and a function in the precise control of cell divisions explains its presence in the syncytial embryo.

The *encore* gene was identified as a female sterile locus by P-element mutagenesis (Hawkins et al, 1996). The recessive $encore^{BB}$ allele shows three classes of mutation when homozygous; the most common possessing twice the normal number of germ cells, seen as 31 nurse cells and 1 oocyte, while a less frequent class shows bipolar egg chambers with additional germ cells and incorrect positioning of the oocyte. The least frequent class of mutation shows fewer than normal germ cells and no oocyte. Since the last two classes of mutation are often seen as adjacent egg chambers containing a total of 32 germ cells, it seems reasonable to predict that the egg chamber that contains less than 16 germ cells is caused by incorrect follicle cell migration during the pinching off of the egg chamber.

The *encore* mutant phenotype arises from an extra round of mitotic cell division in the cystocyte as demonstrated by 2 cells within the 32 germ cell egg chamber carrying five ring canals instead of the normal four (Hawkins et al, 1996).

The *encore* gene appears to interact with the *bag-of-marbles* (*bam*) gene with the expression pattern of *bag-of-marbles* being more expanded in an *encore* mutant than in the wild type. A *bag-of-marbles-encore* double mutant exhibited the tumorous ovary phenotype of *bam* indicating that *bam* is a dominant suppresser of *encore*. These results suggest that *bam* protein influences the cystoblast mitotic cell divisions and that *encore* is required to regulate *bam*. Mutations in *encore* also affect the oocyte nucleus with it becoming more like the nuclei of the nurse cells and showing polyploidy, especially in egg chambers with a normal complement of nurse cells. This observation suggests that *encore* is also required for oocyte maintenance.

Another gene required for germ cell development is *stonewall* (Clark and McKearin, 1996). This gene was identified from a P-element mutagenesis screen and encodes a nuclear protein with a DNA binding domain that shows similarity to the Myb and Adf-1 transcription factors. The $stonewall^{\Delta 95}$ allele, which produces

transcripts but not protein, only affects female ovaries. In this mutant, stem cells divide correctly to produce cystoblasts and cystocytes but the oocyte fails to differentiate implicating a function in oocyte maintenance for the *stonewall* gene.

The *Bicaudal-D* gene, when mutant, produces a number of phenotypes ranging from sterility to the double abdomen embryos which give the gene its name. Sterile mutants produce egg chambers containing 16 nurse cells with no oocyte being specified. Bicaudal embryos are formed because of incorrect localisation of posterior patterning determinants stemming from the *Bicaudal-D* gene.

The *Bicaudal-D* gene has been characterised at the molecular level and encodes a protein with extensive coiled-coil α-helices similar to those in the tail regions of type II myosins (Wharton and Struhl, [1989], Suter et al [1989]). The mRNA from this gene accumulates in the pro-oocyte in germarial region 2a, a process which is dependent on the correct function of the gene itself as mutants producing a 16 nurse cell cyst fail to localise the mRNA (Suter and Steward, 1991). Later results have shown that the localisation of the *Bicaudal-D* protein to the pro-oocyte is dependent on its phosphorylation (Suter and Steward, 1991). Mutations in *Bicaudal-D* that lead to failure in phosphorylation of the protein show loss of protein localisation to the oocyte and failure of oocyte differentiation.

Ran et al (1994) generated null mutations in the *Bicaudal-D* locus and determined that *Bicaudal-D* is essential for oocyte determination, but not for formation of ring canals. These authors also determined that *Bicaudal-D* has a zygotic role and observed the presence of protein in other tissues suggesting a wider role for this gene in the fly.

In severe *Bicaudal-D* mutants the protein distribution is altered with no oocyte development taking place implying that cytoskeletal formation has a role in determining the oocyte. Further evidence supporting this hypothesis comes from Theurkauf et al (1993) who show that the microtubule organising centre (MTOC) forms in the presumptive oocyte shortly after completion of the 16 cell cyst. This MTOC polarises the egg chamber and it is interesting to note that mutations in *Bicaudal-D* disrupt its formation.

The *orb* gene is expressed very early in oogenesis with the transcript localised to the presumptive oocyte during region 2 development of the germarium and then at

both the posterior and anterior of the oocyte at later stages of oogenesis (Lantz et al, 1992). This localisation is dependent on the correct functioning of *Bicaudal-D* but is independent of *oskar* localisation. The *orb* gene encodes a 99kDa protein with two RRM RNA binding domains that differ suggesting different specificity for RNA. Females exhibit a 4.7kb transcript limited to the germ-line while males express 3.2kb and 2.0kb transcripts in late 3rd instar gonads and in the testes. Mutations in *orb* cause defects in abdominal development and can ventralise the embryo showing that *orb's* role later in oogenesis is in the localisation of determinants such as *oskar* and *gurken* (Christerson and McKearin, 1994). There is also an essential function for *orb* early in oogenesis as another set of mutations cause oogenesis to arrest in the germarium (Lantz et al, 1994).

Cytoplasmic bridges are a common feature among developing gametes. In spermatogenesis the role of these bridges may be to synchronise sperm development and to equalise the distribution of gene products (Spradling, 1993b). The role of the cytoplasmic bridges, or ring canals, in the interactions of the *Drosophila* nurse cells and oocyte has been well established (Robinson et al, 1994). Several genes involved in the establishment of the ring canals have been identified and characterised. Xue and Cooley (1993) have described the cloning of the *kelch* gene. In homozygous mutant lines, normal cytoplasmic transport from the nurse cells to the oocyte is disrupted, however, uptake of yolk protein is normal. The ring canals are present and appear normal as shown by actin staining. Interestingly, the *kelch* gene appears to be polycistronic, containing an in-frame stop codon. Immunoblots detected 2 proteins, a strong 80kDa band, the product of the first ORF, and a weak 180kDa band which suggests partial suppression of the stop codon. Antibodies to *kelch* protein localise it to the ring canals leading these authors to propose that the *kelch* protein is involved in the regulation of cytoplasmic flow.

The *Drosophila* gene encoding the homologue of the mammalian cytoskeletal adducin protein has been reported by two groups (Ding et al [1993a] and, Yue and Spradling [1993]). There are two forms of human adducin, α and β, however, *Drosophila* appears to have only one which shares homology with both human forms. Ding et al (1993a) showed that the *adducin-like* mRNA was localised at the anterior of the oocyte and that this localisation was dependent on the *swallow* gene. In humans, adducin promotes the association of spectrin and F-actin, both cytoskeletal components. Therefore, it is reasonable to assume that *adducin-like*

(or *hu-li tao shao* [*hts*] as named by Yue and Spradling, 1993), will be involved in the formation and maintenance of the egg chamber cytoskeleton in *Drosophila.* The *hu-li tao shao* protein is a component of the fusome as demonstrated by antibody staining (Lin et al, 1994). The protein is seen in the follicle cells around the membrane, whereas in the germ-line internal structures are stained. This staining begins with a perinuclear spherical structure in the stem cells and progresses to an elongated and branched structure in 2-8 cell cysts. The mammalian Adducin protein shows association with α-spectrin and consistent with this observation, antibodies detect α-spectrin in these fusomal structures. The fusome is also seen in the germarium of *Bicaudal-D* and *egalitarian* mutants showing that these genes do not control its formation.

During stage 10b of oogenesis the nurse cells dump most of their cytoplasm into the oocyte. In *chickadee* mutants this process is disrupted because the ring canals become blocked with the nurse cell nuclei (Cooley et al, 1992). The *chickadee* gene has been shown to encode a homologue of profilin, a protein involved in the assembly of actin filaments. It has been suggested by these authors that the subcortical actin filaments contract to propel the nurse cell cytoplasm into the oocyte while a second, independent, cytoskeletal set of actin filaments anchor the nurse cell nuclei. The *chickadee* encoded profilin acts exclusively with this second set of actin filaments as cytoplasmic streaming still occurs in *chickadee* mutants.

The *quail* gene also produces a similar phenotype to *chickadee* with the molecular cloning of the gene showing that it encodes a protein with similarity to vertebrate Villin, a protein that can cap, nucleate, sever and bundle actin filaments (Mahajan-Miklos and Cooley, 1994). The predicted protein of 887 amino acids contains six repeats characteristic of actin binding proteins. Interestingly, the *quail* product is more divergent from vertebrate Villin than a similar *Drosophila* actin binding protein, Gelsolin, is from the vertebrate version of Gelsolin. This finding suggests that there are subfamilies of Villin, a finding supported by antibodies to vertebrate Villin recognising a protein in the larval midgut. This protein does not correspond to the *quail* product as *in situ* hybridisation with *quail* does not detect a transcript in the midgut tissue.

A third actin binding protein comes from the *singed* gene, originally named for its mutant bristle phenotype, but a gene which also shows female sterility (Cant et al, 1994). This female sterility manifests itself as a blockage of the ring canals by the

nurse cell nuclei during stage 10 of oogenesis. The *singed* gene product shows similarity (35% identity and 67% similarity) to ectinodermal Fascin, a protein found in actin bundles. The *singed* protein is detected by antibodies in cell extensions of follicle cells that are undergoing migration, for example, the border cells and the centripetal follicle cells migrating around the oocyte anterior. Female sterile mutant phenotypes and mosaic studies also show a requirement for the *singed* gene in the germ-line.

1.3.5 Regulation of the Chorion, Vitelline Membrane, and Yolk Protein Genes

During the later stages of oogenesis the oocyte increases rapidly in size due to uptake of yolk proteins and transfer of cytoplasm from the nurse cells. After this size increase, the vitelline membrane and chorion (eggshell) are synthesised. The expression of the yolk proteins, chorion and vitelline membrane genes is both temporally and spatially regulated. These genes are, therefore, good candidates for the study of molecular mechanisms controlling this regulation.

The yolk protein genes are regulated by two hormones, juvenile hormone and 20-hydroxyecdysone. Both hormones can act on the fat body to stimulate yolk protein synthesis but only juvenile hormone acts on the follicle cells. This action may be indirect as yolk protein expression in the follicle cells appears to be dependent on stimulation of yolk protein uptake from the fat body (reviewed by Bownes, 1994).

Two genes identified to date which are involved in this regulation are *apterous* and *cricklet*. The *apterous* gene is a putative transcription factor but appears to have an indirect role in juvenile hormone signalling as it is not expressed in the brain tissue where juvenile hormone is synthesised, the corpus allatum. However, mutations in *apterous* cause, among other defects, female sterility with oogenesis blocked at stage 7-8 (Cohen et al, 1992). The *cricklet* gene has been proposed to code for a juvenile hormone carrier protein as the female sterile phenotype cannot be rescued with a juvenile hormone analogue (Shirras and Bownes, 1989).

The 3 single copy yolk protein genes are located on the X-chromosome (Barnett et al, 1980). The *yp1* and *yp2* genes are separated by a region of 1.2kb and show divergent expression with the 5' ends facing each other. The *yp3* gene is located more than 1000kb away. (reviewed in Slee and Bownes, 1990).

Molecular studies using reporter gene constructs have begun to identify and localise the mechanisms involved in the expression of the *yp1* and *yp2* genes. A fat body enhancer of both genes has been mapped to a 125bp element that starts at a position 196bp from the cap site of the *yp1* gene. An enhancer element that controls expression of both genes in the follicle cells has been mapped to a 181bp element that is 159bp from the *yp2* cap site. An additional enhancer that affects expression of *yp1* has been located in the first exon of *yp2* . In addition, a novel DNA binding protein with high specificity for a 31bp element that maps 148bp downstream from the *yp1* transcriptional initiation site has been discovered.

In addition to hormonal regulation, the yolk protein genes are also influenced by nutritional conditions. Regulatory elements for 20-hydroxyecdysone have been mapped for *yp3* in upstream sequence, within coding and intron sequence and in sequence 3' to the gene (Bownes et al, 1996). Nutritional regulatory regions have been mapped within the intergenic region of *yp1* and *yp2* (Sondergaard et al, 1995). These genes are also regulated by the sex determination pathway, with both male and female specified versions of the *doublesex* gene product being shown to bind in the intergenic region. As the male form of *doublesex* protein is larger than the female form it is possible that the additional male-specific domain represses binding by other enhancers or transcription factors.

Analysis of the *yp3* gene is less advanced but the fat body and ovarian regulatory elements are contained within a 747bp region 5' to the coding sequence. Two independent elements of 328bp and 419bp within this region have been shown to govern ovarian and fat body expression respectively (Ronaldson and Bownes, 1995).

The chorion is a multilayered structure composed of about 20 different proteins synthesised in the follicle cells during stages 11-14 of oogenesis. The six major proteins, now referred to as Cp15, Cp16, Cp18, Cp19, Cp36 and Cp38, are clustered at position 7F of the X-chromosome (Cp36 and Cp38) and position 66D of the 3rd chromosome (Spradling et al, 1980). High demand for chorion proteins results in selected amplification of the genes in the follicle cells (Delidakis and Kafatos, 1987). Regulatory elements of the Cp36 gene from the X-chromosome were investigated using a P-element reporter gene construct. The results showed distinct *cis*-acting elements regulating anterior and posterior expression of the reporter gene (Tolias and Kafatos, 1990).

Not only are the chorion proteins tightly regulated during oogenesis, they are also regulated with respect to each other with those on the X chromosome expressed before those on the 3rd chromosome. Expression begins in the follicle cells at the anterior-dorsal region and appears to be partially under the control of the dorsal-ventral axis specification pathway (Tolias et al, 1993).

The minor chorion proteins are believed to be required for formation of the specialised eggshell structures; the micropyle, aeropyle, opuculum and dorsal appendages (Parks and Spradling, 1987).

The vitelline membrane is the innermost part of the eggshell and is synthesised during stages 9 and 10 of oogenesis (Fargnoli and Waring, 1982). The vitelline membrane is composed of the products of at least six genes, all on chromosome 2, four at 26A and one each at 32E and 34C. Analysis of *cis*-acting sequences of one 26A gene has identified a 170bp element responsible for follicle cell expression while another element was shown to reduce expression in the border cells (Jingman and Petri, 1993).

1.3.6 Gurken, a Signal that has Two Roles

A role for *gurken*, a gene coding for a TGF-α-like signalling molecule and its receptor *torpedo*, a gene coding for an epidermal growth factor receptor (EGF), in defining the dorsal-ventral axis had been established (reviewed in St Johnston and Nüsslein-Volhard, 1992) before it was realised that these genes also functioned in formation of the anterior-posterior axis. The original theory for anterior-posterior axis formation in the oocyte involved direct signalling from the posterior follicle cells mediated by the *Notch* and *Delta* system (Ruohala-Baker et al, 1991).

Investigations of *spindle-C* mutants showed that anterior type follicle cells could appear at both poles of the egg chamber when the pro-oocyte failed to migrate to the posterior pole (González-Reyes and St Johnston, 1994). This finding suggested that the pro-oocyte was involved in defining the posterior pole of the egg chamber. Use of an enhancer trap line, L53b, and the *slow-border-cells* (*slbo*) gene as markers for anterior type follicle cells showed that *gurken* mutants had anterior type follicle cells at the posterior pole indicating a role for *gurken* in defining the anterior-posterior axis (González-Reyes et al, 1995). Additional evidence comes from flies carrying hypomorphic *torpedo* mutations in conjunction

with one copy of the *rolled* gene, a MAP-kinase which functions downstream of *torpedo*. The hypomorphic *torpedo* alleles allow the flies to survive to adulthood but females show obvious defects in the anterior-posterior polarity of developing egg chambers (González-Reyes et al, 1995). The anterior follicle cell markers, L53b and *slow-border-cells*, were used to investigate whether other genes affecting polarity had a role in establishing the anterior-posterior axis. The genes investigated were: *orb*, *cappuccino* and *spire* which affect polarity of both axes; *fs(1)K10*, which causes a dorsalised phenotype and *cornichon*, a ventralising mutant. Only *cornichon* showed an effect producing a similar phenotype to that seen by *gurken* and *torpedo* mutants (González-Reyes et al, 1995).

The *cornichon* gene has been isolated and analysed at the molecular level by Roth et al (1995) who also described the requirement for *gurken* and *torpedo* in determining the anterior-posterior axis. The *cornichon* protein is small at 144 amino acids and is highly conserved between *D melanogaster* and *D virilis* even though these species are thought to have diverged some 60 million years ago. The *cornichon* protein sequence shows a large hydrophobic domain which may be membrane spanning suggesting a role in the regulation of the *gurken* membrane spanning signalling protein (Roth et al, 1995).

Studies on the formation of the anterior-posterior axis (González-Reyes and St Johnston, 1994, González-Reyes et al, 1995 and Roth et al, 1995) show that the pro-oocyte nucleus fails to migrate from the posterior pole to an anterior-dorsal position, where signalling for dorsal-ventral axis formation occurs, in *gurken*, *torpedo* and *cornichon* mutant lines. A Kinesin-β-galactosidase fusion protein (Clark et al, 1994) was used to identify the ends of the microtubule array showing that rearrangement of the microtubule cytoskeleton accompanies the pro-oocyte nucleus migration from the posterior pole to an anterior-dorsal position.

Thus, a model for establishing the anterior-posterior axis followed by the dorsal-ventral axis mediated by the same signalling and receptor molecules can be envisaged. Initially, the pro-oocyte migrates to the posterior pole of the egg chamber and the microtubules are arranged with their minus-ends at the posterior of the pro-oocyte which localises the pro-oocyte nucleus there. Signalling to the posterior follicle cells via the *gurken*, *torpedo* and *cornichon* system causes these follicle cells to adopt a differentiated posterior state. These posterior follicle cells then signal back to the pro-oocyte by the *Notch* and *Delta* system to begin

rearranging the microtubule cytoskeleton transferring the pro-oocyte nucleus to the anterior-dorsal position where once again *gurken*, *cornichon* and *torpedo* proteins signal to a group of follicle cells to undertake a dorsal fate.

The signal-receptor mediated pathway involving *gurken* and *torpedo* in determining dorsal-ventral polarity was established prior to the role of these molecules in anterior-posterior axis formation. Germ-line activity of *cappuccino*, *spire*, *cornichon*, *fs(1)K10* and *squid* is required to control the gurken signal received by a subset of follicle cells at the anterio-dorsal region of the oocyte via the *torpedo* protein. The *spitz* gene product also shows similarity to TGF-α molecules and Ruohola-Baker et al (1993) originally proposed that the product of the *spitz* gene was the molecule which signalled to the *torpedo* product. It is possible that this molecule is used for an independent signalling pathway.

The product of the *rhomboid* gene, a putative transmembrane protein, is involved in this regulation either by improving follicle cell-oocyte contact or by direct involvement in the signalling pathway (Ruohola-Baker et al, 1993). A further gene that has been shown to be involved in this pathway is the neurogenic locus *brainiac* which is also required in the germ-line in determining the signal to the *torpedo* product (Goode et al, 1992). The *fs(1)K10* RNA is localised specifically to the oocyte nucleus possibly by *cappuccino* and *spire* activity (Cheung et al, 1992). Mutations in *fs(1)K10* show a dorsalised phenotype with the role of the *fs(1)K10* product proposed as being to negatively regulate the signal being supplied from the oocyte, whereas *torpedo* (loss-of-function), *cornichon* and *gurken* mutations are ventralising - no dorsalising signal is formed. The *squid* gene also produces a dorsalised phenotype when mutant (Kelley, 1993). The product of this gene is a putative RNA binding protein that may act by either post-transcriptionally modifying the mRNA encoding the signal or by sequestering the mRNA (Kelley, 1993). On the basis of this information it appears that follicle cells which do not receive the *torpedo* mediated signal adopt a ventralised state by default.

Several factors that operate downstream of the initial dorsal-ventral signalling have been identified including the D-raf serine-threonine kinase and the CF2 transcription factor. A role for D-raf in transmitting the Toll mediated signal has been established by Brand and Perrimon (1994). A Gal-4 enhancer trap system was used to target gain-of-function D-raf to the follicle cells with activation resulting in a dorsalised phenotype.

The CF2 transcription factor is cleared from the anterior-dorsal follicle cells most probably as a result of *gurken-torpedo* mediated signalling leading to its phosphorylation (Hsu et al, 1996). Thus, the CF2 transcription factor is able to act in the ventral follicle cells although it is not known at this stage whether it functions as a direct activator of genes or as a repressor.

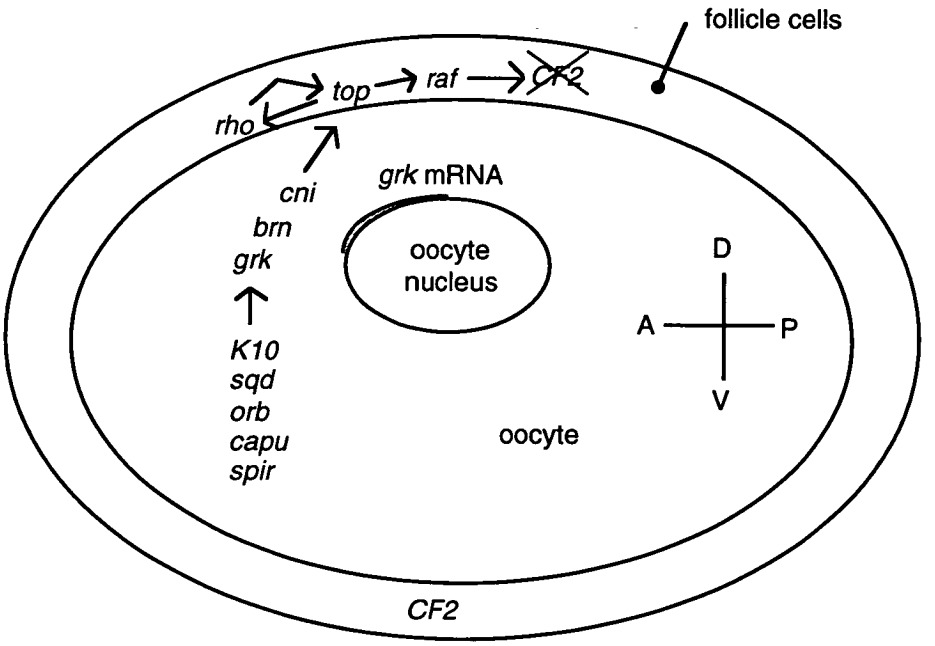


Figure 1.4 Summary of the early events for establishing the dorsal-ventral axis by the *gurken* (*grk*) and *torpedo* (*top*) signalling pathway. *fs(1)K10* (*K10*), *squid* (*sqd*), *orb*, *cappuccino* (*capu*) and *spire* (*spir*) are required to localise the *gurken* mRNA to the dorsal-anterior side of the oocyte nucleus. Production of Gurken signal requires the function of *cornichon* (*cni*) and *brainiac* (*brn*) genes. The Gurken signal activates Torpedo in the follicle cells which, in turn, activates a Ras signalling pathway. The *CF2* transcription factor is only active in the ventral follicle cells as a result of this signalling pathway. Adapted from Morisato and Anderson (1995).

1.3.7 Embryonic Body Patterning and the Maternal Genes that Control It

One area that has received considerable attention in recent years is the formation of the embryonic body pattern which has been shown in *Drosophila* to be controlled by maternally derived gene products. Many of the genes involved in this process have been identified and the mechanisms of their action has been elucidated. This process has been comprehensively reviewed by St Johnston and Nüsslein-Volhard (1992).

Once fertilised, an egg develops through an embryonic stage to emerge as a larvae after about 24 hours. After fertilisation the zygotic nucleus undergoes a series of rapid divisions. After nine divisions the nuclei migrate to the surface of the egg to form a syncytial blastoderm. The pole plasm is a specialised and distinct region of cytoplasm at the posterior pole containing the polar granules. At the syncytial blastoderm stage there are 3-4 nuclei in the pole plasm which form pole cells and ultimately the germ-line. The remaining nuclei are surrounded by membranes after undergoing a further four divisions at the surface giving rise to the formation of the blastoderm.

1.3.7.1 Anterior

Manipulations of the egg cytoplasm showed that it contained maternally derived determinants. Removal of some material from the anterior pole showed defects in the development of the head and thoracic segments. Similarly, removal of cytoplasm from the posterior pole showed defects in abdominal development but interestingly, not in the unsegmented most posterior region (Frohnhöfer et al, 1986).

A number of female sterile mutant lines were identified which showed defects in developing embryos affecting the embryonic body pattern. These mutations have allowed the identification of the genes involved in the control of body pattern formation.

The anterior determinant is encoded by the *bicoid* gene. RNA from the *bicoid* gene is localised at the anterior pole of the egg (Berleth et al, 1988). Once the egg is fertilised, the RNA is translated and the gene product diffuses towards the posterior

pole forming a gradient. The *bicoid* gene product contains a homeodomain defining it as a direct regulator of transcription. It has been shown to control the expression of zygotic genes in the anterior region of the embryo. One of these zygotic targets is the *hunchback* (*hb*) gene (Tautz et al, 1987). In the absence of *bicoid*, no *hunchback* expression is seen. The *bicoid* gene product has been shown to bind to high and low affinity sites in the promoter for the *hunchback* gene (Driever and Nüsslein-Volhard, 1989). Evidence suggests that the *bicoid* product acts as a morphogen; high concentrations of protein can activate high and low affinity sites whereas low concentrations of protein will only activate the high affinity sites. Furthermore, the anterior pattern extends towards the posterior when the dosage of *bicoid* is increased. The zygotic target genes are activated in a specific region of the embryo through a combination of these *bicoid* protein binding sites so that they only respond to a certain concentration of *bicoid* protein. The *hunchback* gene product is also a transcriptional regulator and has been shown to repress the expression of the gap gene *Krüppel* in the anterior region of the embryo. However, the *hunchback* gene is only required for determining part of the anterior region which suggests that there are other targets for the *bicoid* gene product which act more anteriorly than *hunchback*. Candidate genes are *orthodenticle*, *empty spiracles*, *buttonhead*, and *giant*, with each being required in early embryonic development. As yet none have been shown to be directly regulated by the *bicoid* product (St Johnston and Nüsslein-Volhard, 1992).

1.3.7.2 The Posterior

The pole plasm contains polar granules, electron dense structures consisting of protein and RNA at the very posterior of the oocyte. During embryogenesis the pole plasm directs a small subset of cells, the pole cells, to form the germ-line in the developing fly (Mahowald, 1968). The pole plasm contains two signals; the pole cell determinant and the posterior determinant responsible for abdominal development. A set of genes called the posterior group have been identified through extensive screens as being required for the formation of pole plasm (reviewed in St Johnston and Nüsslein-Volhard, 1992). Mutations in many of the posterior group of genes cause defects in both abdominal development and pole plasm formation. These genes are *cappuccino*, *spire*, *staufen*, *oskar*, *valois*, *vasa*, *tudor* and *mago nashi*. Two genes which show abdominal defects but no effect on the pole plasm when mutant are *pumilio* and *nanos*. The *germ-cell-less* gene is probably required only for pole cell formation (Jongens et al, 1992).

Continued investigations into the molecular nature of the posterior group of genes has allowed a picture to emerge of the roles of several members and the molecular interactions which occur.

The primary posterior determinant is the product of the *nanos* gene (Wang and Lehmann, 1991) as shown by injection of *in vitro* synthesised *nanos* mRNA into the abdominal region of mutant embryos. Further experiments have shown that the *nanos* mRNA rescues abdominal defects produced by all of the posterior group of genes.

The *nanos* gene product appears to act in a similar fashion to the *bicoid* product in that a diffusing protein gradient is formed emanating from the posterior pole (Smith et al, 1992). The expression of the gap genes *knirps* and *giant* are dependent on *nanos*, but unlike *bicoid* and *hunchback*, this is an indirect relationship.

Recent experiments have determined that the *caudal* gene functions within the embryo to form a posterior to anterior gradient which is the opposite of that formed by *bicoid* protein and that *caudal* activity is required for the activation of the gap genes *knirps* and *giant* (Rivera-Pomar et al, 1995). Interestingly, the activation of *caudal* mRNA requires *bicoid* protein with cross-linking experiments determining that *bicoid* protein binds to the *caudal* 3'UTR (Dubnau and Struhl, 1996). Further evidence comes from the work of Rivera-Pomar et al (1996) who show that in *bicoid* mutants a *caudal* protein gradient fails to form and that *bicoid* protein can be immunoprecipitated from cross-linked *caudal* mRNA in a 83kDa-71kDA complex along with a 120kDa protein. This result suggests that at least three proteins, one of which is Bicoid, are required for correct *caudal* function.

Creation of germ-line clones lacking both *nanos* and maternal *hunchback* has shown the role of *nanos* protein is to repress the expression of maternally derived *hunchback* transcripts whose product would otherwise act as a transcriptional repressor of both *knirps* and *giant*. The *hunchback* gene is transcribed during oogenesis with the mRNA being found evenly distributed throughout the oocyte. In wild type embryos the maternally derived *hunchback* product is degraded in the posterior half, while *nanos* mutants show an even distribution of both *hunchback* mRNA and protein throughout the embryo (Hülskamp et al, [1989], Irish et al [1989]).

A sequence motif in the 3'UTR of the *hunchback* mRNA has been shown to mediate inactivation by the *nanos* gene product (Wharton and Struhl, 1991). Removal of these motifs, termed *nanos* response elements (NREs) from a *hunchback* transgene resulted in expression of the *hunchback* protein throughout the embryo.

Gavis and Lehmann (1994) have extended the studies on *nanos* translational control by showing that the *oskar* gene is limiting. Increasing the dose of *oskar* from 2 to 4 copies results in more *nanos* protein being translated at the posterior pole although *nanos* mRNA levels remained the same. Replacement of the *nanos* 3'UTR with that from the *α-tubulin* gene created a dominant female sterile mutation with embryos showing defects consistent with ectopic expression of the posterior determinant. A *bicoid* 3'UTR attached to the *nanos* coding sequence produces bicaudal mutant embryos by localising the *nanos* mRNA to the anterior of the oocyte. Inclusion of the *nanos* 3'UTR to make a *nanos-bicoid3'UTR* transgene results in the anterior localisation of *nanos* mRNA but not its translation showing that the *nanos* 3'UTR controls the translation of its mRNA. This result is interesting as the *bicoid* 3'UTR appears dominant over the natural *nanos* 3'UTR in terms of mRNA localisation.

A further role for *nanos* function has been established in the proper formation of the germ cells in the developing embryo (Kobayashi et al, 1996). Pole cells which lack *nanos* activity fail to migrate to the gonad. Gene expression in these mutant pole cells is observed much earlier than in the wild type where gene expression is not seen until the pole cells reach the gonad.

1.3.7.3 The oskar Gene is the Posterior Determinant

Transplantation and injection experiments showed that *oskar* was the determinant for abdominal and pole cell formation (Sander and Lehmann, 1988). Two groups reported the isolation of the *oskar* gene (Ephrussi et al [1991] and Kim-Ha et al [1991]) with several alleles identified as being defective in RNA localisation while others were defective in posterior function. The *oskar* mRNA is seen localised to the oocyte in germarial regions 2b-3 (it is frequently used as an oocyte marker in experiments on oocyte determination) and is localised throughout the oocyte from stages 1-6, but by stage 9 is localised to the posterior of the oocyte. The genes *cappuccino*, *spire* and *staufen* were all determined to be required for posterior

localisation and it was shown that *nanos* function was dependent on *oskar*. In addition, the nonsense mutants showed that maintenance of the *oskar* mRNA at the posterior was dependent on *oskar* protein (Ephrussi et al [1991] and Kim-Ha et al [1991]).

The role of *oskar* as pole cell and abdominal determinant was confirmed by mislocalising it to the anterior by coupling the *bicoid* 3'UTR to the *oskar* coding sequence and introducing this construct as a transgene. Mislocalisation results in bicaudal double abdomen embryos which also have formed pole cells at the anterior (Ephrussi and Lehmann, 1992). Increasing the doses of *oskar* to four copies showed increased levels of *nanos* RNA and *staufen* protein and a concommitant increase of pole cells to 40-60 compared with 10-15 when one copy of *oskar* is present (Ephrussi and Lehmann, 1992). The *oskar-bicoid3'UTR* transgene was also useful in determining which of the posterior group of genes are essential for abdominal development and which are involved in posterior localisation. The genes *cappuccino*, *spire*, *staufen* and *valois* were dispensible for abdominal development while *nanos*, *vasa* and *tudor* were essential.

Further work has recently emerged on the control of *oskar* expression with transgenes used to show that signals which control *oskar* localisation and translational repression reside in the 3'UTR (Rongo et al, 1995). The *oskar* protein is not detected until stage 8 egg chambers at the posterior pole even though the mRNA can be detected much earlier showing that translational repression takes place. Mutants that fail to localise the *oskar* mRNA to the posterior; *cappuccino*, *spire*, *mago nashi* and *staufen*, show no *oskar* protein, whereas protein is detected in *tudor*, *vasa*, and *valois* although at reduced levels in the latter two mutants.

Crosslinking with uv light was used to detect proteins from ovary extracts which bound to oskar mRNA (Kim-Ha et al, 1995). An 80kDa protein named Bruno was identified and shown to bind specifically to 3 sites within the oskar 3'UTR. These sites were called Bruno response elements (BREs). A transgene of *oskar* lacking the BREs was introduced into an *oskar*$^-$ line and shown to rescue the abdominal defects but to also produce anterior defects. This result showed that Bruno mediated translational repression of *oskar* in the ovary. Introduction of the *oskarBRE* transgene into *cappuccino*, *spire*, *mago nashi* and *orb* mutants showed the same effect as with *oskar*$^-$; rescue of abdominal defects but creation of anterior

defects. However, in a *staufen* mutant this transgene continued to show posterior defects showing *staufen* is necessary for *oskar* translation (Kim-Ha et al, 1995).

Webster et al (1994) placed the *Drosophila virilis oskar* gene, *virosk* as a transgene into *D melanogaster* to examine the effect of the *cis*-acting localisation signals. The *virosk* gene was localised correctly to the posterior although a novel intermediate was observed; the 'dot', a small spherical shaped signal that proceeded through the oocyte. Although the *virosk* gene rescued body patterning defects in *oskar- D melanogaster* ovaries, it was impaired in the maintenance of the mRNA signal at the posterior and in pole cell formation. In a wild type background the *virosk* gene showed a dominant negative lethality, probably due to failure to maintain localisation at the posterior and thus producing ectopic anterior signalling (Webster et al, 1994).

Analysis of the *oskar* transcript sequence has revealed two potential translational initiation sites referred to as M1 and M130 by Markussen et al (1995). These initiation sites can give rise to 69kDA or 54kDa proteins which are referred to as LongOskar and ShortOskar respectively. Both forms were detected by Western blots using ovary extracts with an additional band seen at 57kDa which is likely to correspond to a post-translationally modified form of the 54kda *oskar* protein. Mutation of both translational initiation sites showed that the ShortOskar was sufficient to rescue the *oskar* mutant phenotype. The LongOskar was tightly localised to the posterior pole but did not rescue pole plasm mutations. It would be interesting to speculate that the LongOskar represents an early translational product required to initiate tight localisation of *oskar* mRNA to the posterior pole while the ShortOskar is produced later and is the more dominant version.

The *vasa* protein is localised to the perinuclear region of the nurse cells during early oogenesis, then at the pole plasm at the oocyte posterior later in oogenesis. Localisation to the posterior is dependent on *cappuccino*, *spire*, *oskar* and *staufen* but the perinuclear position is not dependent on these genes, although it is abolished in *vasa* mutants. These observations suggest two independent roles for *vasa* (Liang et al, 1994). The *vasa* gene encodes a DEAD box protein with RNA helicase activity. Missense mutants which affect the RNA helicase activity localise to the posterior but cause defects in later processes showing that the RNA interaction is essential. Mutations in many other parts of the gene cause loss of

localisation both at the posterior pole and in the perinuclear region in the nurse cells (Liang et al, 1994).

The *pipsqueak* gene is required for abdominal segmentation and pole cell formation. These defects would appear to be due to a lack of properly localised *vasa* protein in *pipsqueak* mutants (Siegel et al, 1993). The *pipsqueak* gene encodes a protein containing a BTB box at the amino-terminus, a feature often associated with zinc fingers, although *pipsqueak* protein does not appear to possess zinc fingers (Horowitz et al, 1996). The protein contains 34 histidine-lysine tandem repeats raising the possibility that these repeats co-ordinate a metal ion and facilitate protein-protein interactions. There are also four conserved domains at the carboxyl terminus with no known homologies which have been named the *pipsqueak* motif. Antibodies show a nuclear location for the *pipsqueak* protein and, curiously, two spots on the oocyte nucleus during stages 6-10 suggesting the regulation of a specific loci.

Pole cells can be formed at any site within the embryo if the products of *oskar*, *vasa* and *tudor* are present. The *tudor* gene encodes a very large protein of 2516 amino acids which appears to be novel (Golumbeski et al, 1991). Antibodies to recombinant *tudor* protein recognise three different proteins; a 285kDa species in ovarian extracts, a 135kDa species in 0-2 hour embryos and a 205kDa species in 0-4 hour embryos. The *tudor* mRNA is localised to the oocyte in the germarium, similar to that of *oskar*. After stage 8 high levels of transcript are seen in the nurse cells. The protein distribution shows nurse cell perinuclear localisation and is also present in the oocyte prior to vitellogenesis. Localisation in the oocyte is dynamic with a transient anterior position from stages 4-6 followed by concentration to the posterior pole from stages 9-10 (Bardsley et al, 1993).

The *mago nashi* gene was cloned using Restriction Fragment Length Polymorphisms (RFLPs) in combination with a chromosomal walk (Newmark and Boswell, 1994). It encodes a small 147 amino acid protein with no known similarities. In *mago nashi* mutants the polar granules are absent or severely affected.

The *germ-cell-less* gene is required specifically for pole cell formation and encodes a novel 569 amino acid protein (Jongens et al, 1992). Expression of an antisense transgene caused a reduction or loss of pole cells. The *germ-cell-less* mRNA is

localised to the pole plasm and shows dependence on all the posterior group genes required for *nanos* localisation. A transgene resulting in increased expression of *germ-cell-less* produces additional pole cells while localisation to the anterior causes ectopic pole cell precursor formation (Jongens et al, 1994). The protein is located on nuclear pores suggesting that it may have a role in limiting access to or from the nucleus.

1.3.7.4 The Terminal System

The terminal system determines the development of the unsegmented ends of the embryo, the acron at the anterior and the telson at the posterior. Evidence shows that the terminal system is controlled by a different mechanism to that of the anterior-posterior axis (reviewed by St Johnston and Nüsslein-Volhard, 1992). Transplantation of cytoplasm can be used to rescue several terminal system mutations as was the case for the anterior-posterior system but there are notable differences. The site of injection of the rescuing cytoplasm is immaterial as is the position in the donor oocyte from which it was derived. These results indicate that the terminal system does not produce a localised cytoplasmic determinant as has been shown for the anterior-posterior axis. In addition, at least one gene in the female soma is required for determining terminal structures (Stevens et al, 1990).

The *torso* gene has been identified as functioning in determining the terminal system. The gene product of *torso* has a putative membrane spanning domain, a cytoplasmic domain showing homology to tyrosine kinases while the N-terminal region contains a signal peptide, thus, the *torso* product can be classified as a member of the receptor tyrosine kinase family (Sprenger et al, 1989). The *torso* gene is transcribed during oogenesis but is not translated until after fertilisation whereupon it localises throughout the zygotic membrane.

Other genes in the terminal system have been ordered by the use of *torso* gain-of-function mutants which probably produce constitutively active torso receptor. Thus, mutant genes upstream of the gain-of-function *torso* should have no effect while those downstream will show an effect. Genes upstream of *torso* have been shown to be *torso-like*, *trunk*, *fs(1) pole hole* and *fs(1) Nasrat*. A downstream gene identified by this method was *l(I) pole hole* which has been shown to encode the *Drosophila* homologue (D-raf) to the mammalian proto-oncogene *c-raf*, a

serine-threonine kinase involved in signal transduction pathways (Ambrosio et al, 1989).

The zygotic targets of this pathway are the genes *huckebein* and *tailless* but the transcription factor that controls these genes has yet to be identified (Weigel et al. 1990). The current model for the terminal system proposes that there is an interaction between the somatic follicle cells and the oocyte. Of the upstream genes, *torso-like* has been shown to be required in the soma while the other genes have been shown to be germ-line dependent. At this stage it is not known how the ligand for the *torso* product is formed. Two possibilities exist; *torso-like* produces an inactive ligand acted on by the upstream germ-line gene products to produce an active ligand after fertilisation, or perhaps the ligand is the product of one of the germ-line genes and it is activated by the *torso-like* product. The *torso-like* cDNA has been cloned recently (Savant-Bhonsale and Montell, 1993). It encodes a 353 amino acid protein with an N-terminal signal sequence indicating that the protein is secreted. Ectopic expression of *torso-like* produced embryonic phenotypes similar to those seen with gain-of-function *torso* mutants, but ectopic terminal structures were not seen. These authors propose that the torso-*like* product may interact with the *trunk* product to produce the ligand and that *trunk* product may be limiting explaining the lack of ectopic terminal structures when *torso-like* was mis-expressed. Further support for the ligand being the product of the *trunk* gene comes from its comparison with *spätzle* protein (Cassanova et al, 1995). The *trunk* protein contains a domain similar to the proteolytic sites in component 3 of the complement pathway and also contains a cysteine residues in a similar arrangement to the *spätzle* protein. These "cysteine knots" are regions susceptible to proteolysis and are found in a number of signalling molecules.

Perrimon (1993) has reviewed recent developments in understanding this pathway. Activation of D-raf involves $p21^{ras}$, a guanine nucleotide binding protein and the product of the *Ras1* gene, and *Son-of-sevenless* (*Sos*) whose product is the guanine nucleotide releasing/exchange factor. Mutations in *Ras1* or *Sos* can suppress *torso* gain-of-function alleles (Doyle and Bishop, 1993). A target for D-raf activity has been found in the form of the *Dsor1* mutation where a gain-of-function allele bypasses the need for *torso* and D-raf activity. The wild type *Dsor1* product appears to be the *Drosophila* homologue of the tyrosine/threonine kinase Mek. The Mek kinase's target may be a serine/threonine kinase of the MAP class, a mammalian MAP kinase has been shown to translocate to the nucleus and activate

transcription factors such as Jun and Elk-1. One further gene which has been recently identified and appears to be involved in this system is *corkscrew* which codes for a putative protein tyrosine phosphatase (Perkins et al, 1992). A mammalian protein tyrosine phosphatase homologue has been shown to bind to activated receptor tyrosine kinases and to become phosphorylated on tyrosine residues.

One curious aspect of the terminal signal transduction pathway is how do the same set of activated zygotic gap genes, *tailless* and *huckebein*, produce differing structures at the termini, the acron and telson. Cytoplasm transplantations indicate that these gap genes interact with components of the anterior-posterior axis (Sugiyama and Okada, 1990). Wild type cytoplasm from the anterior pole was able to produce anterior structures in a *torso*$^-$ mutant while wild type posterior cytoplasm restored posterior structures in the *torso*$^-$ mutant. This ability was lost in *bicoid* and *nanos* mutants, where terminal structures from the opposite pole were formed indicating a requirement for the correct anterior-posterior determinant for proper terminal structure formation.

A number of signal transduction pathways have now been identified which use a receptor tyrosine kinase and a kinase cascade to transduce their signal - another being the *sevenless* system which is involved in differentiation of the R7 photoreceptor cell in the *Drosophila* eye. These pathways show a high degree of conservation throughout evolution (Perrimon, 1993).

1.3.7.5 The Dorsal-Ventral System

The dorsal-ventral pattern is established by three sequential signalling pathways as defined by the phenotypes of mutations observed to affect this pathway (reviewed in Morisato and Anderson, 1995). The three stages are; 1) maternal effect that disrupt the polarity of the embryo and show defects in the eggshell, 2) those that produce a normal eggshell but show embryonic defects and 3) zygotic mutations that disrupt the dorsal-ventral pattern at more specific parts of the embryo.

The role of *gurken* in establishing the first stage of dorsal-ventral signalling has been discussed earlier (see section 1.3.6). The second stage of signalling involves the dorsal group of genes which will be discussed below. The third stage of dorsal-ventral patterning occurs in the embryo as a result of the primary patterning

set down by the first two stages. As this stage occurs in the embryo, it will only be mentioned very briefly.

The second stage of dorsal-ventral patterning has features similar to the terminal system in that a membrane spanning receptor protein is central to a signal transduction pathway. The transmembrane receptor protein is believed to be the product of the *Toll* gene due to it possessing domains homologous with other receptor proteins such as the human platelet glycoprotein 1b and the interleukin-1 receptor (Hashimoto et al [1988], Schneider et al {1991]). The *Toll* gene product is present throughout the embryonic membrane by the syncytial blastoderm stage (Hashimoto et al, 1991). Phenotypes of *Toll* range from dorsalised embryos resulting from loss-of-function mutants to ventralised embryos resulting from gain-of-function mutants which are probably constitutively active (Schneider et al, 1991). As with the terminal system, the gain of function mutants have enabled other genes in the dorsal-ventral group to be positioned upstream or downstream of *Toll*. Genes which function upstream of *Toll* are *gastrulation defective*, *pipe*, *nudel*, *windbeutel*, *snake*, *easter* and *spätzle*, while those that function downstream include *tube*, *pelle* and *dorsal*. The *cactus* gene is also part of the dorsal-ventral group and its function will be discussed later.

The product of the *dorsal* gene is believed to be the transcription factor that controls dorsal-ventral polarity in the embryo. Several lines of evidence support this conclusion; the rescuing effect of transplanted cytoplasm, the interaction with the *cactus* gene product, and only a *+/dorsal* heterozygote shows a dominant effect, producing a partially ventralised embryo indicating that higher levels of *dorsal* product are required for proper determination of ventral regions (see St Johnston and Nüsslein-Volhard, 1992). The *Toll* signal is transmitted from the ventral side of the embryo and regulates the nuclear localisation of the *dorsal* product. *dorsal* RNA and protein are synthesised during oogenesis and evenly distributed throughout the oocyte cytoplasm. The distribution of the *dorsal* protein changes at about the ninth nuclear division in the zygote (Steward, 1989). In ventral positions *dorsal* protein associates with the nuclei, in lateral positions it is evenly distributed between the nuclei and cytoplasm, while in dorsal positions it is found only in the cytoplasm. During mitosis the protein levels become uniformly distributed as the nuclear membranes breakdown. Localisation of the *dorsal* product occurs by selective nuclear targeting.

The *dorsal* gene product shows homology to subunit of the mammalian nuclear transcription factor NF-κB which is also regulated via a nuclear localisation signal (Ghosh et al, 1990). NF-κB is inactive in the cytoplasm as p65 and p50 subunits bound to the 1κB protein, but upon receiving a signal the 1κB protein is released and the p65 subunit translocates to the nucleus (Nolan et al, 1991). The *cactus* gene product shows homology to 1κB, with loss-of-function *cactus* mutations exhibiting a ventralised phenotype consistent with the hypothesis of nuclear targeting of *dorsal* protein (Roth et al, 1991). However, other factors must also be involved as the *cactus* mutants are not the strongest ventralised phenotypes that can be found. Govind et al (1993) postulate that the *Toll* mediated ventral signal modifies both *dorsal* and *cactus* products, perhaps by phosphorylation, to produce maximal nuclear localisation of *dorsal* product in ventral regions. Thus, the spatially regulated ventral signal may be required to act directly on *dorsal* protein for nuclear localisation. Later results from Belvin et al (1995) show that the *cactus* protein is rapidly degraded in response to the *Toll* mediated signal. The *cactus* protein contains PEST sequences, regions which are susceptible to proteolytic attack (Rogers et al, 1986). Injection of *in vitro* synthesised mRNA to produce *cactus* protein lacking the PEST sequences resulted in retention of more *cactus* protein than in the wild type. However, the PEST sequences are not required for signal-dependent degradation leading Belvin et al (1995) to propose that *cactus* is more stable when attached to *dorsal* protein and that it undergoes both signal dependent and signal independent rapid degradation in the embryo.

Both *tube* and *pelle* have been cloned and sequenced (Letsou et al [1991], Shelton and Wasserman [1993]). The *tube* gene encodes a novel protein of 50kd for which no direct function has been found while *pelle* encodes a 501 amino acid protein which shows a protein kinase domain similar to raf and mos, mammalian kinases with serine/threonine specificity. The *pelle* product may phosphorylate the *dorsal* and/or *cactus* products as suggested by Govind et al (1993).

The role of the *easter* gene has been further defined by Chasen et al (1992), who show that the *easter* product, a serine protease, is in the form of an inactive zymogen which is evenly distributed in the perivitelline space. Therefore, *easter* product is regulated post-translationally to produce active protease restricted to ventral regions. Of the 7 genes upstream of *Toll*, only *spätzle* was bypassed by a dominant active mutant of *easter* in double mutants suggesting that the *spätzle* product may produce the ligand for the Toll receptor when acted upon by the *easter*

serine protease. Morisato and Anderson (1993) have shown that a proteolytically processed form of *spätzle* product is lacking in dorsal group mutants upstream of *spätzle*, but present in those downstream of *spätzle*.

Further evidence that *spätzle* provides the ligand for the Toll receptor comes from purification of a 24kDa form of the protein which is only recognised by antibodies to the carboxyl terminal region (Schneider et al, 1994). This purified protein was able to produce polarising activity on injection into embryos. A deletion mutant in the *spätzle* gene which leaves only 106 amino acids from the carboxyl terminal region was also shown to activate Toll but a deletion leaving 168 amino acids was not (Morisato and Anderson, 1994). This observation suggests that intrinsic properties of the *spätzle* protein prevent it activating the Toll receptor and that these parts of the protein must be removed by proteolytic cleavage.

Epistatic experiments show that the *snake* gene is immediately upstream of *easter* and that *gastrulation defective* is upstream of *snake* (Chasan et al [1992], Roth [1994]). The *snake* protein appears to be secreted as an inactive zymogen. Mosaic studies have demonstrated that the genes *pipe*, *nudel* and *windbeutel* are required in the soma (Stein et al, 1991). The *nudel* gene has been cloned and sequenced with the deduced protein sequence showing similarity to extracellular matrix proteins but also containing a serine protease region (Hong and Hashimoto, 1995). This result suggests that the *nudel* protein resides on the surface of the follicle cells and is somehow activated by the *pipe* and *windbeutel* gene products. The *nudel* protein then initiates a proteolytic cascade which leads to the spatially restricted activation of the *spätzle* protein.

The zygotic targets for the dorsal protein are the genes *zerknüllt* (*zen*), *decapentaplegic* (*dpp*) and *tolloid* (reviewed by Morisato and Anderson, 1995) which act in dorsal regions, and *twist* and *snail* which act ventrally. In the wild type, the exclusion of *dorsal* protein from the nucleus allows expression of *zerknüllt*, *decapentaplegic* and *tolloid*, while in the ventral regions localisation of *dorsal* protein is required for the expression of *twist* and *snail*. In lateral regions the even distribution of *dorsal* protein between the cytoplasm and nucleus prevents expression of any of the above genes. Thus, the *dorsal* protein must act in a similar way to *bicoid* protein in that a threshold concentration has the ability to activate a gene. Binding sites for *dorsal* protein have been found upstream of the

zerknüllt and *twist* genes which suggests that the *dorsal* protein acts as both transcriptional repressor and activator

1.3.8 Localisation of Transcripts

Oogenesis requires the targeted localisation of many proteins. While it is possible for proteins to diffuse to the place they are required, there is increasing evidence for targeting cytoplasmic proteins by the localisation of their mRNAs (reviewed by St Johnston, 1995).

As discussed earlier, the *bicoid* mRNA is localised to the anterior margin of the oocyte implying the existence of a mechanism which causes this localisation. Several mutant genes have been identified that result in the mis-localisation of *bicoid* mRNA. These are *exuperantia* (*exu*), *swallow* (*swa*) and *staufen*. The *staufen* gene product has also been shown to be involved in the localisation of the posterior determinant.

Several steps have been observed in the localisation of *bicoid* mRNA. First, it localises to the apical region of the nurse cells, then it is transported to the cortex of the oocyte at the anterior pole where it is anchored (St Johnston et al, 1989). Mutations in *exuperantia* disrupt the nurse cell localisation phase while *swallow* mutants cause *bicoid* mRNA to dissociate from the cortex. The *swallow* mRNA shows a localisation pattern similar to *bicoid* mRNA during stages 5 to 6 of oogenesis, but it is not clear whether or not *swallow* is involved in *bicoid* localisation at that time (Hedgé and Stephenson, 1993). The *swallow* gene product encodes a putative RNA binding protein and a helical domain that may be involved in protein-protein interaction (Hedgé and Stephenson, 1993) Analysis of the 3' UTR of *bicoid* mRNA has shown a 53nt element (BLE1) that determines the early steps of RNA localisation, while deletions in other regions of the 3' UTR impair later steps of localisation (Macdonald et al, 1993). Therefore, it is probably that localisation of *bicoid* mRNA involves the step-wise use of a number of elements.

The green fluorescent protein (GFP) from the jellyfish *Acquorea victoria* is proving to be a useful tool for the study of many biological systems. The GFP gene has been coupled to the *exuperantia* gene and used as a transgene to study the localisation patterns of *exuperatia* protein (Wang and Hazelrigg, 1994). This study showed that the *exuperantia* protein was part of a ribonucleoprotein complex which

contains *bicoid* mRNA and is targeted to the oocyte anterior probably via microtubule attachment. These authors note that *exuperantia* protein can bind RNA non-specifically in vitro and do not believe that it binds *bicoid* RNA directly.

Transport to the posterior pole is more complex as the molecules must translocate across the oocyte. The earliest members of the posterior group of genes to show localisation are *staufen* protein and *oskar* mRNA (St Johnston et al, 1991). However, the interaction of these two molecules is complex as *oskar* localisation is abolished in a *staufen* mutant but posterior stabilisation of *staufen* protein requires *oskar*, probably as protein as in-frame stop codon mutations that affect the *oskar* protein but not the mRNA exhibit this effect. A similar experiment to that with the 3' UTR of *bicoid* RNA was performed with the 3' UTR of *oskar* mRNA showing that localisation involves several elements acting in a step-wise combinatorial fashion (Kim-Ha et al, 1993).

One further posterior group mRNA that shows an interesting pattern of localisation is that of the *Hsp83* gene (Ding et al, 1993b). The mRNA from this gene is localised to the posterior pole by a combination of RNA degradation and protection - RNA not at the posterior pole is selectively degraded while that at the posterior pole is protected. This gene product is also found at the anterior pole of the zygote where expression is dependent on *bicoid* and the other genes that operate in the anterior regions (Ding et al, 1993b).

Interestingly, not all mRNA localisation is essential as shown with the *fs(1)K10* gene by Serano and Cohen (1995). The *fs(1)K10* mRNA is normally found localised to the anterior cortex of the oocyte during early stages of oogenesis and is required for establishing the dorsal-ventral polarity in the egg chamber (see section 1.3.6). Mutations in *cappuccino* and *spire* that affect the *fs(1)K10* mRNA localisation continue to show normal synthesis and localisation of *fs(1)K10* protein. Transgenes were constructed to produce either a uniformly distributed transcript throughout the oocyte or posterior localisation of transcript. In both cases functional and localised *fs(1)K10* protein was produced.

1.4 Aims of this Thesis

It can be seen from the preceding discussion on the molecular genetics of oogenesis that some areas are particularly well understood, for example, the genes used in oogenesis to initiate the establishment of the embryonic axes. However, there is much more that is poorly understood, for example: what processes determine the oocyte, what are the interactions that occur between the germ-line cells and the follicle cells, what are the interactions between the subtypes of follicle cells, and how does the sex determination pathway influence oogenesis? Thus, there are many aspects of oogenesis which are open to further investigation.

As the oocyte develops it enlarges considerably as a result deposition of many products synthesised in the nurse cells and in the soma. These products from the nurse cells include proteins and mRNAs essential for oogenesis and later embryonic development. A complex targeting and transport system is required for the proper localisation of these molecules. Genetic screens for mutations have allowed many genes involved in these processes to be identified, cloned, and investigated in detail, for example, *gurken*, *oskar*, and *vasa*. However, not all genes involved in oogenesis are amenable to this approach, hence the introduction of the P-element enhancer trapping system. This system allows screening for genes involved in oogenesis on the basis of reporter gene expression. Of particular interest are patterns that show temporal and/or spatially restricted expression indicating that this gene may have a specific and critical function in oogenesis.

This laboratory has been using a P-element based reverse genetics approach to investigate genes involved in oogenesis. Genes with patterns that suggest an important role are investigated further. Occasionally, an expression pattern is observed by *in situ* hybridisation that shows no relationship to the original P-element reporter system, but is considered interesting nonetheless. This Thesis presents a study of a previously uncharacterised gene encoding a class V myosin which was identified from such an observation. In addition, a second gene encoding a putative ion-dependent inorganic phosphate co-transporter with functions in oogenesis was isolated and fully sequenced.

Chapter 2

Cloning of an Unconventional Myosin

2.1 Results Obtained Prior to the Commencement of This Work

A project was initiated to clone genes involved in oogenesis using a reverse genetics approach. Several hundred lines containing P-element reporter insertions P[*lacZ;rosy*] were analysed by dissection of the ovary and stained in a solution of X-Gal for β-galactosidase activity (M Bownes, personal communication, results not shown). Several lines with staining patterns in the ovaries were selected for further study. These lines included two which showed staining in a subset of follicle cells at the anterior and posterior of the egg chamber, one early in oogenesis and the second at a later stage. Another line demonstrated staining only in the germarium with molecular analysis showing that this expression pattern corresponded to the histone H1 gene (Walker and Bownes, personal communication).

A further line named H14 shows an ovary specific β-galactosidase staining pattern in the follicle cells from about stage 7. In later stages, the staining becomes localised to follicle cells at the anterior and posterior of the egg chamber. The border cells, a group of follicle cells which migrate through the nurse cells to lie at the anterior of the oocyte, also show staining. By stage 11-12 the staining is localised to follicle cells at the poles of the oocyte (M Bownes, personal communication).

In order to locate the gene producing this expression pattern, cloning experiments were initiated (R Slee, personal communication). As these enhancer trap lines contained an early version of the P-element reporter transposon, this cloning required construction of a bacteriophage genomic library with recombinant genomic DNA flanking the P-element isolated using a *lacZ* probe. This isolated genomic DNA was used as a probe to obtain cDNAs from a whole body adult library in the bacteriophage NM1149 vector. Three cDNAs were obtained with insert sizes of approximately 1kb, 1.5kb and 1.7kb, and named NMC1, NMC4 and NMC7 respectively. All three cDNAs were used to perform *in situ* hybridisation on whole mount ovarieswith those of NMC1 and NMC4 showing a high degree of staining in the nurse cells. However, the NMC7 cDNA revealed a staining pattern quite different to the original β-galactosidase pattern but highly interesting nonetheless. A temporally and spatially specific pattern showing localisation of a transcript early in the germarium in the presumptive oocyte was observed, see figure 2.1. During stages 3-7 of oogenesis the transcript remains localised in the oocyte. By stage 9 significant amounts of transcript can be seen in the

nurse cells while in the oocyte the transcript is seen as a band running dorso-ventrally at the anterior. Localised transcripts have been shown to be important for establishing the polarity of the oocyte, for example, *bicoid*, *oskar*, *gurken* and *orb* (see Chapter 1). Thus, this gene, in having a localised transcript, warranted further investigation.

A Northern blot using NMC7 cDNA as probe suggested a transcript in the ovaries of approximately 3.5-4kb. (R Slee, personal communication).

The inserted DNA from NMC7 recombinant phage was subcloned as two fragments, both of approximately 800bp, owing to an *Eco*RI site within the phage insert as shown in figure 2.2. The two recombinant plasmids, pNMC7a and pNMC7b, were subsequently sequenced and a deduced partial open reading frame of 984bp and an untranslated 3' region of 661bp (3'UTR) were observed.

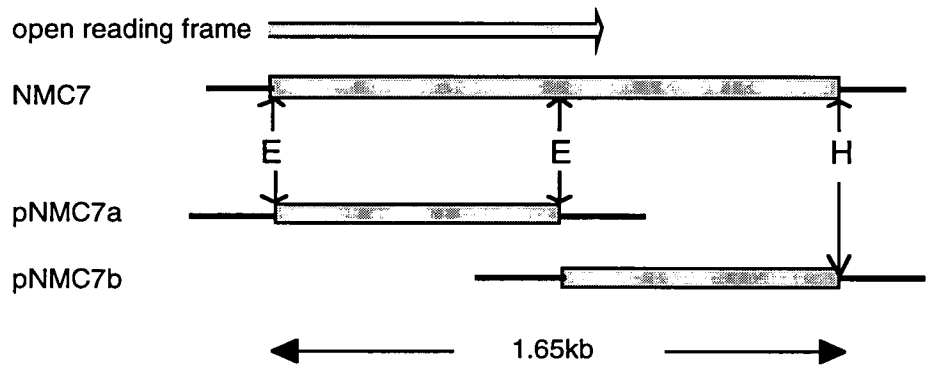


Figure 2.2 Schematic representation of the recombinant bacteriophage NMC7 and the two subcloned plasmids, pNMC7a and pNMC7b, derived from it. The open arrow represents the predicted open reading frame. Restriction endonucleases; E: *Eco*RI, H: *Hin*dIII

The deduced amino acid sequence was compared to the GenEMBL database using the GCG TFASTA program (Devereux et al, 1984). Significant homology was observed to an L-glutamate decarboxylase, the cDNA for which was reported by Huang et al (1990). An alignment of the protein sequences is shown in figure 2.3.

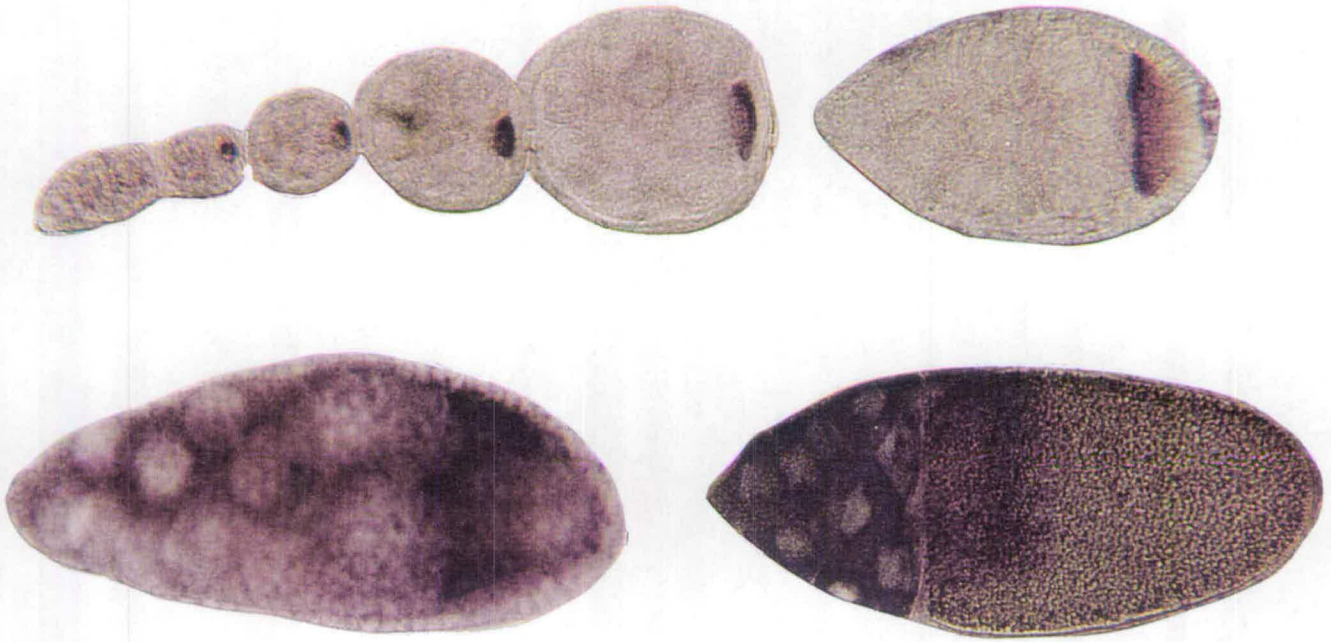


Figure 2.1 *in situ* hybridisation to egg chambers showing that the transcript is localised to the oocyte from very early stages of oogenesis becoming an anterior band at mid stages (6-8). Later stages (9-10) show high levels of transcription with signal apparent in the nurse cells and loss of the anterior banding pattern. The transcript is also detectable in early embryos (not shown). This result was kindly provided by A McCormack.

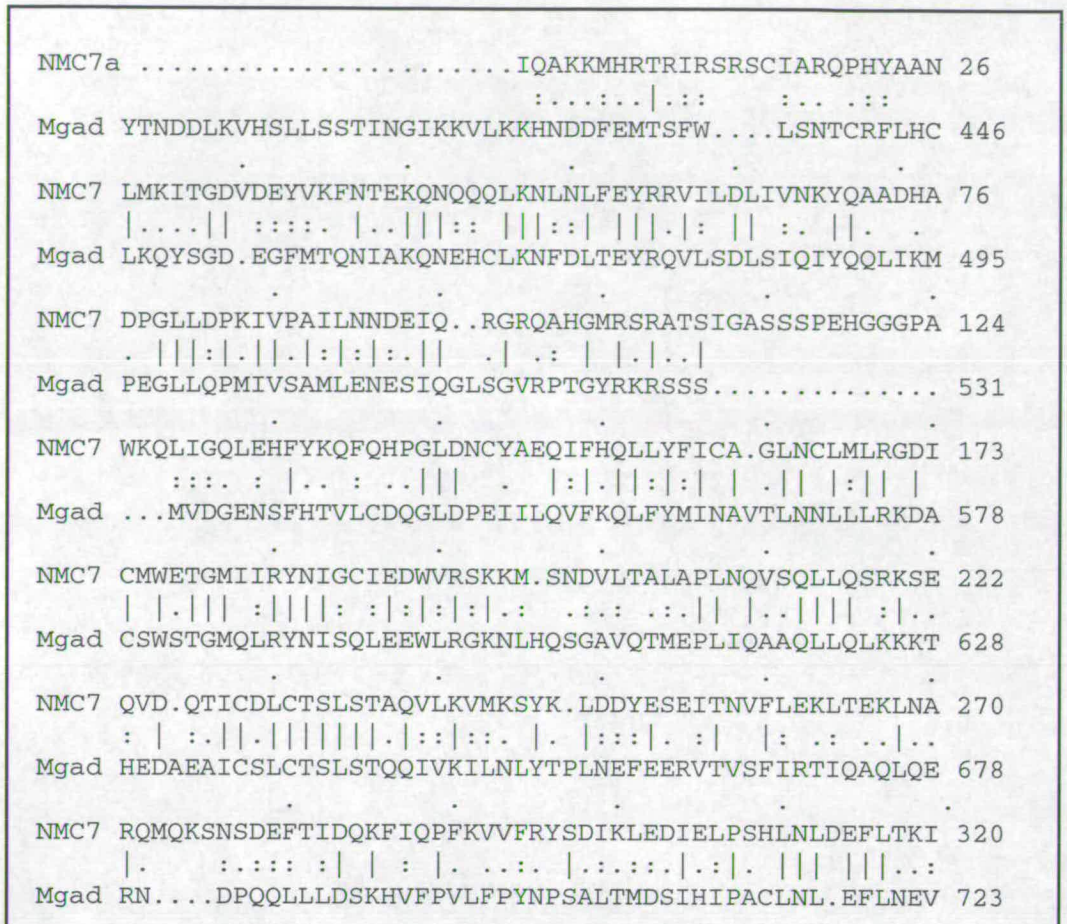


Figure 2.3 Comparison of mouse GAD protein (Mgad) with deduced protein sequence from *Drosophila* cDNA NMC7 using GCG GAP routine. Amino acid idendity is 36% and similarity 58%.

2.2 Work From This Project

All standard materials and methods used in this work are given in Chapter 7 and referred to where relevant. Specific materials and/or methods are described in detail within the appropriate results Chapter.

2.2.1 Isolation of Further cDNAs

The transcript size had been estimated as 3.5 - 4 kb from a Northern blot (R Slee, personal communication). Therefore, an early goal of the project was to isolate further cDNAs representing more of the transcript.

An ovarian cDNA library which had originated from the laboratory of Jan (Jongens et al, 1992) was kindly provided by D Zhao (personal communication). This library was screened using the insert from pNMC7a, an 847bp *Eco*RI fragment that represented 282 amino acids from the open reading frame.

2.2.2 λZAP Library Screen.

Approximately 5 x 10^4 recombinant bacteriophage were plated on a 25cm x 25cm Nalgene plate using standard phage plating techniques (Sambrook et al, 1989) and *E coli* XL1-Blue plating cells (Stratagene). Duplicate lifts to Hybond-N membranes were taken and processed as described in the materials and methods. The membranes were prehybridised at 65°C and probed with the pNMC7a insert which had been labelled with α-^{32}P dCTP as described in 7.5.5.2.

The primary screen yielded 25 positive recombinants in duplicate of which 10 were initially selected for further work. Subsequent screening resulted in 5 of these initial 10 recombinants remaining positive and being plasmid rescued.(7.2.3). An *Eco*RI digestion indicated that four of these 5 plasmids contained an insert of 2.2kb while the fifth was smaller at 1.4kb. These plasmids have been named ZAP-A, -D, -F, -G (2.2kb insert) and ZAP-E (1.4kb insert).

A restriction map of the recombinant plasmids containing the 2.2kb cDNA suggested that they were identical. Initial DNA sequence information derived from the ends of the inserts using universal primers showed that both the 1.4kb and 2.2kb cDNAs were

related to NMC7. This relationship is shown in figure 2.4 which shows that both λZAP derived cDNAs matched NMC7 from the *Eco*RI site within NMC7 and continued towards the 5' end of the gene, that is, neither λZAP derived cDNA contained the last 120bp of protein coding sequence nor the 3'UTR. Thus, combining the results from the NMC7 cDNA and the new λZAP cDNAs approximately 3.2kb of cDNA from this gene was represented.

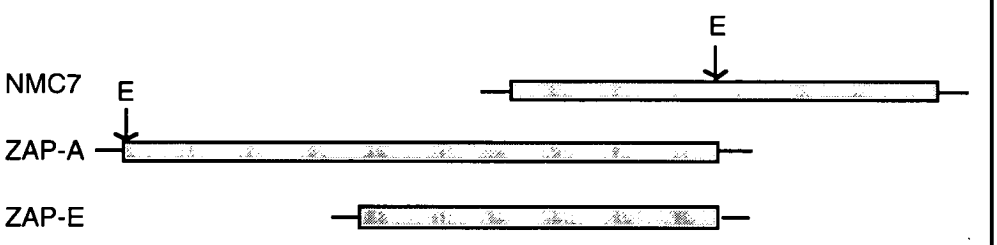


Figure 2.4 A schematic diagram showing the relationship between the original cDNA isolated, NMC7, and the new cDNAs, ZAP-A and ZAP-E, isolated from the λZAP library. Restriction endonuclease; E: *Eco*RI.

2.2.2.1 Sequencing Strategy

In order to determine the DNA sequence of the λZAP derived cDNAs a subcloning strategy was developed. The cDNA from plasmid pNMC7b was also sequenced to ensure that both strands at the 3' end were covered. In addition to this subcloning strategy, several synthetic oligonucleotide primers were obtained to fill gaps not covered by the recombinant plasmids.

The recombinant plasmids generated are listed in Table 2.1. Simple deletions were constructed by utilising a restriction endonuclease site in the vector's multiple cloning site and a similar site within the insert. Thus, by cutting the plasmid, heat denaturing the restriction endonuclease and re-circularising the remaining vector-insert DNA by ligation in a 50μl volume, a deletion construct was easily generated. For situations where the restriction endonuclease was not heat denaturable or additional restriction fragments were to be recovered, the DNA fragments were purified from agarose gels

(7.5.4.4). Recombinant subcloned plasmids were constructed by cleavage of the plasmid followed by purification of the required fragment from an agarose gel by Qiaex. This fragment was ligated into a pBluescript vector which had been cleaved with the appropriate restriction endonuclease and treated with CIAP to reduce occurrences of vector re-ligation events (7.5.2). Figure 2.5 shows a schematic representation of the subcloned recombinant plasmids. Sequencing was carried out using the Sequenase 2.0 system (7.7.1), then later using dye-labelled universal forward (custom made) and reverse primers and dye-terminators on an Applied Biosystems automated sequencer (7.7.2.1).

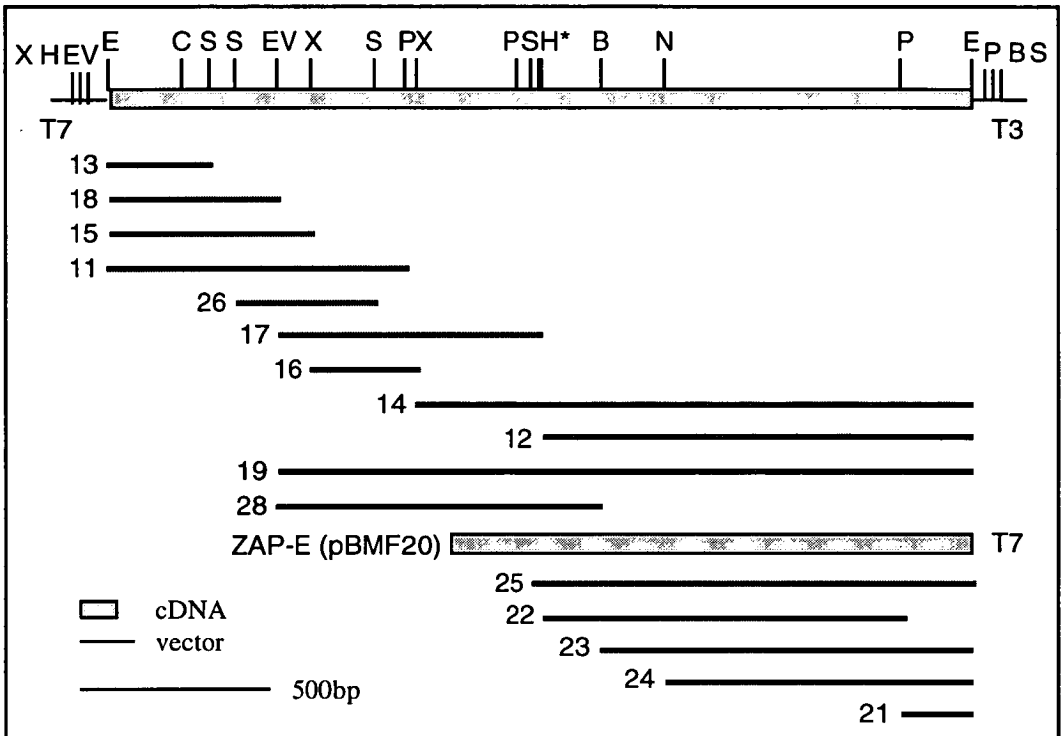


Figure 2.5 Schematic diagram showing restriction map of ZAP-A (pBMF10) as uppermost filled bar. The cDNA ZAP-E (pBMF20) is also represented by a filled bar while all recombinant plasmids subcloned from pBMF10 and pBMF20 for sequencing are represented as shaded lines with the numbers coreesponding to the pBMF series as shown in Table 2.1. Restriction endonuclease sites are: B: *Bgl*II, C: *Cla*I, E: *Eco*RI, EV: *Eco*RV, H: *Hin*dIII, N: *Nsi*I, P: *Pst*I, S: *Sac*I, X: *Xho*I. The *Hin*dIII site is marked with an asterisk (*) to indicate that there are two adjacent *Hin*dIII sites at this position. T7 and T3 represent the respective positions of the T7 and T3 primers with respect to the cDNA inserts.

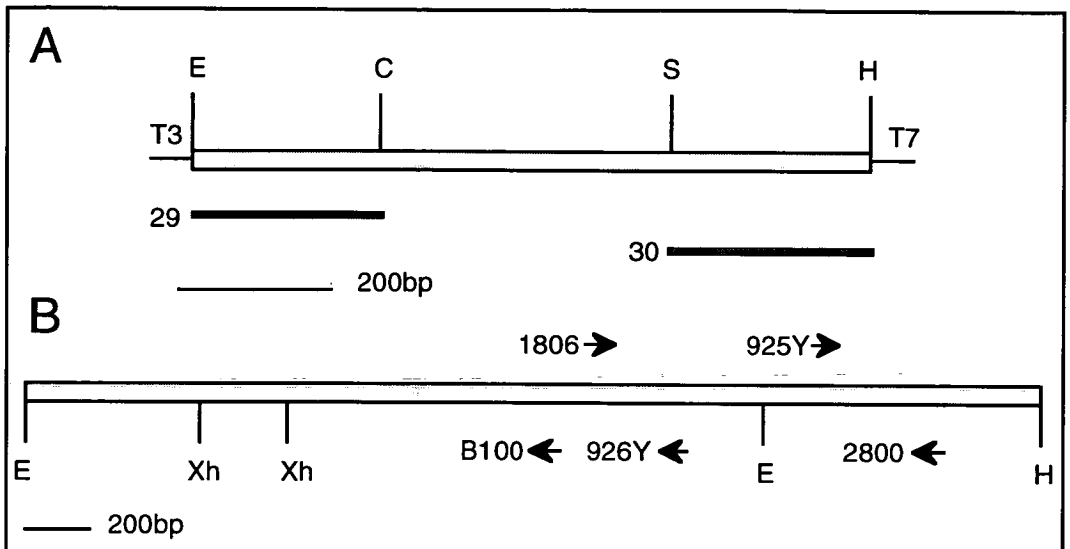


Figure 2.6 A) Schematic diagram of pNMC7b subcloning - see also Table 2.1. T3 and T7 represent the position of T3 and T7 sequencing primers relative to the cDNA insert. B) Primers used for sequencing with all 3.2kb of cDNA represented by the bar. "1806' and "2800" are the approximate distances, in bp, these primers are from the first nucleotide of the cDNA. The remaining primers were designed by R Slee for sequencing pNMC7a and pNMC7b. Restriction endonucleases; C: *Cla*I, E: *Eco*RI, H: *Hin*dIII, S:*Sac*I, Xh: *Xho*I.

Table 2.1 Plasmids Used to Determine the cDNA Sequence. Nucleotide (nt) positions are relative to pBMF10

Plasmid Name	Description	Vector	Primer(s)
pBMF 10	λZAP 2.2kb cDNA ZAP-A, -D, -F, -G	pBS SK-	T7, T3
pBMF 11	*Pst*I deletion of pBMF10, nt 1-799	pBS SK-	T3
pBMF 12	*Hin*dIII deletion of pBMF10, nt 1166-2324	pBS SK-	T7
pBMF 13	*Sac*I deletion of pBMF10 , nt 1-271	pBS SK-	T3
pBMF 14	*Xho*I deletion of pBMF10, nt 826-2324	pBS SK-	T7
pBMF 15	*Xho*I subclone of pBMF10, nt 1-541	pBS SK-	T3, T7
pBMF 16	*Xho*I subclone of pBMF10, nt 541-826	pBS SK-	T3, T7
pBMF 17	*Eco*RV-*Hin*dIII fragment subcloned from pBMF10, nt 457-1159	pBS SK+	T3, T7
pBMF 18	*Eco*RV-*Hin*dIII fragment subcloned from pBMF10, nt 1-457	pBS SK+	T3
pBMF 19	*Eco*RV deletion of pBMF10, nt 457-2324	pBS SK-	T7
pBMF 20	λZAP 1.4kb cDNA ZAP-E	pBS SK-	T3, T7
pBMF 21	*Pst*I deletion of pBMF20, nt 2132-2324	pBS SK-	T3
pBMF 22	*Hin*dIII-*Pst*I subclone of pBMF20, nt 1166-2132	pBS SK-	T3
pBMF 23	*Bgl*II-*Bam*HI deletion of pBMF20, nt 1328-2324	pBS SK-	T3
pBMF 24	*Nsi*I-*Sac*I deletion of pBMF20, nt 1497-2324	pBS SK-	T3
pBMF 25	*Sac*I deletion of pBMF20, nt 1138-2324	pBS SK-	T3
pBMF 26	*Sac*I fragment subcloned from pBMF10, nt 346-718	pBS SK-	T3, T7
pBMF 27	*Sac*I fragment subcloned from pBMF10, contains cloning artefacts	pBS SK-	T3, T7
pBMF 28	*Bgl*II-*Bam*HI deletion of pBMF19, nt 1-1328	pBS SK-	T3
pBMF 29	*Cla*I deletion of pNMC7b	pBS SK-	T7
pBMF 30	*Sac*I deletion of pNMC7b	pBS SK-	T3

2.2.2.2 Sequence Analysis Shows an Incomplete Open Reading Frame and Similarity to the Class V Unconventional Myosins

A database search with the predicted protein sequence from NMC7 revealed significant similarity to a mouse cDNA reported as coding for a L-glutamate decarboxylase (GAD) by Huang et al (1990) as described in section 2.1 above. This sequence will be referred to as the Huang sequence. Glutamate decarboxylases catalyse the conversion of glutamate to GABA (γ-aminobutyric acid), the primary inhibitory neurotransmitter. However, the Huang sequence shows no significant similarity to other glutamate decarboxylases described to date (see section 2.3) casting doubt on this sequence being an authentic glutamate decarboxylase. The 723 amino acid Huang sequence was used to search the database and revealed similarity to the carboxyl end of a family of proteins known as the *dilute* class unconventional myosins, or Class V unconventional myosins.

Structurally, the Class V myosins comprise a myosin head domain, a series of calmodulin binding repeats termed the IQ motifs, and a tail domain which consists of regions predicted to form coiled-coil α-helices and globular domains. At the time of the comparison, three examples of this class were reported; mouse *dilute* (Mercer et al, 1990), chicken p190 (Espreafico et al [1992], Sanders et al [1993]) and budding yeast *MYO2* (Johnston et al, 1991). A schematic representation of the 3 unconventional myosins and the mouse "GAD" protein is shown in figure 2.7

Initially, because of a frame shift at the 5' end of the 2.2kb cDNA, it was thought that the 3.2kb cDNA coded for a 674 amino acid protein lacking a myosin head domain, but with similarity to the tail domain of the Class V unconventional myosins. Further support for this theory came from the estimated size of 3.5-4kb for the transcript (R Slee, personal communication) and the existence of the Huang sequence with a predicted protein of 723 amino acids. However, further analysis of the sequence resolved the frame shift and showed that the predicted open reading frame could extend to the first nucleotide of the cDNA. On the basis of this result, it was decided to perform further Northern blots to verify the size of the transcript.

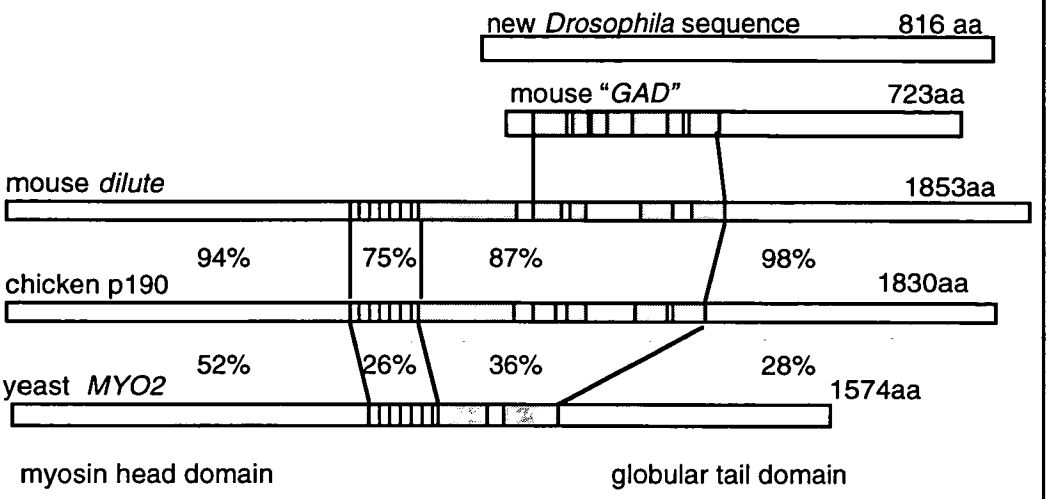


Figure 2.7 Block diagram representing class V unconventional myosins from mouse (Mercer et al, 1991), chicken (Espreafico et al [1992], Sanders et al [1992]) and yeast (Johnston et al, 1991) along with the mouse "GAD" (Huang et al, 1990) and the *Drosophila* protein sequence predicted from cDNA sequencing (see Appendix I). Boxes with vertical lines represent the 'IQ' motifs while shaded regions represent domains predicted to form coiled-coil α-helices. Adapted from Espreafico et al (1992). Percentages indicate the amino acid identity between each region in the myosin V sequences.

2.2.3 Northern Blot Shows A Large Transcript.

RNA was prepared using the Qiagen RNeasy system (7.10) from adult male and female Oregon R flies, from ovary tissue and the female carcass from the ovary dissection. This RNA was then run on an agarose/formaldehyde gel (7.11.2) and transferred to Hydond N. RNA markers from Gibco/BRL were included on the gel to allow sizing. A probe was prepared using the OLB (7.5.5.2) system from the 2.2kb cDNA. The membrane was prehybridised at 42°C for 4 hours and hybridised overnight with probe also at 42°C. Standard washes were carried out (7.11.4) and the membrane exposed to X-ray film. The membrane was then reprobed with labelled bacteriophage λ DNA which would detect the markers. The results shown in figure 2.8 shows a single transcript most evident in the ovary lane that was estimated to be 6kb in size. This result showed that the gene in question was producing a transcript considerably larger than the combined cDNAs obtained to date represented and that in all probability the gene coded for a full length unconventional myosin.

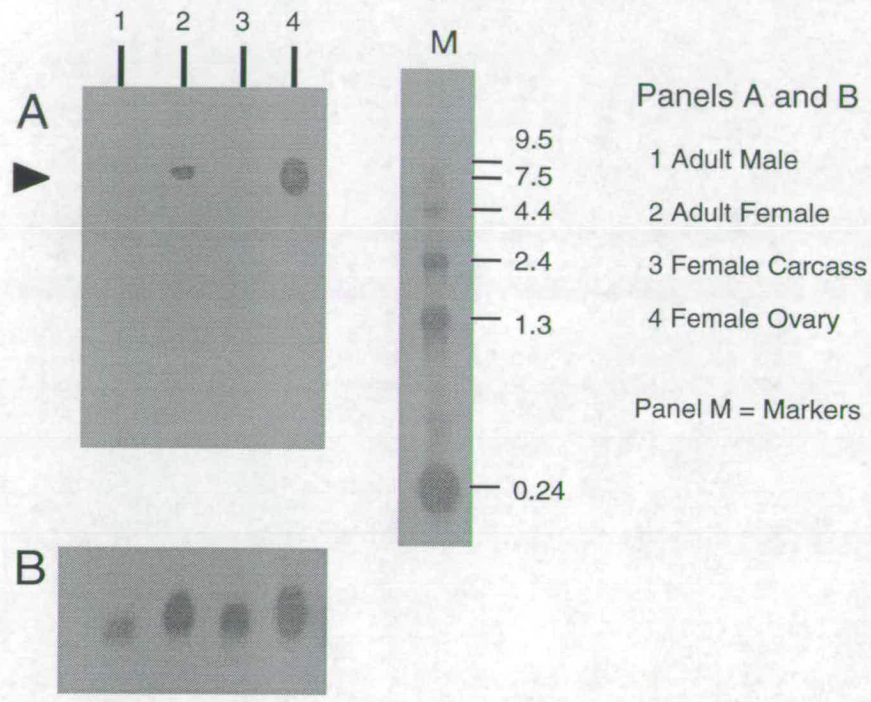


Figure 2.8 Northen Blot Panel A: probed with 2.2kb cDNA. Panel B is the same membrane as panel A probed with RP49 as loading control (transcript size of 0.6kb). Panel M: RNA markers (BRL) with size in kilobases indicated at side, probed with bacteriophage lambda DNA..

One question which remained unanswered at this point was whether the 5' *Eco*RI site in the current 2.2kb cDNA was part of the gene or part of the library cloning technique, that is, an *Eco*RI linker. It was apparent from the way in which both cDNAs from the ovarian λZAP library were truncated at the 3' *Eco*RI site (this site divides NMC7 into two approximately equal fragments) that there was a possibility that the library had been poorly constructed and that *Eco*RI sites internal to the cDNA molecules had not been adequately protected during library construction.

2.2.4 5'RACE Attempt

An attempt was made to obtain the missing 5' cDNA sequence by use of a PCR based technique called 5' RACE (Rapid Amplification of cDNA Ends). This method employs a gene specific primer to reverse transcribe the 5' sequence from the transcript. An anchor is then placed at the 5' end to allow PCR amplification. This anchor can be a specific primer ligated to the reverse transcribed single-stranded DNA molecule by T4 RNA ligase or a polynucleotide tail added by the enzyme terminal deoxynucleotide transferase (TdT). Figure 2.9 is a schematic representation of the RACE system. This technique is frequently used to obtain the last few hundred base pairs from a cDNA whereas in this case it was hoped to amplify up to 3kb of cDNA.

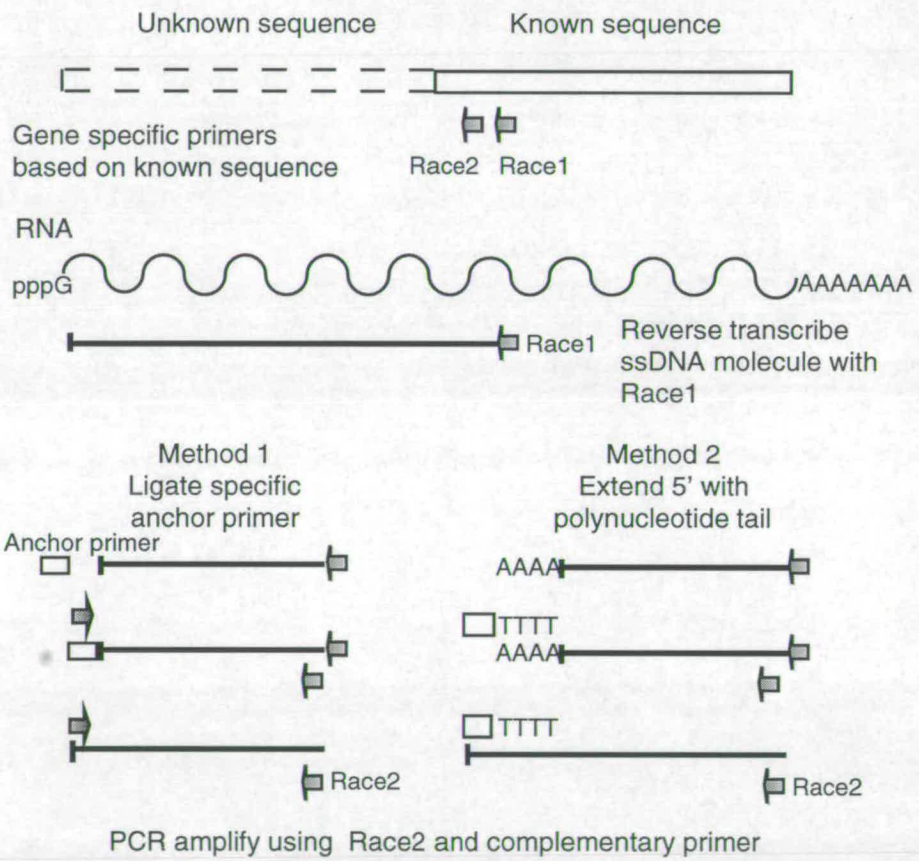


Figure 2.9 The RACE Procedure

The method did not produce the desired 5' end of the cDNA (results not shown) but was the source of 2 gene specific primers, Race1 and Race2, which were used extensively at later stages of the cloning. A 5'RACE procedure was used later to obtain the last few hundred base pairs of the cDNA sequence, see section 2.2.13.1. Figure 2.10 is a schematic representation of the position of primers Race1 and Race2 and additional primers used for sequencing of the cDNA that had been synthesised at this point in the project.

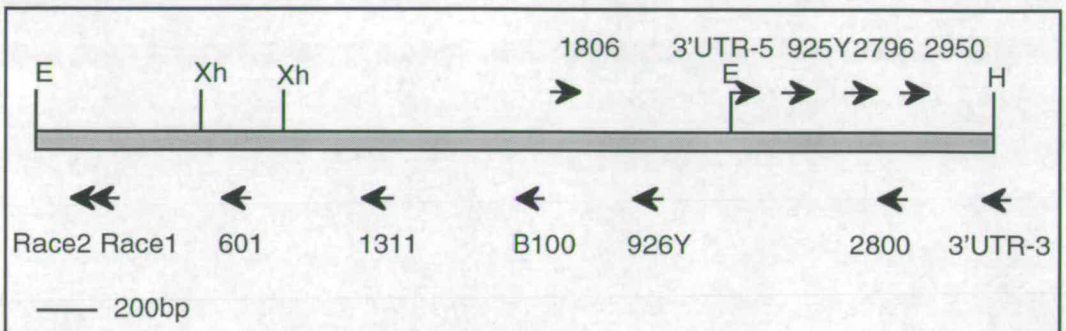


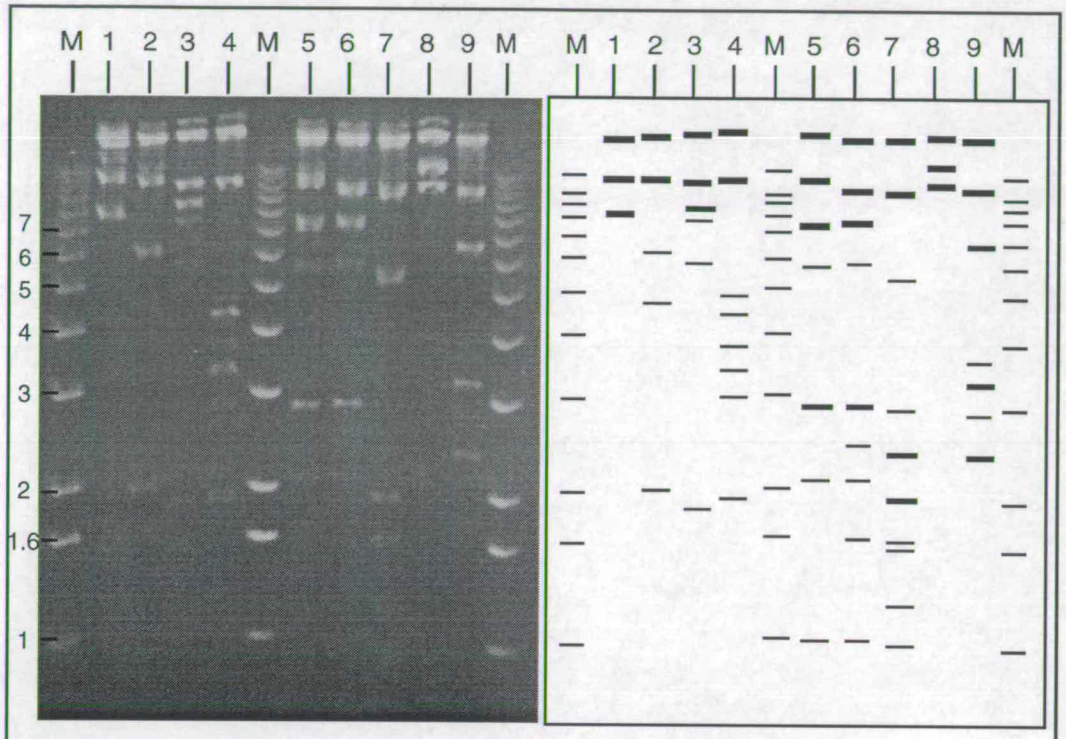
Figure 2.10 Schematic diagram of the cDNA showing positions of oligonucleotide primers. Primers Race1 and Race2 were designed for the 5'RACE procedure. Primers 3'UTR-5, 3'UTR-3, 2796 and 2950 were designed for PCR and sequencing of the 3'UTR (A McCormack, personal communication) but were also used for general sequencing. Restriction endonucleases; E: *Eco*RI, H: *Hin*dIII, Xh: *Xho*I.

2.2.5 Further Library Screens

A general strategy was adopted to screen all libraries in possession in the laboratory for the myosin cDNA. This screening included the genomic DNA library in the λFIX bacteriophage vector. The λFIX genomic library was screened in conjunction with a 3rd instar larval cDNA library by A McCormack (personal communication) using the 2.2kb cDNA as probe. No cDNAs were obtained from the larval cDNA library while a single recombinant phage was recovered from the λFIX genomic library. This recombinant phage is referred to as FIX-1A.

DNA was prepared from FIX-1A using the method described in 7.6.2.5 and restricted with the following restriction endonucleases or combinations of endonucleases; *Bam*HI, *Bam*HI/*Eco*RI, *Bam*HI/*Not*I *Bam*HI/*Xba*I, *Eco*RI, *Eco*RI/*Not*I, *Eco*RI/*Xba*I, *Not*I and *Xba*I. The restriction endonucleases *Not*I and *Xba*I have been engineered

into the λFIX vector and allow the inserted genomic DNA to be isolated. The restriction digests were separated on an 14cm x 14cm agarose gel and blotted to Hybond-N as described in the materials and methods (7.11.1). The result is shown in figure 2.11.



M	Markers	5	*Eco*RI
1	*Bam*HI	6	*Eco*RI + *Not*I
2	*Bam*HI + *Eco*RI	7	*Eco*RI + *Xba*I
3	*Bam*HI + *Not*I	8	*Not*I
4	*Bam*HI + *Xba*I	9	*Xba*I

Figure 2.11 Restriction digests of FIX-1A with a schematic representation of the gel shown on the right. Reproduction quality and aging has resulted in loss of visibilty of lower molecular weight bands on the photograph. Marker sizes are in kb.

It was found that an accurate restriction map could not be generated from the bands visible on the gel. Hybridisation with the 2.2kb cDNA fragment showed that the gene was located at the end of the cloned genomic DNA ligated to the 20kb left arm of the bacteriophage vector. An *Eco*RI digest showed hybridisation to a small ~1kb band in addition to the left arm plus insert band. These results are shown in figure 2.12.

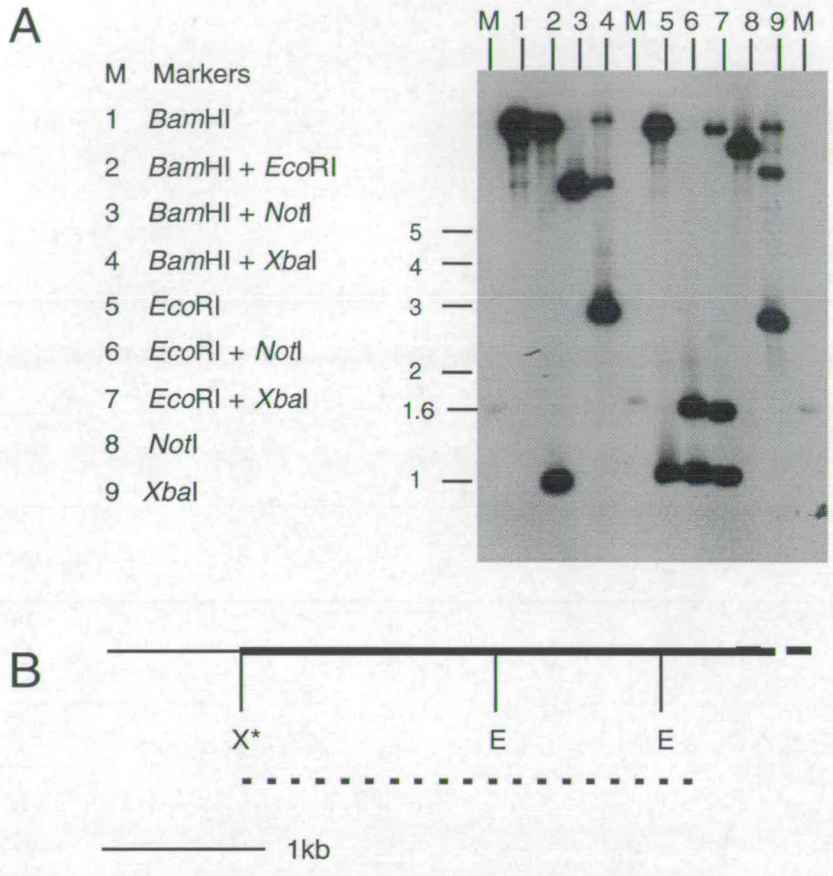


Figure 2.12 A) Hybridisation of restriction digests of FIX-1A with the 2.2kb myosin V cDNA fragment. Marker sizes are in kb. B) A schematic map of the genomic DNA determined from this hybridisation is shown. The heavy line represents inserted genomic DNA which extends to the right, the light line represents vector sequence, the dashed line represents the myosin V gene. Restriction endonucleases; E: *Eco*RI, X: *Xba*I. The *Xba*I site marked with an asterisk (*) is part of the vector cloning site and is not an *Xba*I site within the genome.

2.2.6 Reverse Northern

The principle of a Reverse Northern is to use RNA as a template for producing a probe and then probing DNA to discover which DNA fragments contain exons from the transcript of interest. Use of the Race1 primer allowed a probe that would represent the missing 5' sequence to be generated from total ovary RNA. The probe was prepared as follows:

200ng Race1 primer was annealed to ~10µg total ovary RNA prepared with an RNeasy kit (7.10) in a volume of 20µl by heating to 70°C for 5 minutes, then allowing to cool slowly. The RNA was reverse transcribed by the addition of 8µl Superscript II 5x buffer (BRL), 3µl 0.1M DTT, 500µM each of dATP, dGTP, dTTP, 5µl (50µCi) of α-^{32}P dCTP and water to a total of 39µl. The reaction mix was equilibrated to 42°C in a water bath for 2 minutes before 1µl of Surescript II (BRL) reverse transcriptase was added, then allowed to proceed for 30 minutes at 42°C followed by 5 minutes at 55°C. The unincorporated nucleotides were removed by use of a Pharmacia Nick column as described for general radiolabelling of DNA (7.5.5.2). The RNA-DNA hybrid molecule was denatured in a heating block at 110°C then added to the prehybridised membrane. Hybridisation was carried out at 65°C overnight.

The autoradiograph result is shown in figure 2.13 and shows hybridisation to the following bands; ~8kb *Bam*HI, ~7kb *Eco*RI, ~6kb *Xba*I and ~4kb *Bam*HI/*Xba*I. These bands were clearly different from those seen when hybridisation with the 2.2kb cDNA was used (see above) and suggested that isolation of any one of these large fragments would provide a probe to use for screening cDNA libraries, and in particular, the ovarian λZAP, for larger or 5' cDNAs.

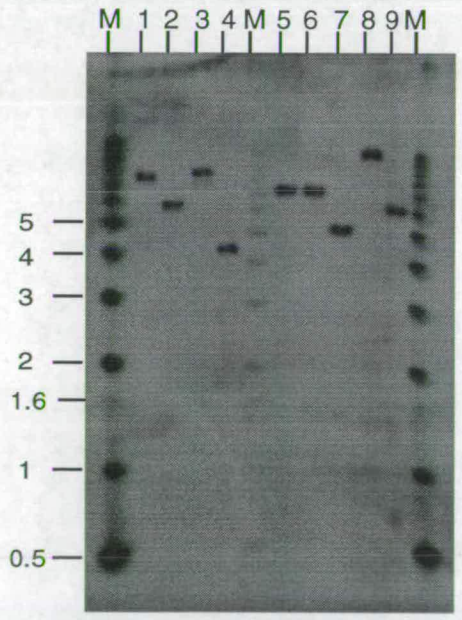


Figure 2.13 Reverse Northern. The restriction digests shown in figure 2.11 were transferred to a Hybond-N membrane and probed with first strand cDNA prepared by reverse transcription from RNA as described in the text. Markers in the outside lanes were labelled with radionucleotide (see 7.4.5.1) with sizes shown being in kb.

Attempts were made to recover all *Xba*I fragments from FIX-1A but this proved only partially successful. However, a small amount of the *Not*I DNA fragments were recovered which were then labelled and used to probe both the ovarian λZAP library and the 0-4 hour embryonic NB40 plasmid library (see below).

2.2.7 An Unrelated Gene, cDNA-34, is Identified

The ovarian library yielded one recombinant plasmid from 14 positives from the primary screen. Initial DNA sequencing using T7 and T3 primers did not produce any sequence identity with the current myosin cDNA. A comparison with the mouse *dilute* peptide sequence by the GCG TFASTA routine using mouse *dilute* as the query peptide and the new cDNA sequence as target produced no similarities. Thus, it was concluded that a cDNA from an unrelated gene had been isolated. How this result came about from the Reverse Northern is still not fully understood, but some theories will be discussed in Chapter 6. The new cDNA, referred to as cDNA-34, has been completely sequenced and analysed and the results are reported in Chapter 5.

2.2.8 Screening of the NB40 Embryonic Library

2.2.8.1 Conventional Screening

On the recommendation of D St Johnston (personal communication), the early (0 - 4 hour) embryonic plasmid library was obtained from N Brown (Brown and Kafatos, 1988). This library was recommended because, although embryonic, it had proved useful for obtaining full length transcripts of genes involved in oogenesis (eg *gurken*, D St Johnston, personal communication). It was also known that the transcript from the myosin V gene was present in the early embryo from Northern analysis (A McCormack, personal communication) The NB40 embryonic libraries have been constructed using the plasmid pNB40, a pUC derived plasmid which allows directional cloning of cDNAs resulting in an SP6 promoter at the 5' end of the cDNA and a T7 promoter at the 3' end.

Information supplied with the library suggested that non saturating amounts of DNA, about 5ng, should be used to transform highly competent cells. The library should be plated on 6 - 10 standard 90mm petri dishes at a density of 15 000 cfu per plate to give sufficient representation of transcripts. Colony lifts should be taken after 3-4 hours growth, before the colonies are visible.

Transformation of *E coli* XL1-Blue cells with 5ng of the library using the standard cold $CaCl_2$ method (7.3.1) produced too few transformants (approximately 15 000 cfu/ml) to reliably screen. Therefore, electroporation was used to transform the XL1-Blue cells to a higher efficiency (7.3.2). A control transformation using 1-2ng of pBluescript KS+ produced a transformation efficiency of 10^9 cfu/µg of DNA. while transformation with 4ng of the library gave a lower efficiency of 4×10^7 cfu/µg resulting in 160 000 cfu available for plating. The library was plated on 10 LB-agar + ampicillin plates at a density of approximately 10 000 cfu per plate, slightly lower than that recommended in the supplied protocol. The plates were incubated for 4 hours at 37°C with standard colony lifts taken on Hybond-N 82mm diameter membranes. The probe used was the *Not*I fragment representing the entire insert from the λFIX-1A cloned genomic DNA. This probe was being used to concurrently screen the ovarian λZAP and NB40 early embryonic libraries for further cDNAs from the putative myosin gene.

The initial signal obtained was weak and it was found to be impossible to identify the positive colonies on the primary plates because the density of colonies was too great. In order to improve the chance of obtaining a positive colony, a small area including the positive colony was scraped with a matchstick and transferred to 200µl LB-broth in a microfuge tube Dilutions were plated out to give plates that were quite dense, but with colonies still clearly separated and these were again screened. No clear positives were obtained

It was decided to rescreen the library by plating at a lower density in order to give clearly defined colonies which could be more readily isolated directly from the primary screen. The library was replated by again transforming *E coli* XL1 Blue cells by electroporation. The efficiency obtained was 1.5×10^7 cfu/µg for the library providing approximately 60 000 cfu which were plated on 20 LB-agar + ampicillin 90mm plates. Standard colony lifts were taken as before and hybridisation was with the 2.2kb cDNA. No positives were obtained. At this point it was decided to investigate the NB40 library using PCR as described below.

2.2.8.2 PCR on the NB40 Embryonic Library

The pNB40 vector contains an SP6 promoter site at the 5' end of the cDNA insert. Therefore, an SP6 primer in combination with a myosin specific primer was an attractive option for obtaining further cDNA information using a PCR approach. Firstly, it was prudent to determine if the NB40 library contained any cDNAs from the

myosin gene. The use of primers designed for sequencing of the existing cDNA showed that the library did contain cDNAs from the myosin V gene. These results are presented in figure 2.14.

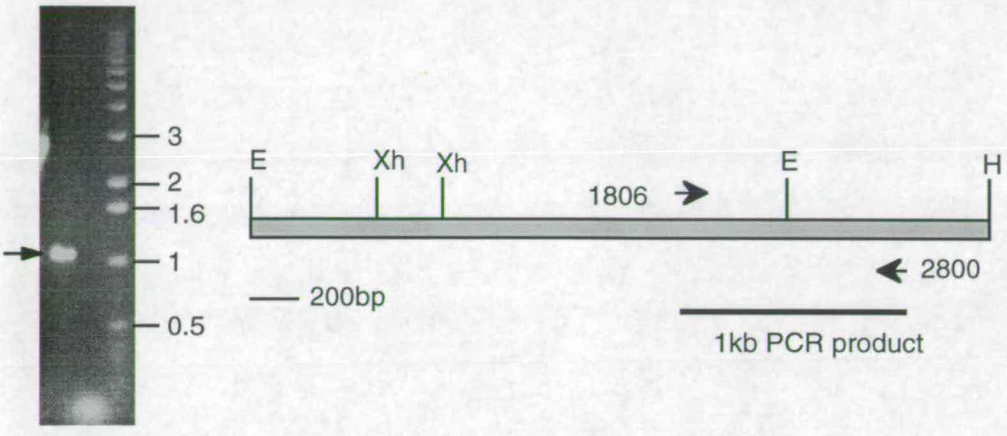


Figure 2.14 PCR with primers 1806 and 2800 showing that myosin V cDNAs are present in the NB40 library. Restriction endonucleases; E: *Eco*RI, H: *Hin*dIII, Xh: *Xho*I. Marker sizes are in kb.

PCR amplification with the myosin specific Race1 or Race2 primers in combination with the SP6 primer yielded inconclusive results. Amplification from the library produced bands of about 1.6kb. However, a control reaction using library template and the SP6 primer also produced a similarly sized band as shown in figure 2.14. The SP6 primer sequence had been copied from that published by Promega Inc to give an 18 nucleotide molecule. The predicted annealing temperature for this SP6 primer was calculated as 48°C using the method of Lathe (1985), that is, 2°C(A + T) + 4°C(G + C). Thus, the PCR was constrained by this relatively low T_m for the SP6 primer.

A new SP6 primer designated SP6/40 was designed with 4 additional bases complementary to the pNB40 vector and 5' to the consensus SP6 sequence to produce a primer with a predicted T_m of 64°C using the formula described above. A PCR was carried out using this new primer in combination with Race1, Race2 and the sequencing primer 601. The distance between the Race primers and primer 601 is approximately 400bp as shown schematically in figure 2.15. The results show that the SP6/40- Race1 or Race2 combinations produced bands of 1.6kb and 1.8kb while the SP6/40 - 601 combination produced bands of 2kb and 2.2kb, that is, 400bp larger (figure 2.15). This result was clear evidence that there were at least 2 cDNA species in this library that extended the current cDNAs by about 1.3kb and 1.5kb (allowing for

the position of the Race primers in the existing sequence and the distance the SP6 promoter is from the start of the cDNA).

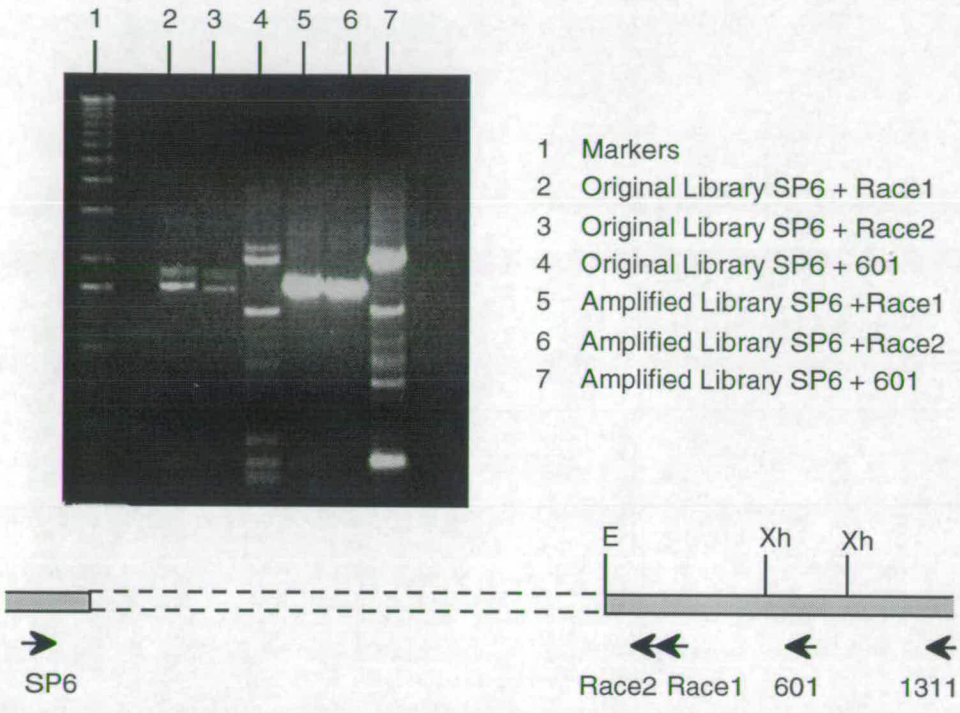


Figure 2.15 PCR result from the NB40 library. Lanes 2, 3 and 4 show 2 bands likely to be from the myosin V gene. The original library was the aliquot received from N Brown. The amplified library was the result of transforming *E coli*, culturing, then performing a plasmid preparation. However, the transformation was not efficient and the amplified library is less than fully representative.

2.2.8.3 Cloning and Sequencing of the PCR Products

Two strategies were used to obtain sequence information from these PCR products: direct sequencing and cloning into pBluescript. Direct sequencing of the PCR product was only successful on one occasion using the Race2 primer but the sequence matched that of the existing cDNA and proved that the *Eco*RI site at the 5' end of the 2.2kb cDNA from the λZAP library was part of the transcript.

The pBluescript SK- vector was prepared for cloning of the PCR products by initially cutting with the enzyme *Eco*RV to produce blunt ends followed by treatment with calf intestinal alkaline phosphatase to reduce vector re-ligation events. The phosphatase was removed by standard phenol/chloroform extraction followed by ethanol

precipitation. To obtain flush ends suitable for ligation the PCR products were treated with an enzyme cocktail as follows:

DNA		20µl
T4 polynucleotide kinase buffer		5µl
2mM ATP		5µl
10mM dNTPs		1µl
T4 DNA polymerase (NEB)		1 unit
T4 polynucleotide kinase	(NEB)	1 unit
Klenow (NBL)		1 unit
Sterile distilled water		to 50µl
37°C for 30 minutes		

The PCR products were separated on 0.8% agarose gels, the bands excised and the DNA purified with Qiaex (7.5.4.4). A ligation reaction was prepared and used to transform *E coli* XL1-Blue cells using the cold $CaCl_2$ method. The cells were plated on LB-agar + ampicillin supplemented with X-Gal and IPTG for blue-white selection purposes. Several ligation-transformation attempts yielded only one colony containing an insert. This recombinant plasmid, pNBpcr1, was used only for sequencing purposes with the emphasis shifting to use of cloned genomic DNA and RT-PCR methods for determining the remaining coding sequence as described later.

2.2.9 Summary of Library Screening

As stated above, all libraries in the laboratory were screened for presence of the myosin V cDNA. Results for the λZAP and NB40 cDNA libraries, and λFIX genomic library have been presented. Screening of other libraries was unsuccessful. Table 2.2 lists all libraries screened and summarises the results.

Table 2.2 Summary of cDNA and Genomic Libraries Screened

Tissue	Vector	Result	Comments
Adult Whole body	λNM1149	1.6kb	The original cDNA isolated
Ovary[1]	λZAP	2.2kb and 1.4kb	*Eco*RI sites not protected, truncated cDNAs
Ovary[2]	λgt22	no positives	
Testes	λgt11	~20 primary positives	no positives on secondary screen
Embryonic	λgt11	no positives	
Embryonic[3]	NB40 (plasmid)	cDNAs present	isolated some 5' sequence by PCR
3rd instar larval		no positives	screened by A McCormack
Genomic DNA[4]	λFIX	6 positives	all lack 5' sequence, 2 appear identical

Notes 1 source D Zhao, Stratagene commercial library
2 source P Tolias (Stroumbakis et al, 1994)
3 source N Brown (Brown and Kafatos, 1988)
4 source Stratagene commercial library

2.2.10 Refinement of the FIX-1A Map

The physical map of FIX-1A could not be obtained from the restriction endonuclease digest information. Therefore, it was decided to subclone fragments to resolve the physical map of FIX-1A by more detailed restriction mapping and by hybridisation. This work was done in conjunction with A McCormack.

A 'shotgun' approach was used to subclone *Eco*RI, *Xba*I and *Eco*RI-*Not*I fragments from the cloned genomic DNA in FIX-1A. DNA prepared from FIX-1A was digested with the appropriate restriction endonucleases (*Not*I and *Eco*RI used an identical buffer system), then the endonucleases removed by phenol/chloroform extraction and the DNA recovered by ethanol precipitation before being ligated to a suitably prepared pBluescript vector (*Eco*RI, *Xba*I or *Eco*RI-*Not*I with CIAP treatment). These ligations were used to transform XL-1 Blue *E coli* by the cold $CaCl_2$ method (7.3.1). White colonies recovered from the transformation were analysed for plasmid content by small scale DNA preparations (7.6.2.1) and digestion with the appropriate restriction endonuclease (eg *Eco*RI for *Eco*RI subcloning, etc). This series of subcloning experiments yielded four distinct *Eco*RI and two distinct *Xba*I recombinant plasmids. No *Eco*RI-*Not*I recombinant plasmids were recovered. The recombinant plasmids obtained are listed in Table 2.3 with restriction maps shown in figure 2.16

Sequence information was obtained from each recombinant plasmid using T3 and T7 primer sites in the vector and compared to the myosin cDNA sequence. These comparisons allowed a partial map of the genomic DNA to be built up and revealed that the FIX-1A cloned genomic DNA contained sequences matching the cDNA from nucleotide 532 of the cDNA, that is, the cloned genomic DNA contained no further 5' sequence , see figure 2.17.

Table 2.3 Recombinant Plasmids Subcloned from FIX-1A

Plasmid Name	Restriction Endonuclease	Insert size	Comments
pE1	*Eco*RI	~1kb	
pE3	*Eco*RI	~3kb	cloned in both orientations
pE5	*Eco*RI	~5kb	cloned in both orientations
pE7	*Eco*RI	~7-8kb	cloned in both orientations
pX2.8	*Xba*I	~2.8kb	
pX3	*Xba*I	~3kb	cloned in both orientations

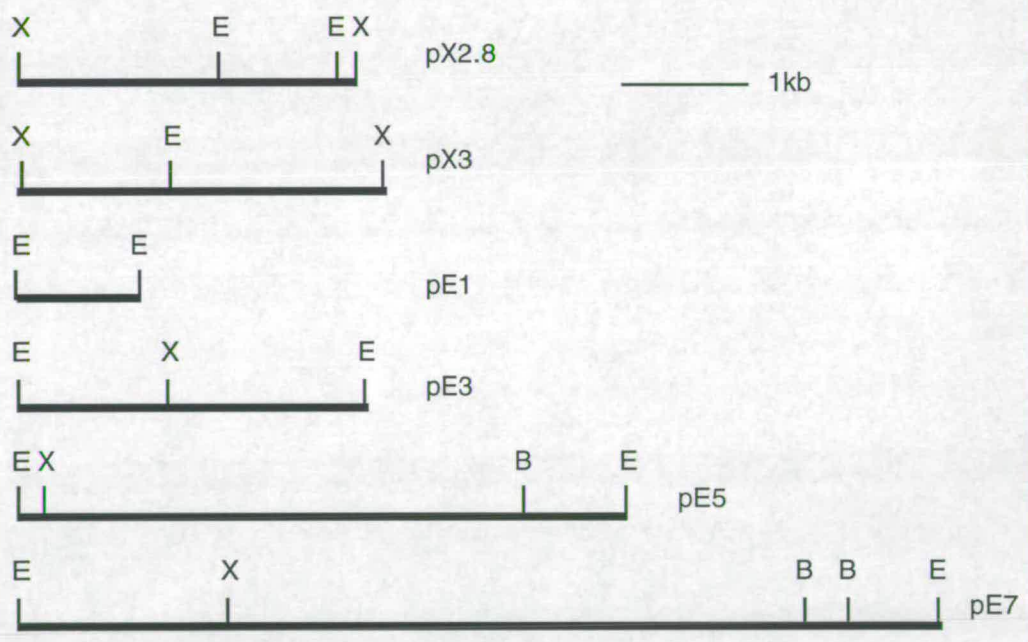


Figure 2.16 Restriction maps of recombinant plasmids obtained from subcloning FIX-1A. Restriction endonuclease; B: *Bam*HI, E: *Eco*RI, X: *Xba*I.

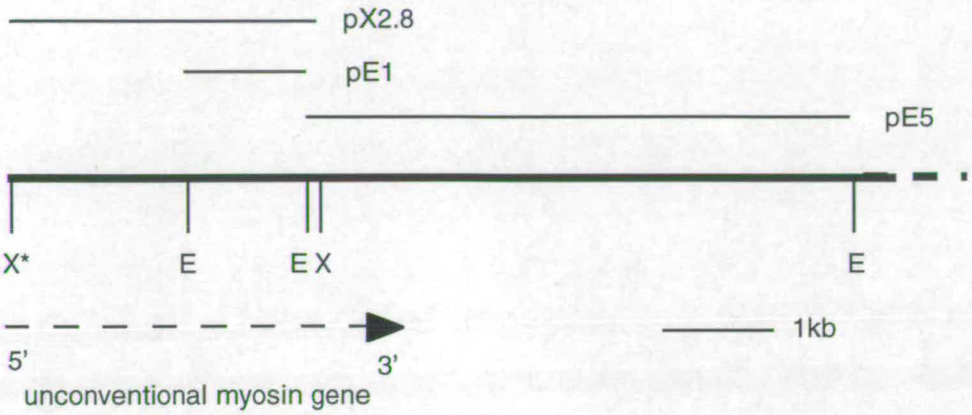


Figure 2.17 Genomic map of the myosin V gene determined from sequencing and restriction mapping of plasmids subcloned from FIX-1A. Restriction endonuclease; E: *Eco*RI, X: *Xba*I. The *Xba*I site marked with an asterisk (*) is part of the λFIX bacteriophage multiple cloning site.

2.2.11 Further Screening of the Genomic λFIX Library

The λFIX genomic library was screened again to isolate further phage that it was hoped would contain sequence 5' to that obtained from FIX-1A. The library was re-screened using the 2.2kb cDNA as probe as has been described previously. This screen yielded 2 positive recombinant phages which have been called FIX-2C and FIX-2F. DNA was prepared from these phage (7.6.2.5) and restricted with *Bam*HI, *Bam*HI/*Not*I, *Bam*HI/*Eco*RI, *Eco*RI *Eco*RI/*Not*I and *Xba*I, separated on a 0.8% agarose gel, transferred to Hybond-N+ and probed with the cDNA. The result is shown in figure 2.18 and it was apparent that FIX-2F contained a further ~1kb of genomic DNA while FIX-2C contained less genomic DNA spanning the putative unconventional myosin gene. A schematic representation of the relationship between the 3 recombinant phages is shown in figure 2.18. The λFIX library was screened once more but the probe used in this screen was a ~460bp *Eco*RV fragment from the 5' end of the cDNA (see figure 2.5) in order to improve chances of isolating phage which contained 5' sequence. A total of 3 phage were isolated and termed FIX-4A, FIX-4B and FIX-4D. Partial mapping has shown that FIX-4B is identical to FIX-2F while FIX-4A contains a further ~3-4kb of 5' DNA compared to FIX-1A, and FIX-4D breaks within the existing cDNA.

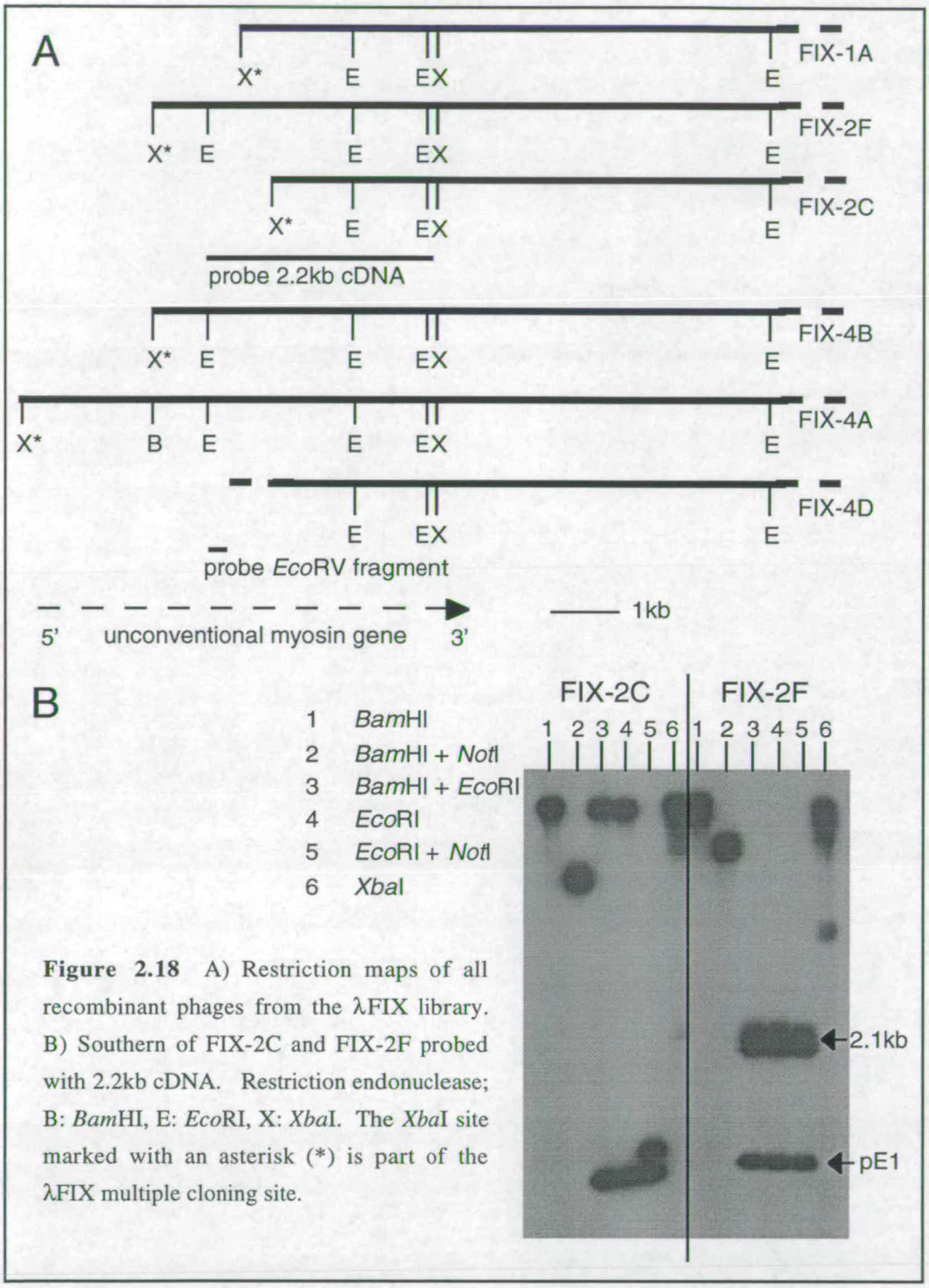


Figure 2.18 A) Restriction maps of all recombinant phages from the λFIX library. B) Southern of FIX-2C and FIX-2F probed with 2.2kb cDNA. Restriction endonuclease; B: *Bam*HI, E: *Eco*RI, X: *Xba*I. The *Xba*I site marked with an asterisk (*) is part of the λFIX multiple cloning site.

These phage from the 3rd screen have been superseded by genomic DNA subcloned from bacteriophage P1 recombinant.

2.2.12 Screening Cosmids and Recombinant P1s

The chromosome position had been determined as 43BC from in situ hybridisation to 3rd instar larvae salivary gland chromosomes as described in Chapter 3. This information was used to obtain recombinant cosmid and bacteriophage P1s containing cloned genomic DNA as described in Table 2.4 from *Drosophila* resource centres.

Table 2.4 Recombinant Cosmid and P1 Bacteriophage Screened

Name	source	Hybridisation position
33H2	Crete	43AC
140C2	Crete	43B
194A4	Crete	43B
P1 DS00126	Cambridge	43B
P1 DS00574	Cambridge	43C
P1 DS08719	Cambridge	43B

The cosmids were obtained prior to the recombinant P1s and were screened by a dot blot of a small scale DNA preparation (7.6.2.1 - kanamycin at 50µg/ml was used for culturing the *E coli*). A positive control of FIX-2F DNA was included. No hybridisation was obtained with the cosmid DNA (result not shown). A similar procedure was used to screen the recombinant P1s when they were obtained with P1 DS00574 showing a positive result.

The recombinant P1s contain a large amount of genomic DNA, in the order of 80-90kb (Hartl et al, 1994). It was decided to undertake a "shotgun" sub-cloning experiment and isolate subcloned recombinant plasmids that hybridised with the cDNA. A medium scale DNA preparation (Promega Wizard™, 7.6.2.3) yielded very low amounts of DNA.

Approximately 1µg of the P1 DS00574 DNA was restricted with *Eco*RI and *Xba*I in 50µl volumes with the restriction endonucleases removed by phenol/chloroform extraction. These digests were added to a ligation reaction containing pBluescript cut with the appropriate restriction endonuclease and treated with phosphatase (7.5.2). Half the ligation was used to transform *E coli* XL1 Blue (7.3.1) and colonies

containing inserts were selected on L-agar plates containing ampicillin, X-Gal and IPTG. All white colonies were picked with small scale DNA preparations made from all 15 *Xba*I subclones and 7 *Eco*RI subclones. These preparations were cut with the restriction endonucleases *Eco*RI or *Xba*I, to isolate the insert, separated on an 0.8% agarose/TBE gel, transferred to Hybond-N+ and probed with the ~460bp *Eco*RV fragment from the cDNA. A subcloned recombinant plasmid containing a ~7kb *Xba*I fragment was identified. This plasmid has been called pP1-X7

A small scale (Promega Wizard™, 7.6.2.3) DNA preparation was made from pP1-X7 and sequenced using automated sequencing procedures (7.7.2.2) with the T3 and T7 primers. The sequence from the T3 end matched that of the cDNA and genomic DNA sequenced to date from the *Xba*I site in the 3'UTR. The sequence from the T7 end was new and did not match any existing cDNA sequence. A comparison of the predicted amino acid sequence from the T7 DNA sequence showed strong similarity (.>44%) to the mouse *dilute* unconventional myosin from amino acids 10 to 170 in the 1st reading frame as shown in figure 2.19. Thus it was concluded that most of the remaining coding sequence for the *Drosophila* unconventional myosin lay within the subcloned 7kb *Xba*I fragment.

```
SCORES    Frame: (1) Init1:  528 Initn:  528 Opt:  588
          44.7% identity in 228 aa overlap

                              10        20        30        40
mdilmy                  MAASELYTKFARVWIPDPEEVWKSAELLKDYKPGDKVLLLHLEEGKD
                          ::| :: |::|:|::: ||:||:| ::|: |:  | :  ::||
x7t7.s SRGGSGITPIHKTCXFGSLLSQGAKIWVPHADLVWESATLEESYRKGAGFLKICTDSGKL
              10        20        30        40        50        60

       50        60        70        80        90       100
mdilmy LEYRLDPKTGELPHLRNPDILVGENDLTALSYLHEPAVLHNLRVRFIDSKLIYTYCGIVL
       | :|:::::::||:||||:||||:||||:|||||||:||||||||| ::::|||||||:|
x7t7.s KEVKLKADGSDLPPLRNPAILVGQNDLTTLSYLHEPGVLHNLRVRFCERQIIYTYCGIIL
              70        80        90       100       110       120

       110       120       130       140       150       160
mdilmy VAINPYEQLPIYGEDIINAYSGQNMGDMDPHIFAVAEEAYKQMARDERNQSIIVSGESGA
       ||||||:::|:|| :||:||:|::|||::|||||:|||||::::|:: | :||:||||||
x7t7.s VAINPYAEMPLYGPSIIRAYRGHAMGDLEPHIFALAEEAYTKLERENCNLNIIISGESGA
              130       140       150       160       170       180

       170       180       190       200       210       220
mdilmy GKTVSAKYAMRYFATVSGSASEANVEEKVLASNPIMESIGN-AKTTRNDNSSRFGKYIEI
       |::   : :|:|||     :  ::|: :     : :|:: : :|:: ::: :::||: ::
x7t7.s GQKSVRQIPMKYFARCWKFQIRNQVQPRCWHLPQSLETLRKCQKNPEXKQFPPLGKFYQT
              190       200       210       220       230       240

       230       240       250       260       270       280
mdilmy GFDKRYRIIGANMRTYLLEKSRVVFQAEEERNYHIFYQLCASAKLPEFKMLRLGNADSFH
       :| ::
x7t7.s AFPRNQNGFLIF
              250

SCORES     Frame: (2) Init1:    50 Initn:    50 Opt:    76
           32.8% identity in 67 aa overlap

            140       150       160       170       180       190
mdilmy GDMDPHIFAVAEEAYKQMARDERNQSIIVSGESGAGKTVSAKYAMRYFATVSGSASEANV
                                    |:: :||:|||||::: :::|::| ||::
x7t7.s VIWNRTSLPWRKKRTRNWSAKTATXTSSSVGNPVPGKKVSAKYPXSTLPAVGSSKSETKS
         150       160       170       180       190       200

            200       210       220       230       240       250
mdilmy EEKVLASNPIMESIGNAKTTRNDNSSRF-GKYIEIGFDKRYRIIGANMRTYLLEKSRVVF
       :  :  |    :: ||||:|:|:|:|:: :::::: |
x7t7.s NPGAGISPNHWKPSGNAKKTQNENNSPLXENFTKLLFPETKMGFXF
         210       220       230       240       250
```

Figure 2.19 TFASTA results comparing mouse *dilute* myosin V with open reading frames from pP1-X7 sequenced with the T7 primer. A significant amount of identity is seen in two frames as a result of a frame shift occurring after about 540nt. The ATP-binding domain is boxed in both frames.

2.2.13 Determining the Genomic and cDNA Sequence

The identification of genomic DNA which contained coding sequence near the 5' end of the gene enabled a reverse-transcription-PCR (RT-PCR) based sequencing strategy to be adopted in order to obtain most of the coding sequence. This strategy did not yield the most 5' sequence as this sequence extended beyond the pP1-X7 recombinant plasmid, so 2 new primers were designed for a 5'RACE procedure. The 3 primers initially synthesised were designated Race3 and Race4 for the 5'RACE procedure and did2 for the PCR sequencing procedure. The did2 primer was designed based on genomic sequence that, upon translation, showed amino acids identical to those from the mouse *dilute* peptide sequence. It was also decided that the sequence of pP1-X7 would be determined in full. At this time, the PCR fragment obtained from the NB40 library (section 2.2.8.3) had been subcloned into pBluescript and was also being sequenced.

The gene has been given the working name *didum*, an acronym from *dilute*-like *Drosophila* unconventional myosin. All primers designed specifically for sequencing have been given a 'did' notation.

2.2.13.1 RT-PCR and 5'RACE

Total ovary RNA was prepared using the Trizol method (7.10). Two reverse transcription reactions were prepared using approximately 5µg of RNA and either Race1 primer or Race3 primer. Material from the Race3 reaction was used to perform the 5'RACE procedure using a Gibco-BRL kit. The Race1 reaction was used to perform normal PCR in combination with the did2 primer. A product of approximately 400bp was obtained from the 5'RACE procedure as shown in figure 2.20. This product was re-amplified and sequenced directly after agarose gel purification. A primer, did7, was designed using the GCG PRIME program to match sequences as 5' as possible in the RACE product.

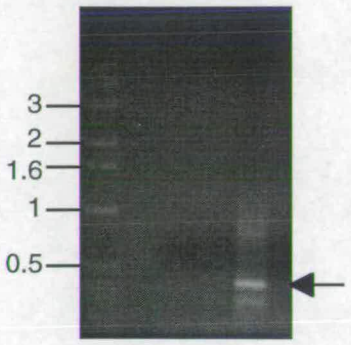


Figure 2.20 Result of 5'RACE PCR. A major 400bp band was obtained. Marker sizes are in kb.

2.2.13.2 PCR Sequencing

Primers were designed using the GCG PRIME program to obtain an optimal primer with constraints of 50% G/C content, T_m at least 50°C, and 3' base either G or C. Primer length was set at 18 nucleotides. Initial sequence obtained from the Race1-did2 PCR product and the NB40 cloned PCR product allowed further primers to be designed. Successive rounds of sequencing and primer design was used to determine the cDNA sequence generated by RT-PCR. These primers were also used to determine the genomic DNA sequence from recombinant pP1-X7.

2.2.13.3 Subcloning pP1-X7

Approximately 2.8kb of sequence from pP1-X7 had been obtained from the FIX-1A recombinant plasmids pE1, pE5 and pX2.8, (see also figure 2.17) A restriction map of pP1-X7 showed that a number of smaller fragments were obtained with the restriction endonucleases *Eco*RV and *Xho*I. The known sequence contained mapped *Eco*RV and *Xho*I sites and enabled an approximate map of the remaining sites in pP1-X7 to be determined. The *Eco*RV digest yielded 5 fragments of 0.4kb, 0.5kb, 0.9kb, 1.0kb and 1.2kb in addition to the fragment remaining with the vector. These 5 smaller fragments were successfully subcloned into pBluescript digested with *Eco*RV and treated with CIAP using identical methods described previously. A concurrent attempt to subclone the 0.5kb, 1.3kb (doublet) and 1.5kb *Xho*I fragments was not successful possibly to nuclease contamination resulting in damaged restriction endonuclease 'sticky ends' preventing ligation.

Figure 2.21 shows a schematic representation of pP1-X7 with subcloned fragments indicated along with primers used for sequencing using RT-PCR.

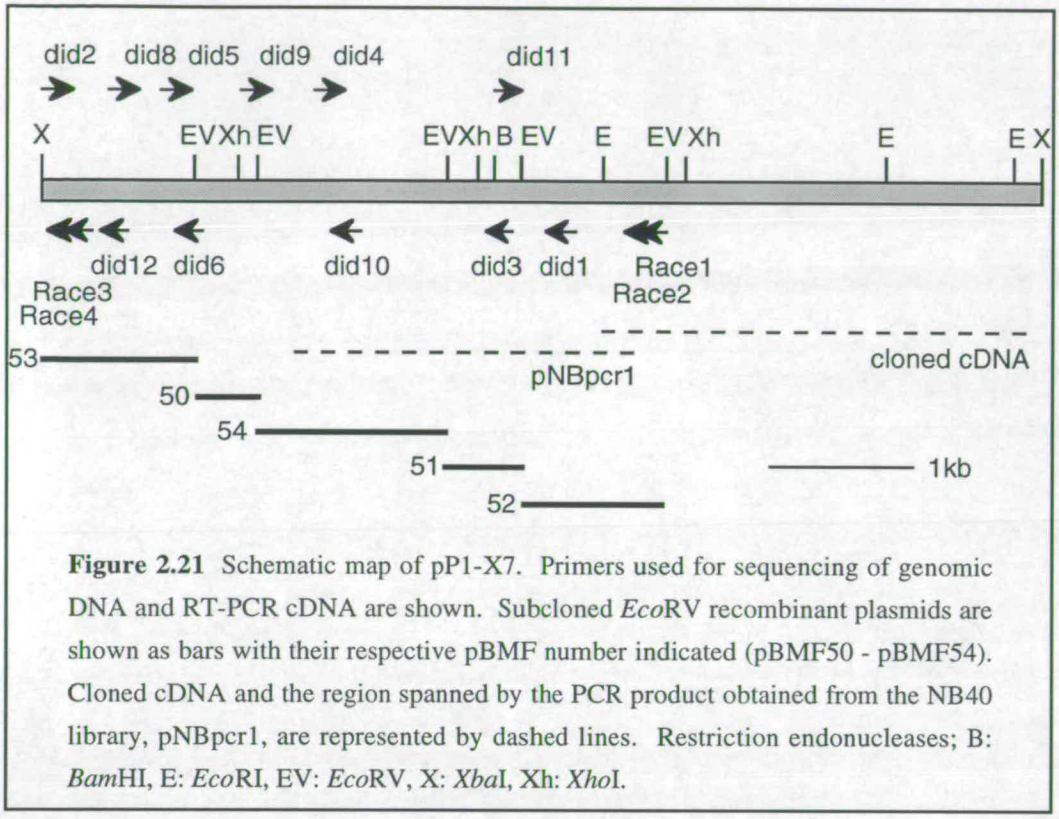


Figure 2.21 Schematic map of pP1-X7. Primers used for sequencing of genomic DNA and RT-PCR cDNA are shown. Subcloned *Eco*RV recombinant plasmids are shown as bars with their respective pBMF number indicated (pBMF50 - pBMF54). Cloned cDNA and the region spanned by the PCR product obtained from the NB40 library, pNBpcr1, are represented by dashed lines. Restriction endonucleases; B: *Bam*HI, E: *Eco*RI, EV: *Eco*RV, X: *Xba*I, Xh: *Xho*I.

2.2.14 The RACE Product may not be the 5' End

The 400bp PCR product obtained from the 5'RACE procedure was sequenced using the Race4 primer by direct PCR sequencing. The sequence obtained allowed a new primer, did7, to be designed as near as possible to the 5' end using the PRIME program. This primer has been used in conjunction with did6 for RT-PCR using single stranded template primed from both did6 and Race1. The product of this PCR is about 950bp and produces sequence with did7 primer but only very poor sequence with did12 primer. The did12 primer should produce sufficient sequence information to provide an overlap between the sequence obtained from the did2-Race1 RT-PCR and the 5'RACE PCR, but no overlap was obtained. The electrophoretogram from did12 shows many positions with 2 possible base calls suggesting that two different but related PCR products may be present.

2.2.15 Sequence Analysis

The cDNA sequence determined to date is shown in Appendix I and has been submitted to the GenEMBL database with accession number Y08160. The genomic sequence is shown in Appendix II. All primers used in this work have been noted on the sequence in Appendix I.

The essential features of the class V unconventional myosins are; a myosin head domain with ATP-binding and actin-binding domains, a series of 'IQ' repeats, and a tail domain with regions that are predicted to form coiled-coil α-helices. Analysis of the predicted amino acid from 5940bp of cDNA sequence demonstrates all features of a class V unconventional myosin. Figure 2.22 shows the predicted amino acid sequence with these features highlighted while figure 2.23 is an alignment of all class V myosins using the GCG PILEUP program with the mouse 'GAD' sequence also included.

The coiled-coil regions were predicted using the algorithm of Lupas et al (1991) in the form of the Macstripe program adapted for Apple Macintosh computers by Knight (1994). Coiled-coils show a heptad repeating unit containing hydrophobic and hydrophilic residues and are considered to represent regions of protein-protein interaction. Figure 2.24 shows a sample of the output from the Macstripe program.

Comparison of the genomic DNA and cDNA reveals that the gene contains at least 16 exons. A representation of this genomic structure is shown in figure 2.25.

It was observed that the original cDNA sequenced, NMC7, included a 3' untranslated region of 661bp These UTRs are responsible for the localisation of RNA transcripts from a number of genes involved in oogenesis and embryogenesis, for example, *hu-li-tao-shao* (Yue and Spradling, 1992), *bicoid* and *oskar* (St Johnston and Nüsslein-Volhard, 1992). The transcript is seen localised to the developing oocyte by *in situ* hybridisation to whole mount ovaries, see figure 2.1. Analysis of the 3'UTR and its role in localisation of the transcript has been carried out by A McCormack.

2.2.15.1 A Polymorphism in the 3'UTR

A comparison of cDNA and genomic sequence showed a 20bp region missing from the genomic DNA within the 3'UTR. This genomic sequence was obtained from the pE5 plasmid subcloned from FIX-1A. Sequence from pP1-X6, subcloned from P1 DS00574, showed this 20bp to be present. The λFIX library was prepared from a

Canton-S fly line and while the DNA source for the P1 library is unknown, it is presumably not Canton-S. This 20bp of sequence represents a polymorphism between fly lines and would suggest that this region of the 3'UTR is not critical for its function.

A

```
  1  VKEVELKADG SDLPPLRNPA ILVGQNDLTT LSYLHEPGVL HNLRVRFCKR

 51  QIIYTYCGII LVAINPYAEM PLYGPSIIRA YRGHAMGDLE PHIFALAEEA

101  YTKLERENCN LSIIVSGESG AGKTVSAKYA MRYFAAVGGS ESETQVERKV

151  LASSPIMEAF GNAKTTRNDN SSRFGKFTNL LFRNQMGVMF LQGPTMHTYL

201  LEKSRVVYQA QGDRNYHIFY QLCAARSKYP ELVLDHQDKF QFLNMGGAPE

251  IERVSDAEQF NETVQAMTVL GFSIQQIADI VKILAGILHL GNIQVSKKFN

301  EGSEEEDSDS CDIFHNDIHL QITADLLRVS ADDLRRWLLM RKIESVNEYV

351  LIPNSIEAAQ AARDALAKHI YAKLFQYIVG VLNKSLNNGS KQCSFIGVLD

401  IYGFETFEVN SFEQFCINYA NEKLQQQFNQ HVFKLEQEEY LKEGITWTMI

451  DFYDNQPCID LIESRLGVLD LLDEECRMPK GSDESWAGKL IGKCNKFPHF

501  EKPRFGTTSF FIKHFSDTVE YDVNGFLEKN RDTVSKELTQ VLSESNMSLA

551  KQVMTLEEID TLCVDSAKSS TLGGRVVISA GRKQVVPSKQ HRKTVGSQFQ

601  ESIASLISTL HATTPHYVRC IKPNDDKVAF KWETAKIIQQ LRACGVLETV

651  RISAAGFPSR WLYPDFYMRY QLLVYRSKLD KNDMRLSCRN IVMKWIQDED

701  KYRFGNTQIF FRAGQVAFLE QVRANLRKKY ITIVQSVVRR FVYRRQVLRI

751  QKVINGIQKH ARGYLARERT QKMREARAGL ILSKYARGWL CRRRYLRLRH

801  SISGIQTYAR GMLARNKFHA MRDHYRAVQI QRFVRGALAR RAYQKRRRNI

851  IICQAAIRRF LARRKFKRMK AEAKTISHME NKYMGLENKI ISMQQRIDEL
```

Figure 2.22 A) The deduced peptide sequence for the Drosophila class V unconventional myosin. The sequence lacks approximately the first 50 amino acids (based on the alignment in figure 2.23). The boxed region beginning at residue 115 represents the ATP-binding domain, that beginning at residue 604 represents the actin-binding domain. The open boxes with residues highlighted are the IQ motifs. The highlighted underlined residues show a very high probability (>90%) of forming a coild-coil α-helix based on the algorithm of Lupas et al (1991). B) An alignment of the IQ motifs showing the consensus residues.

```
 901  NRDNSNLKHK TSEISVLKMK LELKKTLEAE FKNVKAACQD KDKLIEALNK
 951  QLEAERDEKM QLLEENGHAQ EEWISQKQTW RQENEELRRQ IDEIIDMAKN
1001  AEVNQRNQED RMLAEIDNRE LNEAYQRAIK DKEVIENENF MLKEELSRLT
1051  AGSFSLHGRK ASNASSQNED DVGYASAKNT LDINRPPDLL SKNYSYNDST
1101  SLVVKLRSIL EEEKQKHKVL QEQYIKLSSR HKPTEDSFRV SELEVENEKL
1151  RRRYDQLRTS IKHGVEINEL NAQHAALQEE VRRRREECIQ LKAVLLQQSQ
1201  SMRSLEPESL QMRGNDVNEL MEAFHSQKLI NRQLESELKA ITEEHNSKLV
1251  EMTQEIERLN NEKDELQKVM FESIDEFEDS NVDTLRQNDR YLRRELQKAV
1301  AQFLLVQEEL KLANAKLKAY RQDGGQLEHK IEEEMIRNKS NGTSADVGAN
1351  VTKQKSQNPQ GLMKFHSSDL DKILQRLLSA LTPRTVVGLL PGFPAYLIFM
1401  CIRYTDLTNA DDDVRELLSK FVIQIKKMHR TPHPIENRVI WLVNSITLLN
1451  LMKQYGDVDE YVKFNTEKQN QQQLKNFNLF EYRRVILDLF VNLYQALIMQ
1501  IQGLLDPKIV PAILNNDEIQ RGRQAHGMRS RATSIGASSS PEHGGGPAWK
1551  QLIGQLEHFY KQFQHFGLDN CYAEQIFHQL LYFICAVALN CLMLRGDICI
1601  WETGMIIRYN IGCIEDWVRS KKMSNDVLTA LAPLNQVSQL LQSRKSEQDV
1651  QTICDLCTSL STAQVLKVMK SYKLDDYESE ITNVFLEKLT EKLNARQMQK
1701  SNSDEFTIDQ KFIQPFKVVF RYSDIKLEDI ELPSHLNLDE FLTKI*
```

B

```
725     NLRKKYITIVQSVVRRFVYRRQVL
749      RIQKVINGIQKHARGYLARERTQKM
774      REARAGLILSKYARGWLCRRRYL
797      RLRHSISGIQTYARGMLARNKFHAM
822      RDHYRAVQIQRFVRGALARRAYQ
845      KRRRNIIICQAAIRRFLARRKFKRM

consensus         IQ   RG   RR
```

Figure 2.22 continued

```
       1                                                            50
dilute maaselytkf arvwipdpee vwksaellK. dykpGdkvLl LhlEeGkdl.
chp190 maaselytky arvwipdpee vwksaellK. dykpGdkvLq LrlEeGkdl.
  mgad .......... .......... .......... .......... ..........
 didum .......... .......... .......... .......... .......vk.
 ymyo2 ....msfevg trcwyphkel gwigaeviKn efndGkyhLe LqlEddeivs
 ymyo4 ....msfevg tkcwyphkeq gwiggevtKn dffeGtfhLe LklEdGetvs
  mya1 .maapviivg shvwvedphl awidgevtri d....ginvh vktkkGktv.

       51                                                           100
dilute ....eyrlDp ktgeLPhLRN PdILvGenDL TaLSYLhEPA VLHNLrvRfi
chp190 ....eyclDp ktkeLPpLRN PdILvGenDL TaLSYLhEPA VLHNLkvRfi
  mgad .......... .......... .......... .......... ..........
 didum ....evelka dgsdLPpLRN PaILvGqnDL TtLSYLhEPg VLHNLrvRfc
 ymyo2 vdtkdlnnDk dqs.LPlLRN PpILeateDL TsLSYLnEPA vlhaikqRys
 ymyo4 ietnsfenDd dhptLPvLRN PpILestdDL TtLSYLnEPA vlhaikkRym
  mya1 vtnvyfpkDt eaps...... ....gGvdDm TkLSYLhEPg VLRNLetRye

       101                                                          150
dilute dsklIYTYcG IVLvaiNPYe qlp.iygedi InAYsGqnmG dmdPHiFAvA
chp190 dsklIYTYcG IVLvaiNPYe qlp.iygedi InAYsGqnmG dmdPHiFAvA
  mgad .......... .......... .......... .......... ..........
 didum krqiIYTYcG IiLvaiNPYa emp.lygpsi IrAYrGhamG dLePHiFAlA
 ymyo2 qln.IYTYsG IVLiatNPfd rvdqlytqdm IqAYaGkrrG eLePHlFAiA
 ymyo4 ngq.IYTYsG IVLiaaNPfd kvdhlysrem IqnYsskrkd eLePHlFAiA
  mya1 lne.IYTYtG niLiavNPfq rlphiyetdm meqYkGialG eLsPHvFAig

       151                                                          200
dilute EEAYkqMard erNQSIIVSG ESGAGKTVSA KYaMRYFAtV sGsasea...
chp190 EEAYkqMard erNQSIIVSG ESGAGKTVSA KYaMRYFAtV sGsasea...
  mgad .......... .......... .......... .......... ..........
 didum EEAYtklere ncNlSIIVSG ESGAGKTVSA KYaMRYFAaV gGseset...
 ymyo2 EEAYrlMknd kqNQtIvVSG ESGAGKTVSA KYiMRYFAsV eeensatvqh
 ymyo4 EEAYrfMvhe kaNQtvvVSG ESGAGKTVSA KYiMRYFAsV qesnn..reg
  mya1 daAYraMine gkNnsIlVSG ESGAGKTett KmlMRYlAfl gGrsg.....
                          ATP-binding domain
```

Figure 2.23 Alignment of class V unconventional myosins using the GCG PILEUP program. Positions where amino acids are greater than 50% identical have been highlighted by capitalising. The ATP-binding domain, actin-binding domain and IQ motifs have been boxed. The sequences compared are; *dilute* from mouse (Mercer et al, 1991), chp190 from chicken (Espreafico et al, 1992), mgad from mouse (Huang et al, 1990), didum, the new *Drosophila* sequence (this work), ymyo2 from yeast (Johnston et al, 1991), ymyo4 from yeast (Haarer et al, 1994) and mya1 from *Arabidopsis* (Kinkema and Schiefelbein, 1994)

```
      201                                                        250
dilute .....nvEeK VLASNPIMEs iGNAKTTRND NSSRFGKYiE igFDKryrIi
chp190 .....nvEeK VLASNPIMEs iGNAKTTRND NSSRFGKYiE igFDKryrIi
  mgad ..........  ..........  ..........  ..........  ..........
 didum .....qvErK VLASsPIMEA FGNAKTTRND NSSRFGKftn llfrnqmgvm
 ymyo2 qvemsetEqK iLAtNPIMEA FGNAKTTRND NSSRFGKYlE ilFDKdtsIi
 ymyo4 evemsqiEsq iLAtNPIMEA FGNAKTTRND NSSRFGKYlq ilFDenttIr
  mya1 .vegrtvEqq VLeSNPvlEA FGNAKTlRNn NSSRFGKfvE iqFDKngris

      251                                                        300
dilute  ...GAnmRTY LLEKSRVVfQ aeeERNYHIF YQLCAsaklp efkmLrLgna
chp190  ...GAnmRTY LLEKSRVVfQ aeeERNYHIF YQLCAsaalp efktLrLgna
  mgad  ..........  ..........  ..........  ..........  ..........
 didum  flqGptmhty LLEKSRVVyQ aqgdRNYHIF YQLC..aars kypeLvLdhq
 ymyo2  ...GAriRTY LLErSRlVyQ ppiERNYHIF YQLmAglpaq tkeeLhLtda
 ymyo4  ...GskiRTY LLEKSRlVyQ petERNYHIF YQileglpep vkqeLhLssp
  mya1  ...GAaiRTY LLErSRVcqi sdpERNYHcF YLLCAapped ikk.ykLenp

      301                                                        350
dilute  dsfHYtkQGG spmIeGvDDA kEmahTrqAc TLlGISesyQ mgIFrILAGI
chp190  nyfHYtkQGG spvIdGiDDA kEmvnTrqAc TLlGISdsyQ mgIFrILAGI
  mgad  ..........  ..........  ..........  ..........  ..........
 didum  dkfqflNmGG apeIervsDA eqfneTvqAm TvlGfSiqqi adIvkILAGI
 ymyo2  sdyfYmNQGG dtkInGiDDA kEykiTvdAl TLvGItketQ hqIFkILAal
 ymyo4  kdyHYtNQGG qpnIaGiDeA rEykiTtdAl sLvGInhetQ lgIFkILAGl
  mya1  hkfHYlNQss cykldGvDDA sEyleTrrAm dvvGISneeQ eaIFrvvAaI

      351                                                        400
dilute  LHLGNvgfas r........D sdsctippkh epltifCdLm gvDyeemchW
chp190  LHLGNvefas r........D sdscaippkh dpltifCdLm gvDyeemahW
  mgad  ..........  ..........  ..........  ..........  ..........
 didum  LHLGNIqvsk kfnegseeeD sdscdifhnd ihlqitadLL rvsaddlrrW
 ymyo2  LHIGNIeikk .....trn.D aslsadep.. .nlklaCeLL giDaynfakW
 ymyo4  LHIGNIemkm .....trn.D aslsseeq.. .nlqiaCeLL giDpfnfakW
  mya1  LHLGNIdfgk .....geeiD ssvikdkdsr shlnmaaeLL mcnaqsleda

      401                                                        450
dilute  Lchrklatat Etyikpiskl qAtnARDaLA KhIYakLFnW IVdhvnq...
chp190  Lchrklatat Etyikpiskl hAinARDaLA KhIYanLFnW IVdhvnk...
  mgad  ..........  ..........  ..........  ..........  ..........
 didum  Llmrkiesvn Eyvlipnsie aAqaARDaLA KhIYakLFqy IVgvlnk...
 ymyo2  vtkkqiitrs Ekivsnlnys qAlvAkDsvA KfIYsaLFdW LVenintvlc
 ymyo4  ivkkqivtrs Ekivtnlnyn qAliARDsvA KfIYstLFdW LVdninktly
  mya1  Lirrvmvtpe Eiitrtldpd nAiasRDtLA KtIYshLFdW IVnkint...
```

Figure 2.23 continued

```
        451                                                        500
dilute  ..aLhsavkq hSFIGVLDIY GFEtFEiNSF EQFCINYANE KLQQqFNmhv
chp190  ..aLhstvkq hSFIGVLDIY GFEtFEiNSF EQFCINYANE KLQQqFNmhv
  mgad  .......... .......... .......... .......... ..........
 didum  ..sLnngskq cSFIGVLDIY GFEtFEvNSF EQFCINYANE KLQQqFNQHV
 ymyo2  npavn..dqi sSFIGVLDIY GFEhFEkNSF EQFCINYANE KLQQeFNQHV
 ymyo4  dpeLdqqdhv fSFIGiLDIY GFEhFEkNSF EQFCINYANE KLQQeFNQHV
  mya1  ..sigqdprs kSiIGvLDIY GFEsFkcNSF EQFCINftNE KLQQhFNQHV

        501                                                        550
dilute  FKLEQEEYmK EqIpWtlIdF yDNQPCINLI EskL.giLdL LDEECkmPKG
chp190  FKLEQEEYmK EqIpWtlIdF ydNQPCINLI Eakm.gvLdL LDEECkmPKG
  mgad  .......... .......... .......... .......... ..........
 didum  FKLEQEEYlK EgItWtmIdF yDNQPCIDLI EsrL.gvLdL LDEECrmPKG
 ymyo2  FKLEQEEYvK EeIeWsfIeF nDNQPCIDLI EnkL.giLsL LDEEsrlPaG
 ymyo4  FKLEQEEYvK EeIeWsfIeF sdNQPCIDLI EnkL.giLsL LDEEsrlPsG
  mya1  FKmEQEEYtK EeIaWsyieF iDNQDvleLI EkkpggiisL LDEaCmfPks

        551                                                        600
dilute  tDdtWaQKLY nthlnkca.. lFeKPRmsnk aFIikHFAdk VEYqcEGFlE
chp190  SDdtWaQKLY nthlnkca.. lFeKPRlsnk aFIikHFAdk VEYqcEGFlE
  mgad  .......... .......... .......... .......... ..........
 didum  SDEsWagkli gkc.nkfp.. hFeKPRfgtT sFfikHFsdt VEYdvnGFlE
 ymyo2  SDEsWtQKLY qtldksptnk vFsKPRfgqT kFIvsHYald VaYdvEGFiE
 ymyo4  SDEsWasKLY safnkppsne vFsKPRfgqT kFIvsHYAvd VEYevEGFiE
  mya1  thEtfsQKlf qtfke...he rFaKPklsrT dftisHYAge VtYqsnhFid

        601                                                        650
dilute  KNkDTVfeeq ikVLKss.kf kmlpeLfqdd ekaisptsat ssgrtpltrv
chp190  KNkDTVyeeq ikVLKsskkf kllpeLfqde ekaisptsat psgrvplsrt
  mgad  .......... .......... .......... .......... ..........
 didum  KNrDTVskel tqVLsesnms lakqvmtlee idtlcvdsak sst...lggr
 ymyo2  KNrDTVsdgh leVLKastne tliniLegle kaakkleeak kleleqagsk
 ymyo4  KNrDsVslgh ldVfKattnp ifkqiLdn.. ...relrsdd apeeqntekk
  mya1  KNkDyivaeh qalftasnck fvaglfhalh edssr..... ..........

        651                                                        700
dilute  pvkptkgrpg qtakehKkTv GhqFrnsLhl LMeTLnaTtP HYvRCIKPNd
chp190  pvkpakarpg qtskehKkTv GhqFrnsLhl LMeTLnaTtP HYvRCIKPNd
  mgad  .......... .......... .......... .......... ..........
 didum  vvisagrkqv vpskqhrkTv GSqFqesLas lisTLhaTtP HYvRCIKPNd
 ymyo2  kpgpir.... ..tvnrKpTl GSmFkqsLie LMnTinsTnv HYiRCIKPNa
 ymyo4  imipar.... ..lsqkKpTl GSmFkksLge LMaiinsTnv HYiRCIKPNs
  mya1  .......... ...sskfssi GSrFkqqLhs LMesLngteP HYiRCIKPNn
```

actin-binding domain

Figure 2.23 continued

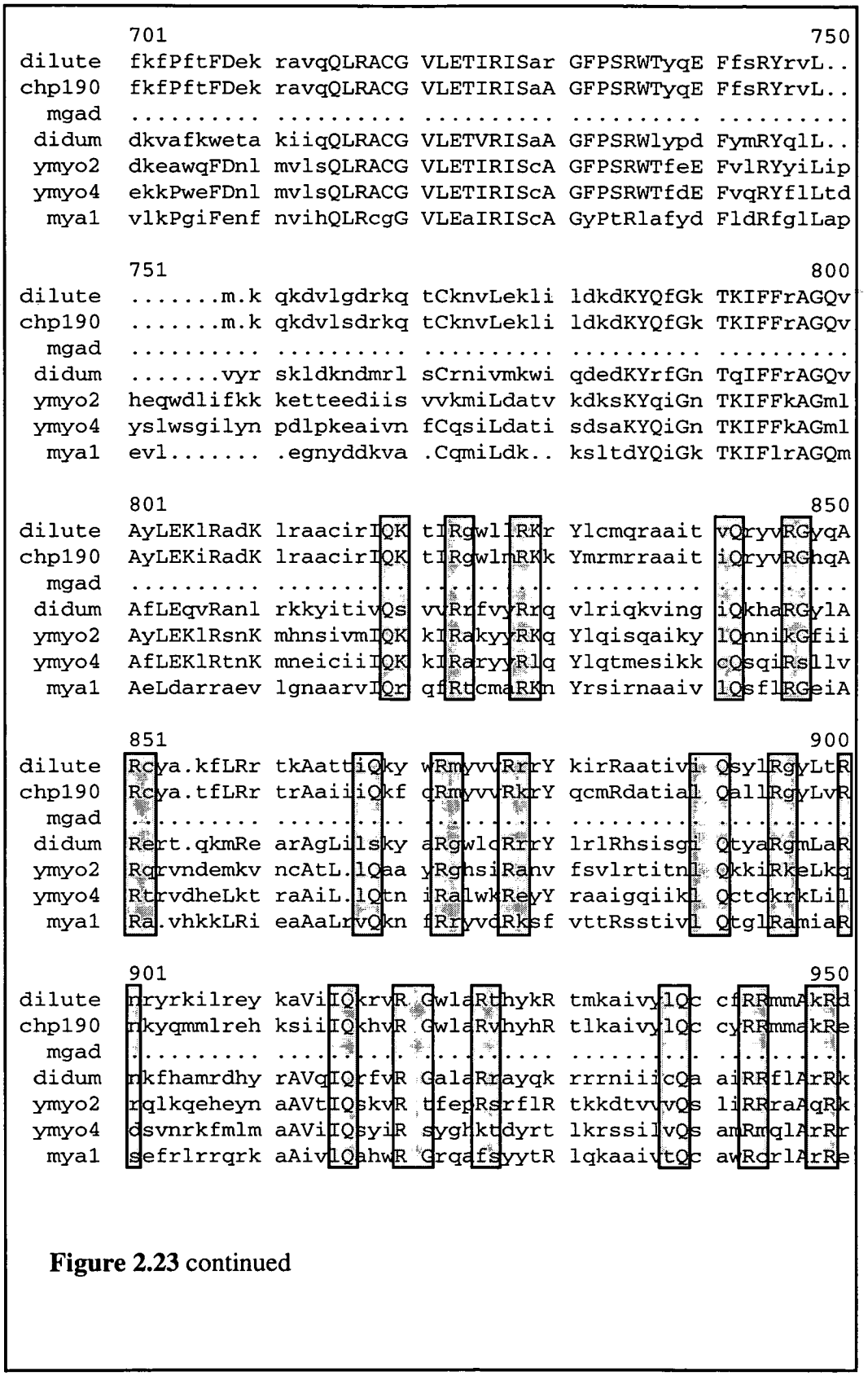


Figure 2.23 continued

```
       951                                                        1000
dilute  vKkLKiEars veryKklhiG mENKimqLqr kvdeqnkdyk clmekltnle
chp190  lKkLKiEars veryKklhiG LENKimqLqr kideqnkeyk sllekmnnle
  mgad  .......... .......... .......... .......... ..........
 didum  fKrmKaEakt ishmenkymG LENKiismqq ridelnrdns nlkhk.....
 ymyo2  lKqLKadaks vnhlKevsyk LENKvieLtq nl........ ...askvken
 ymyo4  yivLqkEvee rnirasygiG Lleeaiefkn sf........ ..........
  mya1  lrmLKmaard tgalKdaknk LEqrveeLsl rlhlekrlrt dleeakvqev

       1001                                                       1050
dilute  gvynsEtekL rndvErlqls EeeaKvatgr vlslqeeIak LrKdLeqtrs
chp190  itystEtekL rsdvErlrms EeeaKnatnr vlslqeeIak LrKeLhqtqt
  mgad  .......... .......... .......... .......... ..........
 didum  ...tsEisvL kmklElkktl EaefKnvkaa cqdkdklIea LnKqLeaerd
 ymyo2  kemterikeL qvqve..... .......... .......... ..........
 ymyo4  ......ilnL emlnd..... .......... .......... ..........
  mya1  aklqealhtm rlqlk..... .......... .......... ..........

       1051                                                       1100
dilute  EKksieErad kykqEtdqlv snLkeentll kqeketlnhr ive.qakemt
chp190  EKktieEwad kykhEteqlv seLkeqntll ktekeelnrr ihd.qakeit
  mgad  .......... .......... .......... .......... ..........
 didum  EKmqllE... ....Enghaq eewisqkqtw rqeneelrrq ideiidmakn
 ymyo2  .......... ....Esaklq etLenmkkeh lididnqksk dmelqkt.ie
 ymyo4  .......... ....sytrlt qlLqg..... ..dlsnipsk qrqeyet.iv
  mya1  .......... ....Ettamv vkeqeaarva ieeassvnke pvvvedteki

       1101                                                       1150
dilute  etmerklvee tkqlElDlnd erlrYQnlln efsrlEeryd dLKeEmtlml
chp190  etmekklvee tkqlElDlnd erlrYQnlln efsrlEeryd dLKdEmnlmv
  mgad  .......... .......... .......... .........m eLrdEqt...
 didum  aevnqrnqed rmlaEiDnre lneaYQraik dkeviEnenf mLKeElsrlt
 ymyo2  nnlqsteq.. .......... .......... .......... ..........
 ymyo4  ngyndkis.. .......... .......... .......... ..........
  mya1  dslsneid.. .......... .......... .......... ..........

       1151                                                       1200
dilute  nvpkPGHkRt dSthSS.NES eytfs....S EfaetEDiap rtE.epiEkK
chp190  sipkPGHkRt dSthSS.NES eytfs....S EiteaEDlpl rmEqepsEkK
  mgad  ....PGHrkn pSnqSS.lES dsnypsistS EigdtEDalq qvEeigiE.K
 didum  agsfslHgRk aSnaSSqNEd dvgyasak.. ...ntlDinr ppdllsknys
 ymyo2  .......... .......... .......... .......... ..........
 ymyo4  .......... .......... .......... .......... ..........
  mya1  .......... .......... .......... .......... ..........
```

Figure 2.23 continued

```
        1201                                                    1250
dilute  vplDMSLFLK LqkrvtEleQ ekqlmQdeLd r..keeqvfr skaKeeerp.
chp190  aplDMSLFLK LqkrvtEleQ ekqsLQdeLd r..keeqaLr akaKeeerp.
  mgad  aamDMtvFLK LqkrvrEleQ erkkLQaqLe k..gqqdskk gqveqqnngl
 didum  yndstSLvvK LrsileEekQ khkvLQeqyi k.......Ls srhKpte...
 ymyo2  .........t Lk........ ...daqleLe dmvkqhdeLk eesKkqleel
 ymyo4  .........K Lk........ ...tLQveim ntlnkknaLk erkKkqssli
  mya1  .........r Lkgllsseth kadeaQhayq salvqneeLc kkleeagrki

        1251                                                    1300
dilute  .qirgaeleY esLKrqELES ENKKLKnEln elrkalseks apevtapgap
chp190  .pirgaeleY esLKrqELES ENKKLKnEln elqkaltetr apevtapgap
  mgad  dvdqdadiaY nsLKrqELES ENKKLKndln elrngvadqa mqdnsthssp
 didum  .......... dsfrvsELEv ENeKLrrryd qlrtsikhgv eine......
 ymyo2  eqtkktlveY qtLn.gdLqn EvKsLKeE.. .i..arlqta mslgtvttsv
 ymyo4  qshmqslaai kgnKpsrLsd EvKsmKqE.. .l..afienv iaqdfttt..
  mya1  dqlqdsvqrf qe.KvfsLES EnKvLrqqtl ti..spttra lalrpkttii

        1301                                                    1350
dilute  .ayrvLmeQl tsvsEEldvR kEEvliLrsq LvsQKeaiqp kddkntmtds
chp190  .ayrvLldQl tsvsEElevR kEEvliLrsq LvsQKeaiqp kedkntmtds
  mgad  dsyslLlnQl klanEElevR kEEaliLrtq imnadqrrls gknmepnina
 didum  .....LnaQh aalqEEvrrR rEEciqLkav LlqQsqsmrs lepeslqmrg
 ymyo2  lpqtpLkdvm gggasnfnnm mlEnsdLspn dlnlKsrstp ssgnnhidsl
 ymyo4  ........ys ankndkvkgl giagqqvkpk Lvnv...... ..........
  mya1  qrtpekdtfs ngettqlqe. .pEtedrpqk slnQKqq... enqelllksi

        1351                                                    1400
dilute  tilledvqkm kDkgeiaqAY iGlkeTNRLL EsqLqsqkrs HeneaeaLrg
chp190  tilledvqkm kDkgeiaqAY iGlkeTNRLL EsqLqsqkks HeneLesLrg
  mgad  rtswpnsekh vDqedaieAY hGvcqTNRLL EaqLqaqsle HeeevehLka
 didum  .......... nDvnelmeAf hsqkliNRqL EseLkaitee HnskLvemtq
 ymyo2  sv........ .......... .......... drengvnatq ineeLyrLle
 ymyo4  .i........ .......... .......... rresg..... .npdLleLlm
  mya1  se........ .......... .......... digfsegkpv aacliykcli

        1401                                                    1450
dilute  eiqsLkeEnn rQQQllaQnL qLpPEAriEa slQheItRLT NENLyfeely
chp190  eiqsLkeEnn rQQQllaQnL qLpPEAriEa slQheItRLT NENL......
  mgad  qveaLkeEmd kQQQtfcQtL lLsPEAqvEf gvQqeIsRLT NENL......
 didum  eierLnnEkd elQkvmfesi defedsnvdt lrQnd..Ryl rreL......
 ymyo2  dteiLnqEit e......... .......... .......... ..........
 ymyo4  dlncytlEvt e......... .......... .......... ..........
  mya1  hwrsfevErt s......... .......... .......... ..........
```

Figure 2.23 continued

```
          1451                                                    1500
dilute    addpkkyqsy rislykrmiD lmEqLEKqdK tvRKLKKQLK vfaKKIgelE
chp190    .......... .........D lmEqLEKqdK tvRKLKKQLK vfaKKIgelE
  mgad    .......... .........D fkElvEKleK neRKLKKQLK iymKKvqdlE
 didum    .......... .qkavaqfll vqEeLklana klkayrqdgg qlehKIeeem
 ymyo2    .......... .......... .......... .......... gllKgfevpd
 ymyo4    .......... .......... .......... .......... gylKKvnvtE
  mya1    .......... .......... .......... .......... ifnriIetia

          1501                                                    1550
dilute    vgqmenispg qiidepirpv niprKEKDFQ GMLEYkrEde qkLvkNLile
chp190    vgqmenispg qiidepirpv niprKEKDFQ GMLEYkkEde qkLvkNLile
  mgad    aaqalaqs.d rrhheltrqv tvqrKEKDFQ GMLEYhkEve alLirNLvtd
 didum    irnksngtsa dvganvtkq. ....KsqnpQ Glmkfhssdl dkilqrLlsa
 ymyo2    agvaiqlskr dv........ .......... .......... ..........
 ymyo4    vngd...... nv........ .......... .......... ..........
  mya1    saiemq.ens dv........ .......... .......... ..........

          1551                                                    1600
dilute    LKPrgvavnL iPglPAYilf MCvRhaDylN dDqkVrsLLt stInsIKKvL
chp190    LKPrgvavnL iPglPAYilf MCvRhaDylN dDqkVrsLLt stIngIKKvL
  mgad    LKPq.mllgt vPclPAYily MCiRhaDytN dDlkVhsLLs stIngIKKvL
 didum    LtPrtv.vgL lPgfPAYlif MCiRytDltN aDddVreLLs kfviqIKK.m
 ymyo2    .........v yPariliivl semwrfgltk qsesflaqvl ttIqkvvtqL
 ymyo4    .........L gPihvittvv sslvrnglli qsskfiskvl ltvesivmsL
  mya1    .........L c......... ...ywlsnsa tllmflqrtl kagatgsitt

          1601                                                    1650
dilute    KKrgddfEtv sFWLSNtcrf ..LhcLKQYS geegfmkhNt srQNehcLtN
chp190    KKrgddfEtv sFWLSNtcrf ..LhcLKQYS geegfmkhNt prQNehcLtN
  mgad    KKhnddfEmt sFWLSNtcrf ..LhcLKQYS gdegfmtqNi akQNehcLkN
 didum    hrtphpiEnr viWLvNsitl ..LnlmKQYg dvdeyvkfNt ekQNqqqLkN
 ymyo2    Kgndlipsgv .FWLaNvrel ysfvvfalnS ilteetfkNg mtdeeykeyv
 ymyo4    pKdetmlggi .FWLSN...l srLpafaan. ..qktlyean ggdekdkLtl
  mya1    prrrgmpssl fgrvsq..sf rgspqsagfp fmtgraiggg ldelrqveak

          1651                                                    1700
dilute    FdLaEYRqVL sDLaiqIYQq LvrvleniLQ PmIVsgmLeh EtIQgvsGvk
chp190    FdLaEYRqVL sDLaiqIYQq LvrvleniLQ PmIVsgmLeh EtIQgvsGvk
  mgad    FdLtEYRqVL sDLsiqIYQq LikmpeglLQ PmIVsamLen EsIQglsGvr
 didum    FnLfEYRrVi lDLfvnLYQa LimqiqglLd PkIVpaiLnn deIQ..rGrq
 ymyo2    sLvtelkddf eaLsynIYni wlkklqkqLQ kkainav... vi.....ses
 ymyo4    iylndlenet lkvfdkIYst wlvkfmkh.. ..asahi... Ei.....fdm
  mya1    ypallfkqqL taflekIYgm irdkmkkeis Pllasci... qvprtprsgl
```

Figure 2.23 continued

```
        1701                                                    1750
dilute  ptGlrkRtsS ia......de GtytldsIlr qLNSFhsvmC qhGmDpElIk
chp190  ptGlrkRtsS ia......de GtytldsIir qLNSFhsvmC qhGmDpElIk
  mgad  ptGyrkRssS mv......dg e......... ..NSFhtvlC dqGlDpEiIl
 didum  ahGmrsRatS igassspehg Ggpawkqlig qLehFykqfq hfGlDncyae
 ymyo2  lpGfsagetS gflnkifant eeytmddIlt ffNSiywcmk sfhienEvfh
 ymyo4  v......... .lneklfkns Gdekfaklft fLNeFdavlC kfqvvdsmht
  mya1  vkGrsqntq. ..nnvvapkp miahwqnIvt cLNghlrtmr anyvpsllIs

        1751                                                    1800
dilute  QVvkQmFYiv gAitLNnLlL RkDmCSWskG MqiRYNvsqL EEWlrdKnlm
chp190  QVvkQmFYiI gAvtLNnLlL RkDmCSWskG MqiRYNvsqL EEWlrdKnlm
  mgad  QVfkQlFYmI nAvtLNnLlL RkDaCSWstG MqlRYNisqL EEWlrgKnlh
 didum  QifhQllYfI cAvaLNcLmL RgDiCiWetG MiiRYNigci EdWvrsKkms
 ymyo2  aVvttllnyv dAicfNeLim krnflSWkrG lqlnYNvtrL EEWckthg..
 ymyo4  kifndtlkyl nvmlfNdLit kcpalnWkyG yevdrNierL vsWfepr...
  mya1  kvfgqiFsfI nvqlfNsLlL RreccSfsnG eyvktglaeL EkWchdatee

        1801                                                    1850
dilute  n.sgaketLe PLIQAAQLLQ vKKKtddDae aI.CsmCnal tTAQIvKvLn
chp190  n.sgaketLe PLIQAAQLLQ vKKKtdeDae aI.CsmCnal tTAQIvKvLn
  mgad  q.sgavqtme PLIQAAQLLQ lKKKtheDae aI.CslCtsl sTqQIvKiLn
 didum  n..dvltaLa PLnQvsQLLQ .srKseqDvq tI.CdlCtsl sTAQvlKvmk
 ymyo2  .ltdgtecLq hLIQtAkLLQ vrKytieDid .IlrgiCysl tpAQlqKlis
 ymyo4  .iedvrpnLi qiIQavkiLQ lKisnlnefk .llfdfwyal npAQIqaiLl
  mya1  fvgsawdeLk hirQavgfLv ihqKpkkslk eittelcpvl siqQlyrist

        1851                                                    1900
dilute  lytpvnefEe rvsvsFirti qmrLrdrk.. ..dspqllmD aKhifPvtfp
chp190  lytpvnefEe rvlvsFirti qlrLrdrk.. ..dspqllmD aKhifPvtfp
  mgad  lytplnefEe rvtvsFirti qaqLqern.. ..dpqqlllD sKhvfPvlfp
 didum  syk.lddyEs eitnvFlekl tekLnarqmq ksnsdeftiD qKfiqPfkvv
 ymyo2  qyqvady.Es pipqeilryv adivkkea.. .......als ssgndskghe
 ymyo4  kykpankgEa gvpneilnyl anvikrenls lpgkmeimls aqfdsaknhl
  mya1  mywddkygth svsteviatm raevsdvsks aisn...sfl ldddssipfs

        1901                                                    1950
dilute  fnpSslalet IqiPasLgL. gFIarv*... .......... ..........
chp190  fnpSslalet IqiPasLgL. gFIsrv*... .......... ..........
  mgad  ynpSaltmds IhiPacLnL. eFlnev*... .......... ..........
 didum  frySdikled IelPshLnLd eFltki*... .......... ..........
 ymyo2  h..Sssifit petgpftdpf slIk.trkfd qveayipawl slpstkrivd
 ymyo4  rydtsaitqn snteglatvs kiIkldrk*. .......... ..........
  mya1  lddisksmqn vevaev.dpp plIrqnsnfm fllersd*.. ..........

        1951     1962
 ymyo2  lvaqqvvqdg h*
```

Figure 2.23 continued

1	2	3	4	5	6
855	A	5	d	1.179511	0.020656
856	A	5	e	1.179511	0.020656
857	I	5	f	1.179511	0.020656
858	R	5	g	1.179511	0.020656
859	R	5	a	1.179511	0.020656
860	F	5	b	1.179511	0.020656
861	L	5	c	1.179511	0.020656
862	A	5	d	1.259497	0.055190
863	R	5	e	1.259497	0.055190
864	R	5	f	1.259497	0.055190
865	K	5	g	1.259497	0.055190
866	F	1	e	1.263676	0.058086
867	K	1	f	1.486036	0.559921
868	R	1	g	1.502049	0.615287
869	M	1	a	1.533937	0.716957
870	K	1	b	1.544777	0.747823
871	A	1	c	1.544777	0.747823
872	E	1	d	1.544777	0.747823
873	A	1	e	1.544777	0.747823
874	K	1	f	1.544777	0.747823
875	T	1	g	1.544777	0.747823
876	I	1	a	1.544777	0.747823
877	S	1	b	1.544777	0.747823
878	H	1	c	1.630408	0.913367
879	M	1	d	1.752887	0.985820
880	E	1	e	1.840007	0.996441
881	N	1	f	1.840007	0.996441
882	K	1	g	1.840007	0.996441
883	Y	1	a	1.840007	0.996441
884	M	1	b	1.840007	0.996441
885	G	1	c	1.840007	0.996441
886	L	1	d	1.840007	0.996441
887	E	1	e	1.840007	0.996441
888	N	1	f	1.840007	0.996441
889	K	1	g	1.840007	0.996441
890	I	1	a	1.840007	0.996441
891	I	1	b	1.840007	0.996441
892	S	1	c	1.840007	0.996441
893	M	1	d	1.840007	0.996441
894	Q	1	e	1.840007	0.996441
895	Q	1	f	1.840007	0.996441
896	R	1	g	1.840007	0.996441
897	I	1	a	1.840007	0.996441
898	D	1	b	1.840007	0.996441
899	E	1	c	1.840007	0.996441
900	L	1	d	1.840007	0.996441
901	N	1	e	1.738845	0.982383
902	R	1	f	1.738845	0.982383
903	D	1	g	1.738845	0.982383
904	N	1	a	1.738845	0.982383
905	S	1	b	1.738845	0.982383
906	N	1	c	1.738845	0.982383
907	L	1	d	1.738845	0.982383
908	K	1	e	1.646830	0.931083
909	H	1	f	1.547187	0.754389
910	K	1	g	1.547187	0.754389
911	T	1	a	1.547187	0.754389
912	S	1	b	1.547187	0.754389
913	E	1	c	1.547187	0.754389
914	I	1	d	1.514030	0.655133
915	S	1	e	1.404267	0.288094
916	V	3	d	1.329672	0.128225
917	L	3	e	1.395349	0.263607
918	K	3	f	1.395349	0.263607
919	M	3	g	1.395349	0.263607
920	K	3	a	1.395349	0.263607
921	L	3	b	1.395349	0.263607
922	E	3	c	1.495420	0.592596
923	L	3	d	1.495420	0.592596
924	K	3	e	1.495420	0.592596
925	K	3	f	1.495420	0.592596
926	T	3	g	1.495420	0.592596
927	L	3	a	1.495420	0.592596
928	E	3	b	1.495420	0.592596
929	A	3	c	1.495420	0.592596

1 Residue number
2 Amino acid
3 Heptad frame (1-7)
4 McLachlan position (abcdefg)
5 Score from Lupas method
6 Probability from Lupas method

Figure 2.24 Example of output from the Macstripe program (Knight, 1994) for predicting coiled-coil structures in peptide sequences. Residues 876 to 910 are highly likely to form a coiled-coil domain.

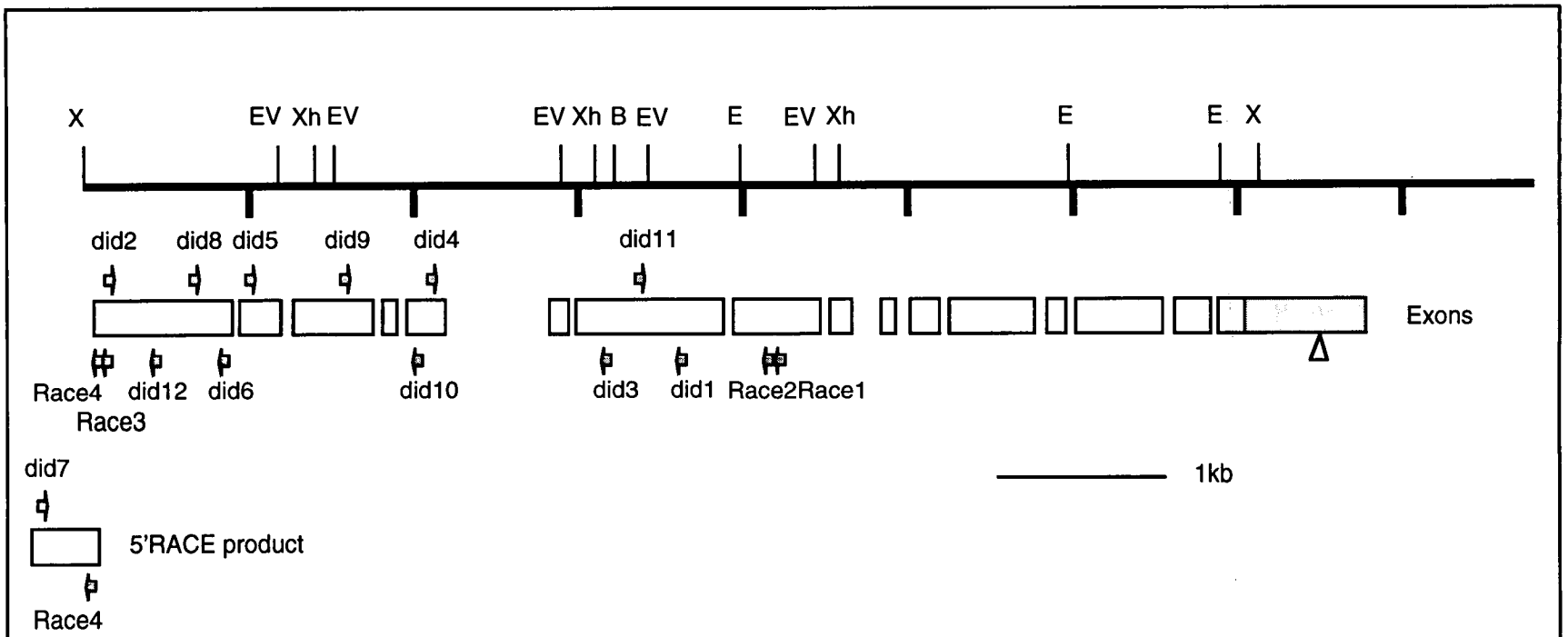


Figure 2.25 Schematic diagram of genomic structure of the class V myosin (*didum*) gene. Primers used for determining the 5' cDNA sequence are shown. The shaded part of the last exon represents the 3'UTR. The open triangle represents the position of the 20bp polymorphism within the 3'UTR. The 5'RACE product is also represented. Restriction endonucleases; B: *Bam*HI, E: *Eco*RI, EV: *Eco*RV, X: *Xba*I, Xh: *Xho*I.

2.3 The Mouse "GAD" is a Class V Myosin

It was clear that the mouse "GAD" described by Huang et al (1990) was unique among the glutamate decarboxylase from the sequences that were deposited in the GenEMBL database, see figure 2.26. In particular, all glutamate decarboxylases show a requirement for the cofactor pyridoxal phosphate, apart from the Huang "GAD" (for review, see Erlander and Tobin [1991], see also Bu et al [1992]). The cofactor binds to a conserved motif with amino acid sequence NPHK (binding is to the lysine [K]), a sequence motif not found in the Huang "GAD". The sequence identity between the Huang "GAD" and the class V myosins strongly suggested that the original sequence reported was a truncated cDNA.

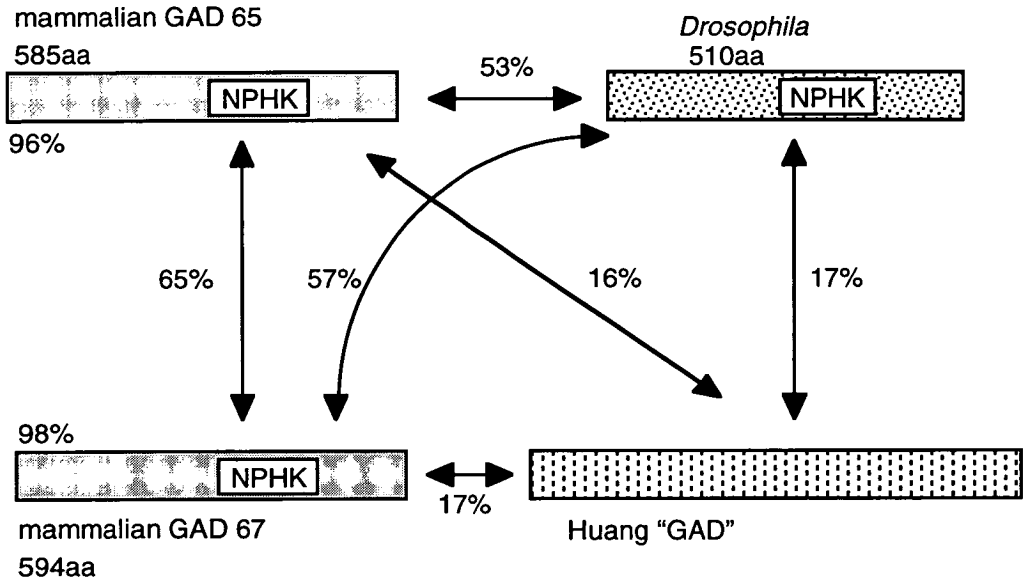


Figure 2.26 Schematic representation of glutamate decarboxylases and the Huang "GAD". Mammalian GADs fall into 2 classes, a 65kDa group and a 67kDa group. Homology within each group is 96% and 98% respectively at the amino acid level. There is also significant sequence similarity between the mammalian GADs and the *Drosophila* GAD (Jackson et al, 1990). The Huang "GAD" shows no obvious sequence similarity to any of these GADs. Comparisons done with the GCG GAP program.

2.3.1 Isolation of Mouse Myosin V cDNAs

A project was carried out by an udergraduate, A Wilke, and continued by a postgraduate on rotation, P Loke, to isolate and characterise cDNAs from a mouse brain library that were homologous to that described by Huang et al (1990). A summary of this project is given below.

Two oligonucleotide primers, Mgad1 and Mgad2, were designed based on the sequence described by Huang et al (1990) to produce a product of 1.2kb by PCR. Total RNA was isolated from liver dissected from mice by the Trizol method (7.10). Liver tissue was chosen as there was an abundance of material but it was not known if the gene was expressed in the liver.

A reverse transcription reaction was performed with Mgad2 primer followed by PCR with both primers and a 1.2kb product was obtained. Restriction with the endonuclease *Sac*I produced a 600bp doublet as seen on agarose gel electrophoresis, a strong indication that this product was derived from the same gene as the Huang sequence. The PCR product was purified, labelled with α-^{32}P-dCTP and used to screen a mouse brain cDNA library in the λZAP vector (Stratagene commercial library). A mouse brain library was chosen to screen as the Huang sequence was isolated from a brain cDNA library. Approximately 4 x 10^5 recombinant bacteriophage were screened with 17 initial positives identified. Of these positives, 7 were able to be plasmid rescued after sucessive rounds of screening. These recombinant plasmids have been analysed by restriction mapping and DNA sequencing.

The insert from the plasmid named ZAP-12 is approximately 4.5-5kb and DNA sequencing has shown that this plasmid contains a region that codes for an actin-binding domain. This finding is sufficient proof that the Huang "GAD" is indeed a class V myosin. However, further evidence comes from a recent publication by Zhao et al (1996) who describe a sequence from rat which encodes a class V myosin and which shows 96% identity at the amino acid level to the Huang "GAD". An alignment between these two sequences is shown in figure 2.27.

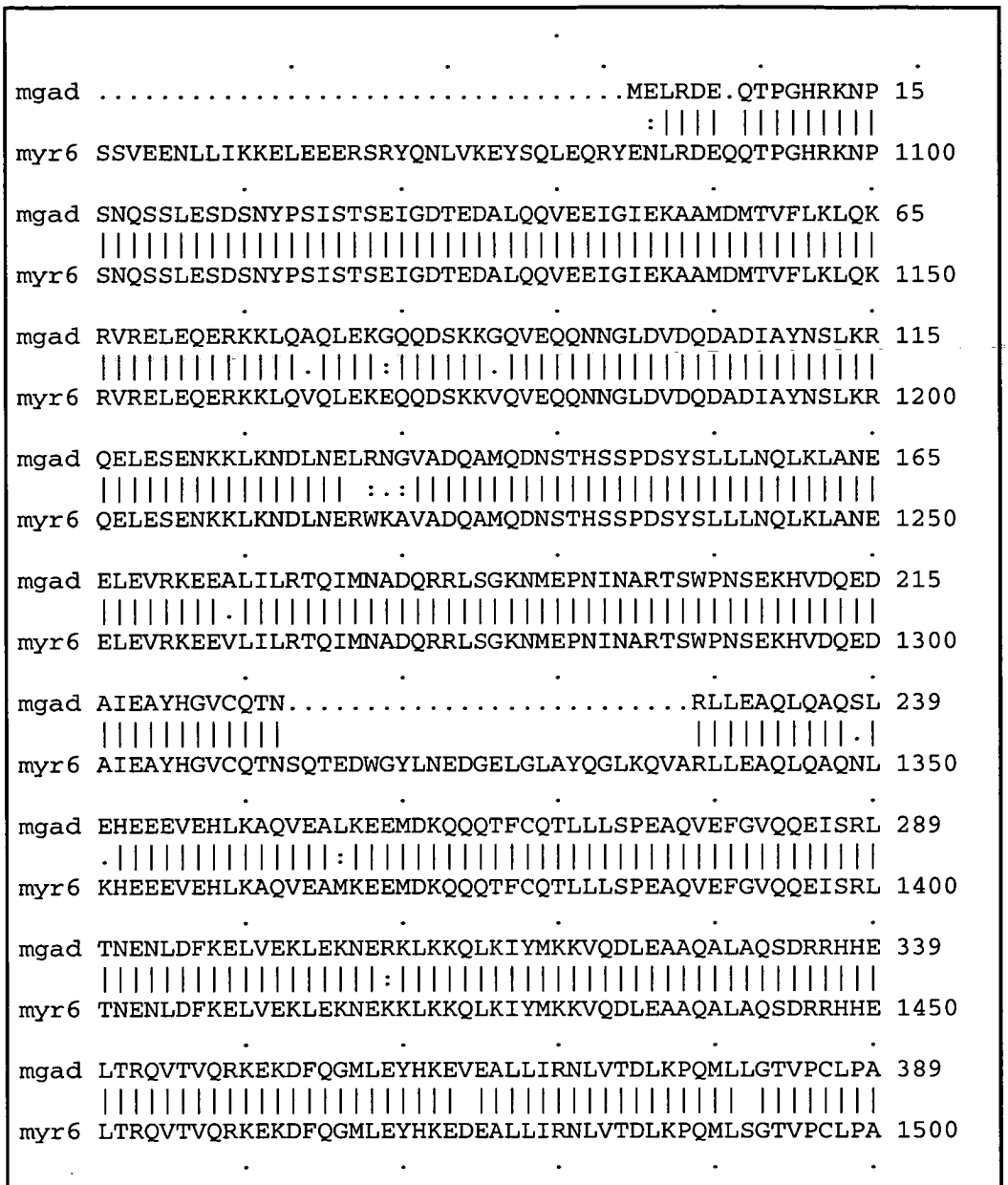


Figure 2.27 Comparison of the Huang "GAD" (mgad) deduced amino acid sequence to that from myr6, a class V myosin from rat (Zhao et al, 1996) using the GCG GAP program. Identity is 96%.

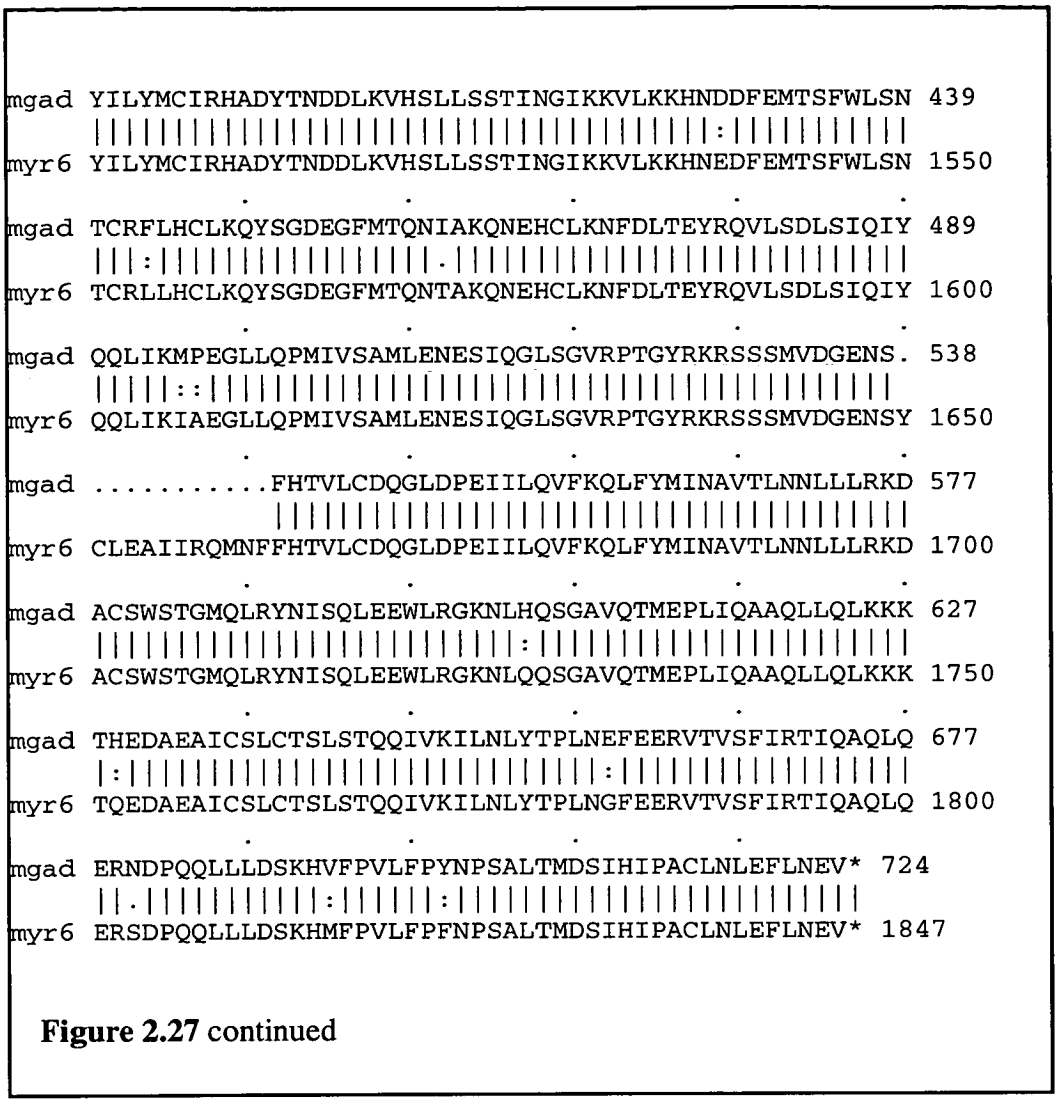


Figure 2.27 continued

2.4 Summary

A reverse genetics screen based on P-element enhancer traps led to the isolation of a cDNA which showed an interesting expression pattern in the ovaries by *in situ* hybridisation. This pattern showed that the gene was expressed early in oogenesis with the transcript localised to the developing oocyte. Localisation was seen as an anterior band in the oocyte during mid stages of oogenesis with strong expression seen in the nurse cells at late stages and loss of localisation in the oocyte. However, the gene is not associated with the P-elements present in the line originally examined.

The original cDNA showed similarity to the class V unconventional myosins. Further cDNAs were isolated from an ovarian library to give a total of 3.2kb but Northern analysis showed a transcript size of approximately 6kb. Further screening of cDNA libraries failed to yield any additional cDNAs. Screening of a genomic library resulted in 6 recombinant phage being isolated from 3 separate screens. However, a more useful subcloned genomic DNA fragment came from a recombinant P1 phage. Sequence analysis showed that this recombinant plasmid contained most of the 5' sequence absent from the cDNAs. A 5'RACE experiment yielded, potentially, the remaining cDNA sequence but it has not been possible to merge this to the bulk of the cDNA sequence.

Analysis of the sequence confirms that the predicted protein shows all the major features of a class V myosin with an ATP-binding domain, an actin-binding domain, IQ motif repeats and regions within the carboxyl tail which show a high probability of forming coiled-coil α-helices.

Attempts to isolate further cDNAs resulted in the identification of two unrelated genes with these results presented in Chapter 5.

Chapter 3

Genetic Analysis

3.1 Introduction

This project used a reverse genetics approach that resulted in the isolation of an class V unconventional myosin and a putative inorganic phosphate cotransporter (cDNA-34, see Chapter 5) from *Drosophila*. To characterise these genes further, it is essential to undertake a genetic analysis with the aim being to match the cloned genes to mapped genes. Mutations that exist in the mapped genes would enable experiments to be undertaken to investigate the role of these genes in the organism. All the work in this chapter was done in collaboration with A McCormack.

3.2 Chromosome Position of the Unconventional Myosin

Salivary glands were dissected from third instar OregonR larvae and chromosomes prepared as described in the materials and methods (7.15). A digoxigenin labelled probe was prepared from the 2.2kb cDNA as described (7.5.5.3) with approximately 100ng of probe used per slide in a 20µl volume. Hybridisation was carried out overnight at 55°C. Detection was carried using HRP-conjugated anti-digoxigenin antibodies and the DAB detection system. The chromosomes were then counterstained with giemsa.

Analysis of the chromosomes using brightfield and phase contrast microscopy indicated that the myosin gene position was on the right arm of the second chromosome. Use of the published cytological maps refined the position to the 43BC region (Lindsley and Zimm, 1992). This position was corroborated by R Saunders (personal communication) on examination of the slides. Although the signal is readily identifiable under the microscope, it is difficult to detect on photographs and therefore, this result is not shown.

The 43 region has been extensively mapped at the genetic level with 27 complementation groups identified (Heitzler et al, 1993). This publication, along with the *Drosophila* Handbook (Lindsley and Zimm, 1992) and the *Drosophila* Database (Flybase - see references) were used to identify a number of deficiency lines, P-element insertions and candidate genes for further examination. A schematic representation of the 43 region is shown in figure 3.1. Deficiency lines obtained and used in this work are listed in Table 3.1.

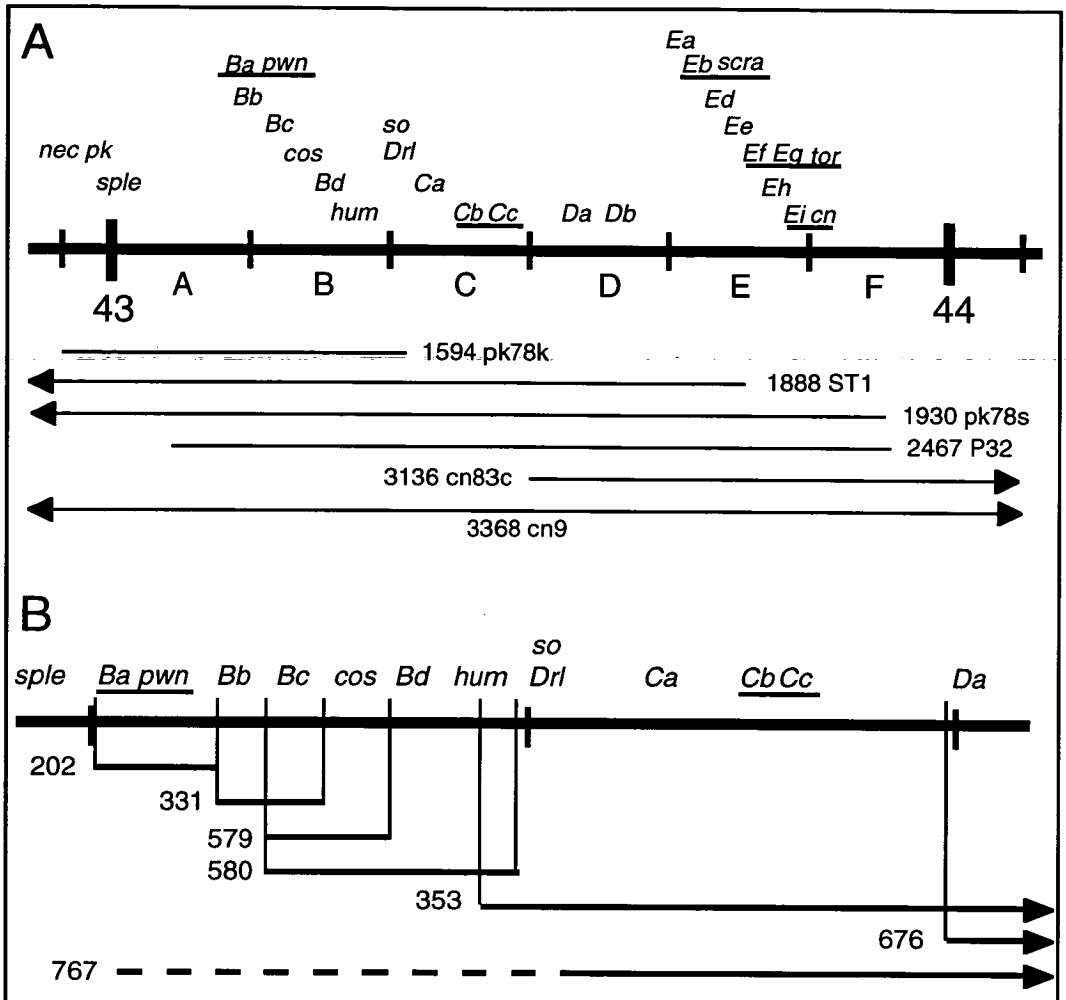


Figure 3.1 Schematic and complementation map of the 43 region modified from Heitzler et al (1993). Part A shows the overall map with genes shown above the main line and the large deficiencies shown below the main line. Complementation groups shown underlined are unordered with respect to each other. Part B shows a more detailed map of the 43BC region with further deficiencies indicated. Arrows indicate that the deficiency extends beyond the 43 region. The 2-767 deficiency is reported as mapping within 43B1-C1 and thus is represented by a dashed line in this region. See also Table 3.1.

3.3 Deficiencies place the genes at 43B

Deficiency breakpoints cause the alteration in size of restriction fragments which may be detected by Southern hybridisation as an additional band. A series of Southern blots were performed using chromosomal DNA from lines carrying chromosomes deficient for parts of the 43 region.

Table 3.1 Deficiency Lines Used in this Work

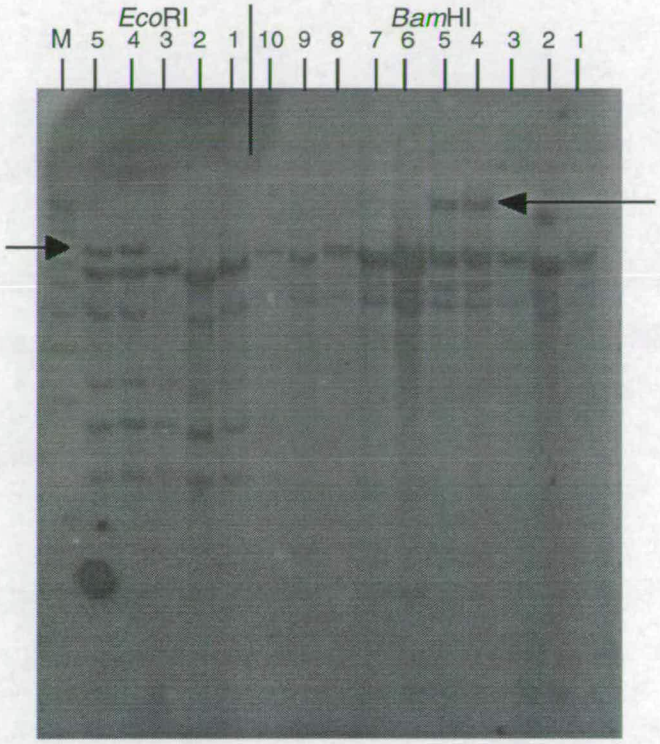
Line No	Name	Balancer	Region Lost
1594	Df(2R) pk78k	*CyO*	42E3; 43C3
1888	Df(2R) ST1	(*pr, cn*?) SM6a?	43B3-5; 43E1-8
1930	Df(2R) pk78s	not known	42C1-2; 43F5-8
2467	Df(2R) P32	possibly *cn*	43A3; 43F6
3136	Cn83c	SM5	43C5-D1; 44B5-C1
3368	Cn9	*Cy, Roi*	42E3; 44C3
2-202	Df(2R) pwn^5	In(2LR)O	[43B]; [43B]
2-331	Df(2R) Ew60	In(2LR)O	[43B]; [43B]
2-353	Df(2R) NCX9	In(2LR)O	43C3-7; 43F2
2-579	Df(2R) cos-2	In(2LR)O	[43B]; [43B]
2-580	Df(2R) cos-3	In(2LR)O	43B1; 43B1
2-676	hum^2 bw^D	In(2LR)O + Df(2R) cnS6	43C3-7; 43F2-8
2-767	Df(2R) cn84h	SM5	43B1-C1; 44A6-B1

Genomic DNA was prepared from these lines (7.6.1), restricted with *Bam*HI, *Eco*RI and *Xba*I and separated on 0.8% agarose/TBE gels. The DNA was transferred to Hybond-N and probed with labelled FIX-1A genomic DNA isolated from a *Not*I digest. The restriction endonuclease *Not*I cuts in the multiple cloning site of the λFIX bacteriophage allowing ready isolation of all inserted genomic DNA. Standard hybridisation and washing procedures were used. The results are shown in figure 3.2A.

Two lines, 2-579 and 2-580, showed additional bands on both *Bam*HI and *Eco*RI digests. The *Xba*I digest contained numerous bands possibly due to partial digestion and thus making identification of band shifts difficult. An additional *Eco*RI band was also observed in line 2467. However, no additional bands were seen in the *Bam*HI digest of line 2467 and, therefore, this result was probably due to a partial digestion.

Hybridisation of the same membranes, once stripped of the preceding probe, with the *didum* 2.2kb cDNA did not reveal the band shifts in lines 2-579 and 2-580 (figure 3.2C - note the autoradiographs shown use genomic DNA spanning the myosin V as probe, but the result was the same). However, hybridisation with a 1.8kb *Pst*I fragment from cDNA-34 did reveal these band shifts (figure 3.2B). Both 2-579 and 2-580 are shown to break at the same point proximal to the centromere on the Heitzler map, consistent with the identical band size seen on the Southern blots (see figure 3.1 part B). These two lines are deficient for the *43Bc* and *costa* genes with *43Bc* being proximal to *costa* (figure 3.1). On the basis of this result the gene encoding cDNA-34 was placed at *43Bc* and the myosin at *43Bb*. The possibility that the myosin was *costa*, and distal to *43Bc*, was ruled out as *costa* has recently been cloned and encodes a kinesin heavy chain-related protein (Sisson and Scott, 1996).

A



Lane	Line
1	202
2	331
3	353
4	579
5	580
6	676
7	767
8	1594
9	2467
10	3136

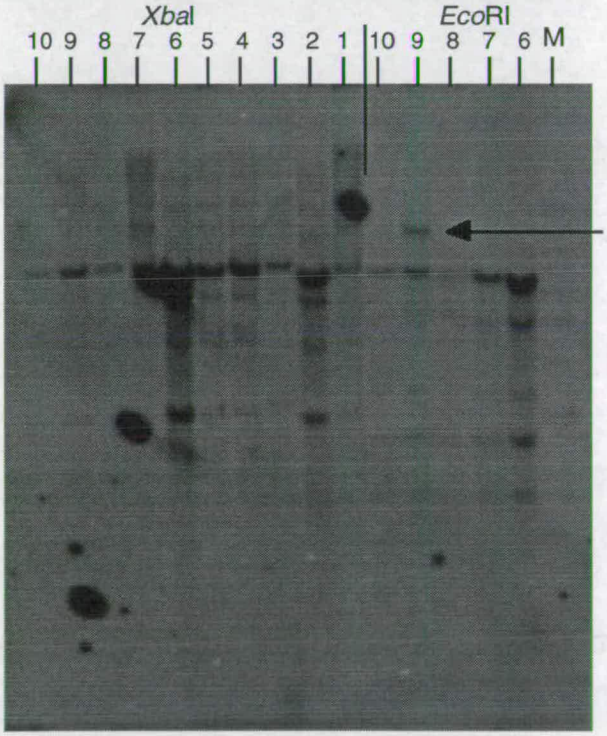
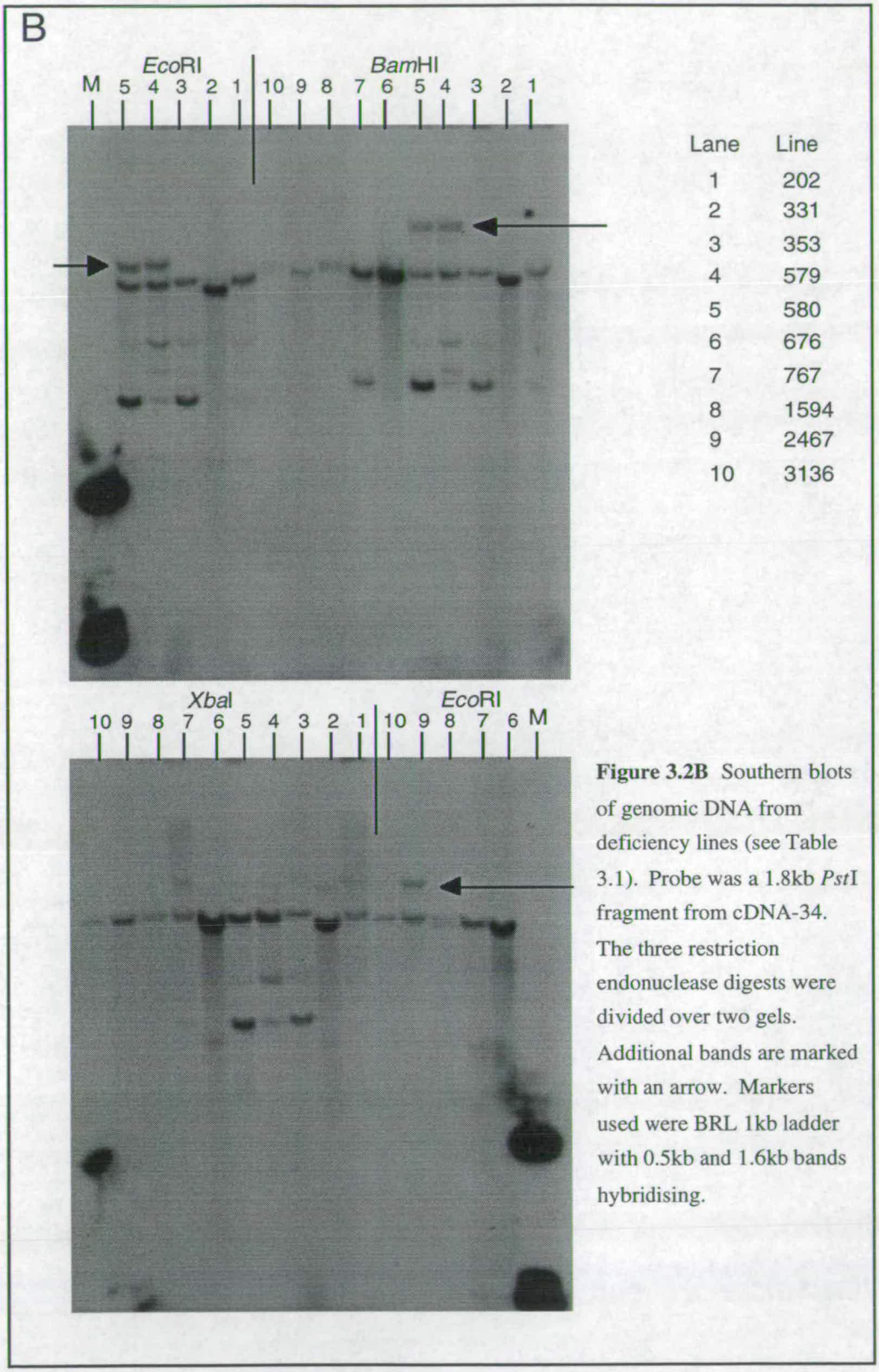


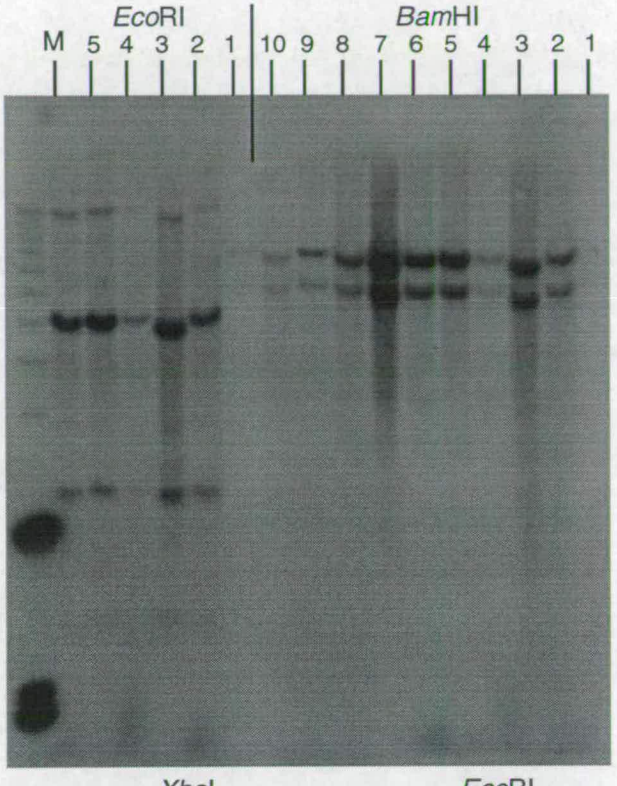
Figure 3.2A Southern blots of genomic DNA from deficiency lines (see Table 3.1). Probe was cloned genomic DNA from FIX-1A. The three restriction endonuclease digests were divided over two gels. Additional bands are marked with an arrow. Markers used were BRL 1kb ladder with 0.5kb and 1.6kb bands cross-hybridising.



Lane	Line
1	202
2	331
3	353
4	579
5	580
6	676
7	767
8	1594
9	2467
10	3136

Figure 3.2B Southern blots of genomic DNA from deficiency lines (see Table 3.1). Probe was a 1.8kb *Pst*I fragment from cDNA-34. The three restriction endonuclease digests were divided over two gels. Additional bands are marked with an arrow. Markers used were BRL 1kb ladder with 0.5kb and 1.6kb bands hybridising.

C



Lane	Line
1	202
2	331
3	353
4	579
5	580
6	676
7	767
8	1594
9	2467
10	3136

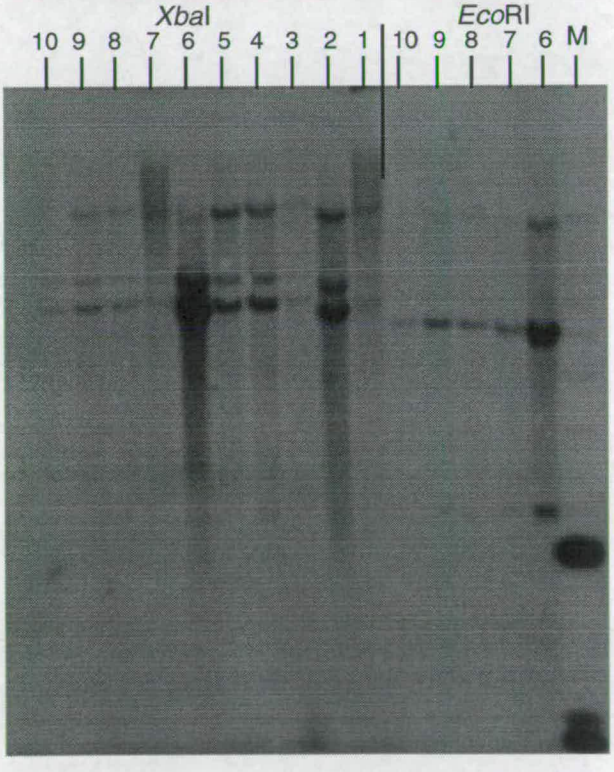


Figure 3.2C Southern blots of genomic DNA from deficiency lines (see Table 3.1). Probe was combined 6kb and 7kb *Xba*I fragments from P1-X6 and P1-X7 covering the *didum* gene. The three restriction endonuclease digests were divided over two gels. Markers used were BRL 1kb ladder.

3.4 A PCR Based Approach

To verify these results, a series of PCR reactions were carried out with deficiency lines. The hypothesis was that a cross between flies carrying Df/+ second chromosomes would produce one-quarter of the progeny as homozygotes with second chromosome deficiencies. These progeny would probably be embryonic lethal as the number of genes lost in the deficiencies was large (>10). DNA from these embryos could then be prepared and examined by PCR using gene specific primers. Absence of a gene specific band would indicate that the gene was located within the region covered by the deficiency. The two lines examined by this method were 1594 and 2467 (Table 3.1 and figure 3.1). Two controls were included. First, primers for a putative kinase gene located on the third chromosome (C MacDougall, personal communication) would indicate the presence of genomic DNA in the embryonic preparations. The second control consisted of primers designed against sequence reported for the *So* gene (Cheyette et al, 1994), a gene located distal to *43Bd* and *humilis* and proximal to *43Ca* (see figure 3.2) The *So* gene should not be present in the 2467 deficiency and appears close to the distal breakpoint in line 1594. The cross is shown in figure 3.3.

3.4.1 Method

The deficiency lines were balanced with a chromosome that results in a curly wing phenotype from the *CyO* gene. Initially, a cross was established between virgin females from the deficiency line and OregonR males. These flies were placed on vials of Staffan food at 18°C and changed to fresh food every 2-3 days. The progeny of this cross were either +/Df or +/balancer, with the flies carrying the balancer chromosome readily identified and removed on the basis of the curly wing phenotype. Vials were examined twice per day for newly eclosed flies with those showing a clear curly wing phenotype removed to separate vials from those showing wild-type straight wings. Flies exhibiting a Df/+ phenotype, that is, a straight wing, were separated on the basis of sex with vials containing females examined for the presence of larvae before being used in the subsequent cross. Any vial which contained larvae indicated that one or more females in that vial had mated before being collected. Flies not demonstrating a clear phenotype were placed individually into separate vials until their phenotype was obvious. When sufficient Df/+ flies had been collected (approximately 20 males and females), a

cross was established between these flies in a small cage with food supplied in the form of an Adh plate (7.4.2) supplemented with a small amount of yeast paste to encourage egg laying. The Adh plate was exchanged for a fresh plate every 24 hours with the used plate incubated at 25°C for a further 24-36 hours to allow all viable embryos to hatch. The hatched larvae were collected and moved to vials to prevent excessive 'churning' of the food and loss of the unhatched embryos. The larvae were also a valuable control as any which developed into a fly with a curly wing phenotype would indicate that the balancer chromosome was still present.

Unhatched embryos were collected and rinsed with water to remove as much food as possible then stored in Ringer's solution at 4°C overnight. Genomic DNA was prepared using the same method as that used for adult flies, see materials and methods, section 7.6.1.

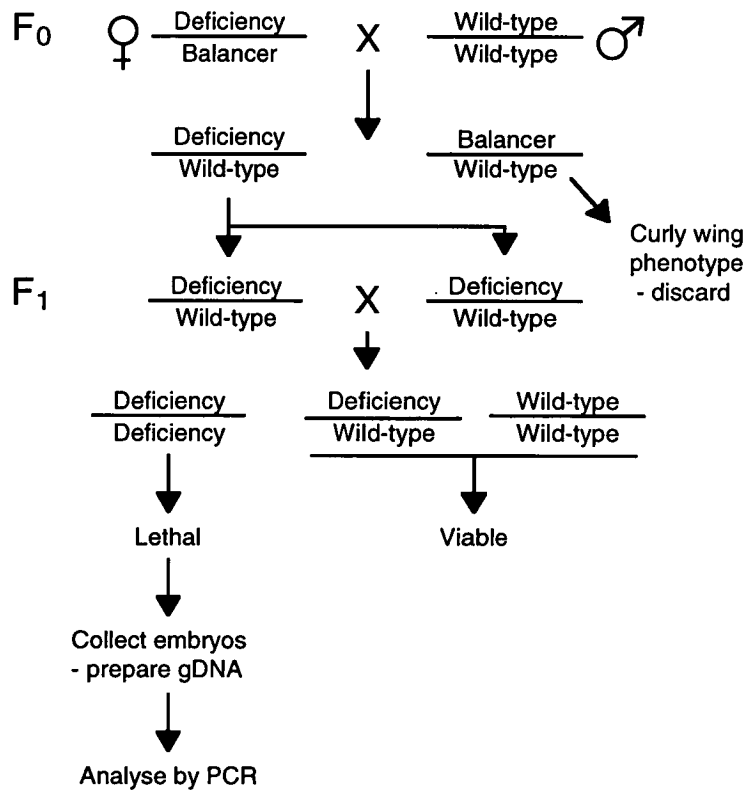


Figure 3.3 Cross between Deficiency lines and OregonR wild type flies to remove the balancer chromosome (F0). Progeny from the F1 generation exhibiting straight wings were then crossed among themselves to give viable flies carrying a wild type chromosome or a homozygote deficiency resulting in embryonic lethality.

3.4.2 PCR amplification

A standard 50µl PCR was set up using Promega 10x buffer and a final Mg^{2+} concentration of 2mM. The primers used were 1806 and 2800 for the myosin gene (see figure 2.6), so1 and so4 for the *So* gene. For the kinase gene on the third chromosome, primers 1240 and 1241 which produce a gene specific product of 0.8kb were provided (C MacDougall, personal communication).

The PCR conditions used were:
94°C for 2 minutes and hold at 80°C for addition of Taq DNA Polymerase,
thermal cycling for 35 rounds at 94°C for 30 seconds
55°C for 30 seconds
72°C for 2 minutes
final 'polish' step of 72°C for 10 minutes.

The results are presented in figure 3.4 and show myosin gene specific bands present in the OregonR and 1594 lines but at a much reduced concentration in line 2467. The *So* gene primers produced strong bands of about 0.9kb in both the OregonR and 1594 lines but also a number of spurious bands.

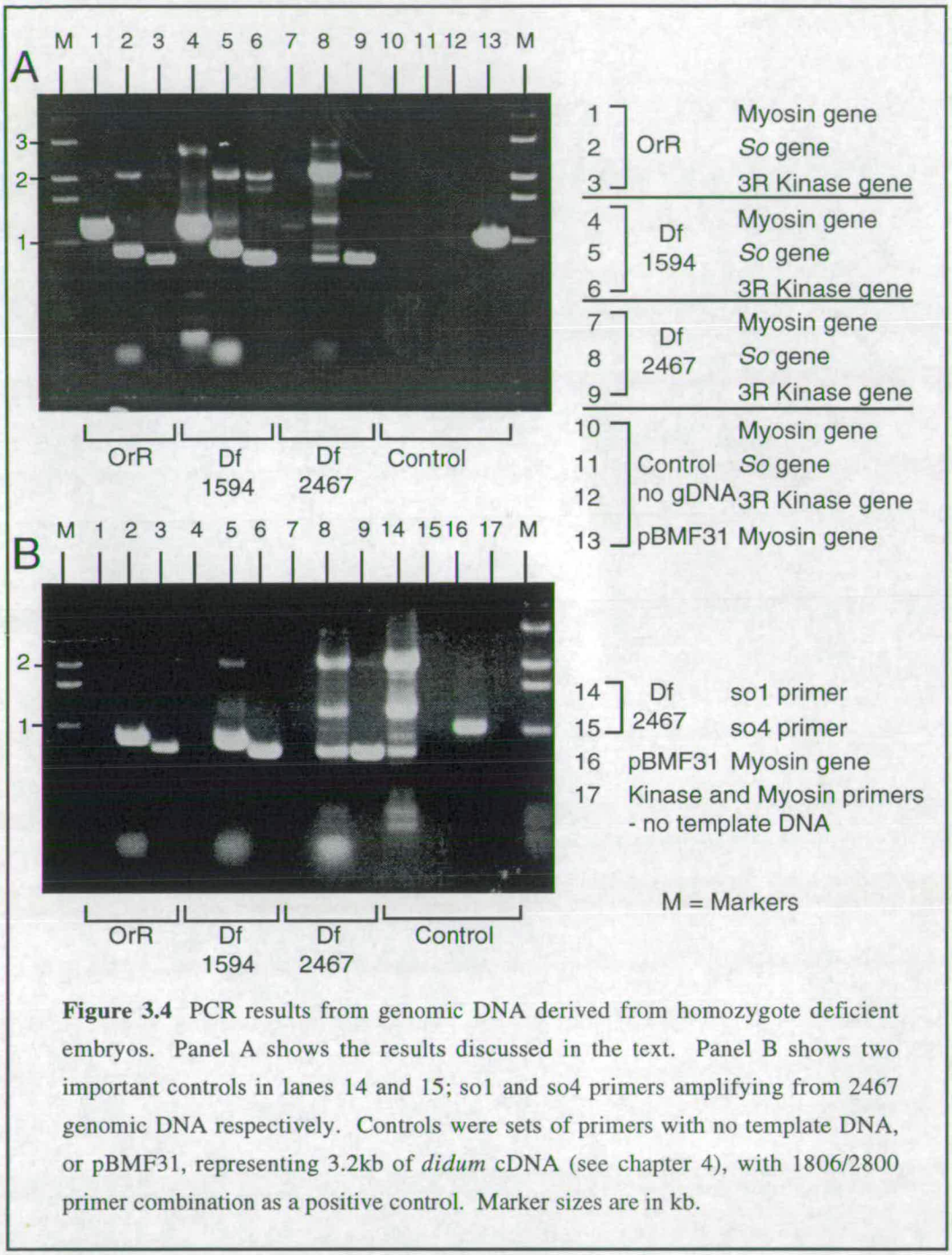


Figure 3.4 PCR results from genomic DNA derived from homozygote deficient embryos. Panel A shows the results discussed in the text. Panel B shows two important controls in lanes 14 and 15; so1 and so4 primers amplifying from 2467 genomic DNA respectively. Controls were sets of primers with no template DNA, or pBMF31, representing 3.2kb of *didum* cDNA (see chapter 4), with 1806/2800 primer combination as a positive control. Marker sizes are in kb.

The *So* primers produced numerous bands in the 2467 line but the 0.9kb band was much reduced in concentration. According to the genetic maps the *So* gene should be lost in a 2467 homozygote deficient embryo. These results can be interpreted as showing both the myosin gene and the *So* gene falling within deficiency 2467. The presence of a gene specific band at reduced levels can be explained by a small amount of wild type DNA being present from wild type embryos that failed to

hatch. Ashburner (1989) suggests that normal embryo mortality in wild type lines is in the order of 5-10%, supporting this conclusion. The presence of numerous additional bands in the *So* lanes can be explained by non-specific priming arising from each individual primer. A previous experiment utilising identical conditions but with each *So* primer used individually with 2467 genomic DNA resulted in numerous bands as shown in figure 3.4. It is reasonable to speculate that DNA from homozygous 2467 embryos is more likely to produce spurious bands as the designed gene target has been lost.

The result from the 1594 line is more difficult to interpret. The distal breakpoint for 1594 is shown to be between *So* and *43Ca* on the Heitzler map (figure 3.1). However it is possible that the 1594 breakpoint occurs within the *So* gene removing essential *So* elements but leaving the two priming sites intact. This suggestion would enable the *So* primers to produce a gene specific product but would still result in deficiency 1594 failing to complement *So* mutant alleles in genetic tests. Another possibility is that the 1594 line had lost the deficiency but this was unlikely as the cross of 1594/+ flies produced embryonic lethals.

3.5 Cytological Position of P1 Supports 43C Location

The above PCR results indicate that the myosin gene is located in the 43C region. This result is supported by the reported position of 43C for P1 DS00574 (reported in Flybase) which was used to isolate 5' genomic sequence (see section 2.2.12 in Chapter 2). However, these results contradicted those obtained for the deficiency Southern blots. Additional results have clarified this situation

3.6 Mapping of FIX-1A Shows It is Composed of Two Independent Recombinant Phage.

As stated in Chapter 2, a definitive restriction map for the cloned genomic DNA from FIX-1A could not be generated from the restriction endonuclease digests because of the number of fragments generated and size inconsistencies. Table 3.3 lists the fragments identified from digests of FIX-1A with the total amount of genomic DNA represented being 18-23kb depending on the digest examined. Figure 3.6 shows schematic maps of the subcloned genomic fragments and a restriction digest of FIX-1A DNA. In order to clarify the physical map, a series of

cross hybridisations with subcloned genomic fragments from FIX-1A were performed.

Table 3.3 Restriction Fragments Identified from Digests of FIX-1A

*Bam*HI	*Bam*HI/ *Not*I	*Eco*RI	*Eco*RI/ *Not*I	*Xba*I
8	8-9	8 (pE7)	8	6
	6-7	5 (pE5)	5	3.5
	5.5	2.9 (pE3)	2.9	3 (pX3)
		2.1 (p2.1)	2.1	2.8 (pX2.8)
		1 (pE1)	1	2.2
			2.5	
			1.6	
Total		19 kb	23 kb	17.5 kb

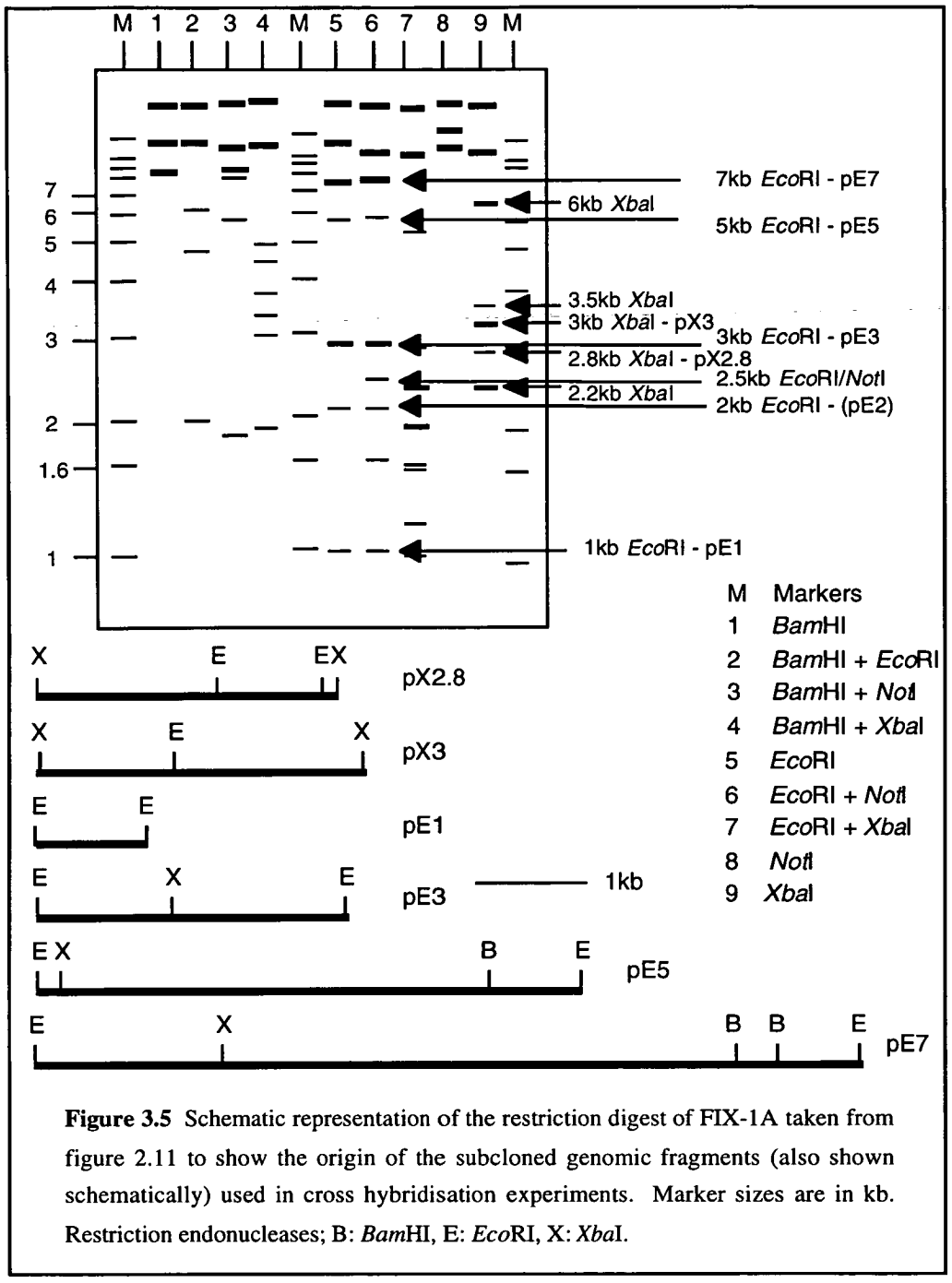


Figure 3.5 Schematic representation of the restriction digest of FIX-1A taken from figure 2.11 to show the origin of the subcloned genomic fragments (also shown schematically) used in cross hybridisation experiments. Marker sizes are in kb. Restriction endonucleases; B: *Bam*HI, E: *Eco*RI, X: *Xba*I.

The 5kb genomic fragment from pE5 (contains sequences from the 3'UTR of the unconventional myosin gene) allowed the isolation of a ~6kb *Xba*I fragment (named pP1-X6) and a ~8kb *Bam*HI fragment (named pP1-B8) from the P1 DS00574 mini-libraries. The relative positions of these fragments is shown in figure 3.6

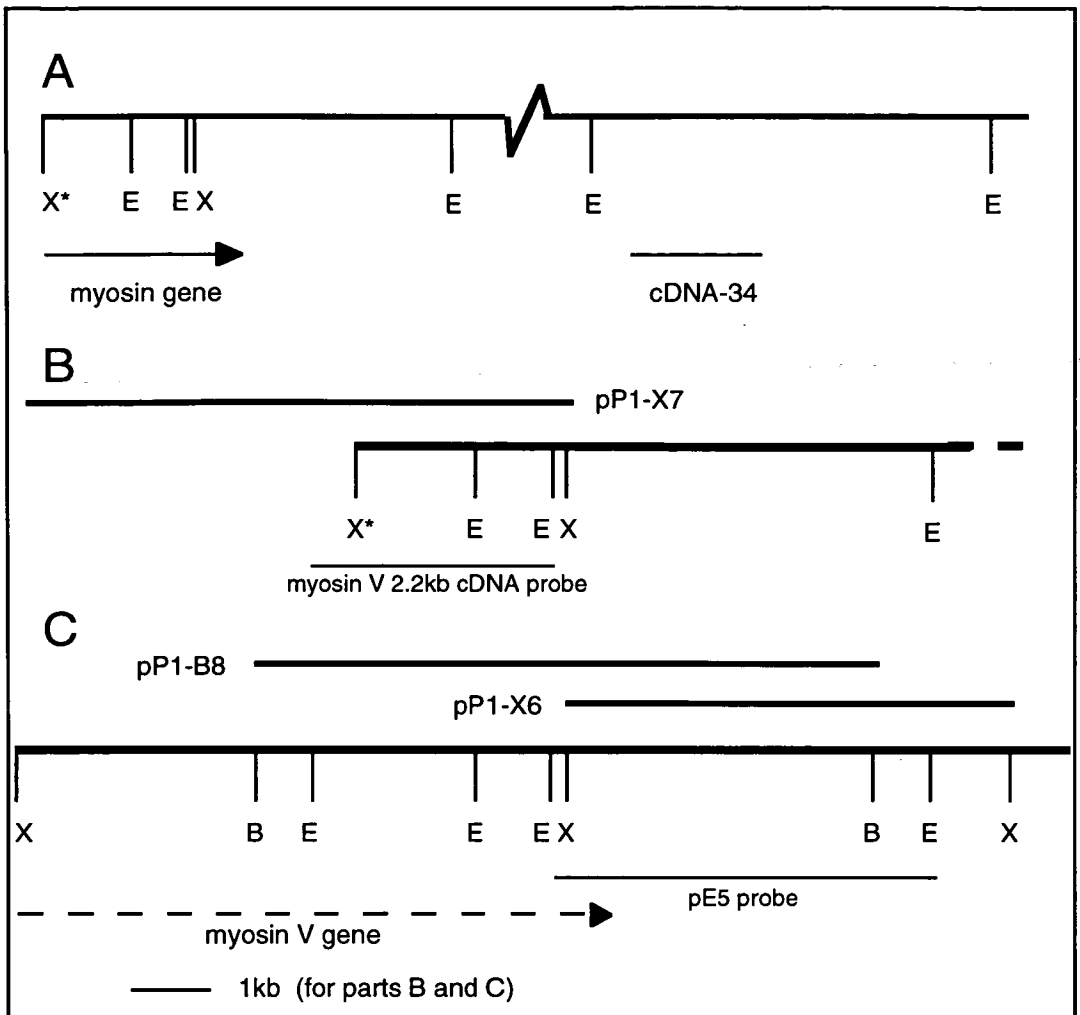


Figure 3.6 Part A) An initial map showing that the myosin V and cDNA-34 genes were thought to reside on a contiguous section of genomic DNA with the map of the intergenic region undefined - a hypothesis that was disproved by the cross hybridisation mapping experiments (figures 3.7 to 3.10). Part B shows the position of the 7kb *Xba*I genomic DNA insert from pP1-X7 relative to the myosin V gene. Part C shows two further cloned genomic fragments, a 6kb *Xba*I fragments (pP1-X6) and an 8kb *Bam*HI fragment (pP1-B8) obtained by screening P1 DS00574 mini-libraries with the 5kb *Eco*RI insert from pE5.

Use of the 6kb *Xba*I fragment (pP1-X6) identified overlapping 2.1kb*Eco*RI and 6.5kb *Bam*HI fragments and resolved the map 3' of the myosin gene as shown in figure 3.7. Cross hybridisation of fragments from pE3, pE7 and pX3 showed that these were related and encompassed the cDNA-34 gene, see figures 3.8 and 3.9. Thus for FIX-1A to be a contiguous section of genomic DNA, it was expected that the pE3 fragment would show some overlap with the 6.5kb *Bam*HI fragment identified from genomic DNA, FIX-2 and FIX-4 Southern blots. However, no

cross hybridisation was observed. The mapped DNA from FIX-1A surrounding the myosin gene represented approximately 10.5kb while that surrounding cDNA-34 represented approximately 11kb with 2 restriction fragments visible on the original agarose gel unaccounted for; a 3.5kb *Xba*I fragment and a 2.5kb *Eco*RI/*Not*I fragment. Thus the total amount of DNA from FIX-1A was at least 25kb (11kb +10.5kb +3.5kb = 25kb) exceeding the maximum of 23kb of DNA that can be cloned into the λFIX vector (Sambrook et al, 1989). These results show that FIX-1A is a combination of two independent recombinant bacteriophage with *didum* 3' sequences residing on one phage and cDNA-34 being located on the second phage. Experiments have yet to be undertaken to separate the two recombinant bacteriophage.

3.7 The az2 Gene

A screen of the ovarian cDNA λZAP library with the 5kb *Eco*RI pE5 fragment to isolate further cDNAs representing the 3'UTR of the *didum* gene resulted in the discovery of a novel cDNA which has been given the working name az2 (A McCormack, personal communication). Sequence analysis indicates that the az2 gene encodes a zinc-finger like protein and the transcript size is in the order of 1.7kb (A McCormack, personal communication). The az2 gene appears to map entirely within the 5kb *Eco*RI fragment from pE5 with this finding supporting the conclusion that *didum* and cDNA-34 genes cannot be *43Bb* and *43Bc*, as originally proposed from the deficiency Southern blot results.

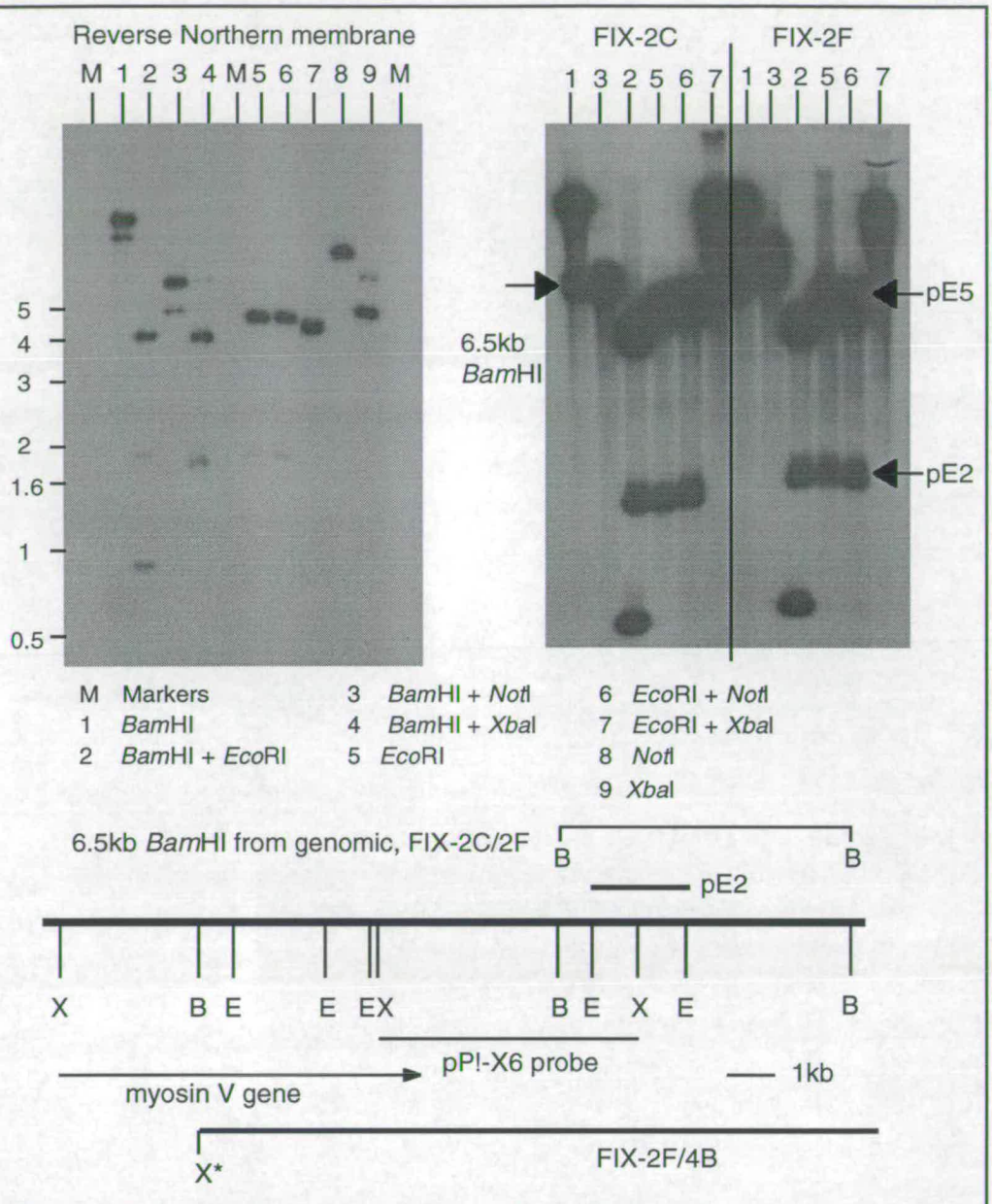


Figure 3.7 Reverse Northern membrane and FIX-2C/2F Southern blot hybridised with pP1-X6. A 2kb *Eco*RI fragment (subcloned as pE2 from FIX-4A) and 6.5kb *Bam*HI fragment (on FIX-2C/2F) were found to overlap (marked with arrows). Marker sizes are in kb and relate only to the Reverse Northern.

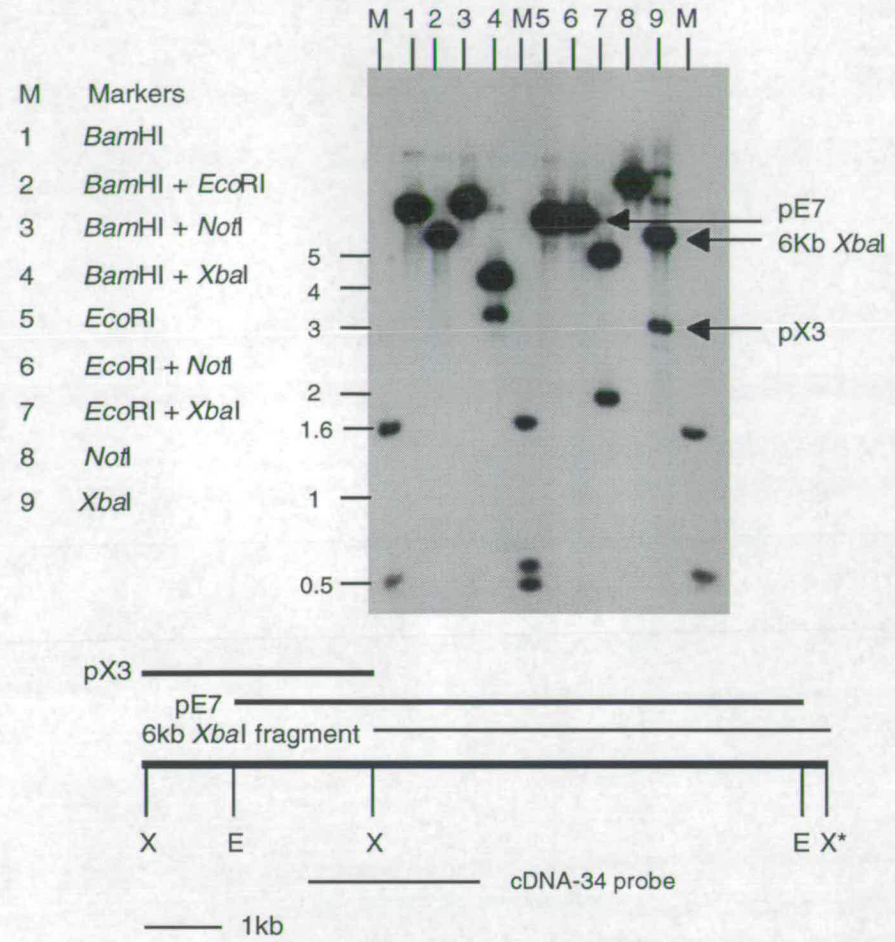


Figure 3.8 Hybridisation of the Reverse Northern membrane with a *Pst*I fragment from cDNA-34. Overlapping fragments identified were 3kb *Xba*I (pX3), 6kb *Xba*I and 7kb *Eco*RI (pE7). The relationship between pX3 and pE7 was confirmed from their restriction maps see also figure 3.5. Marker sizes are in kb.

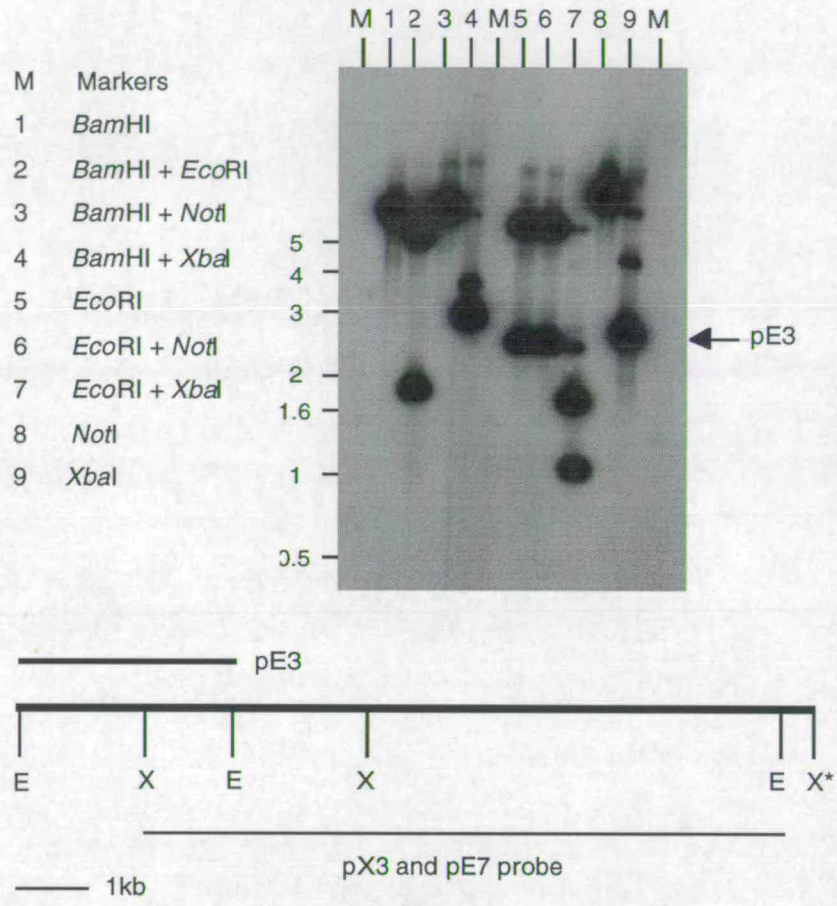


Figure 3.9 Refining the genomic map surrounding cDNA-34. Hybridisation to the Reverse Northern membrane with a combined pX3/pE7 probe identified an overlap with the pE3 fragment. Marker sizes are in kb.

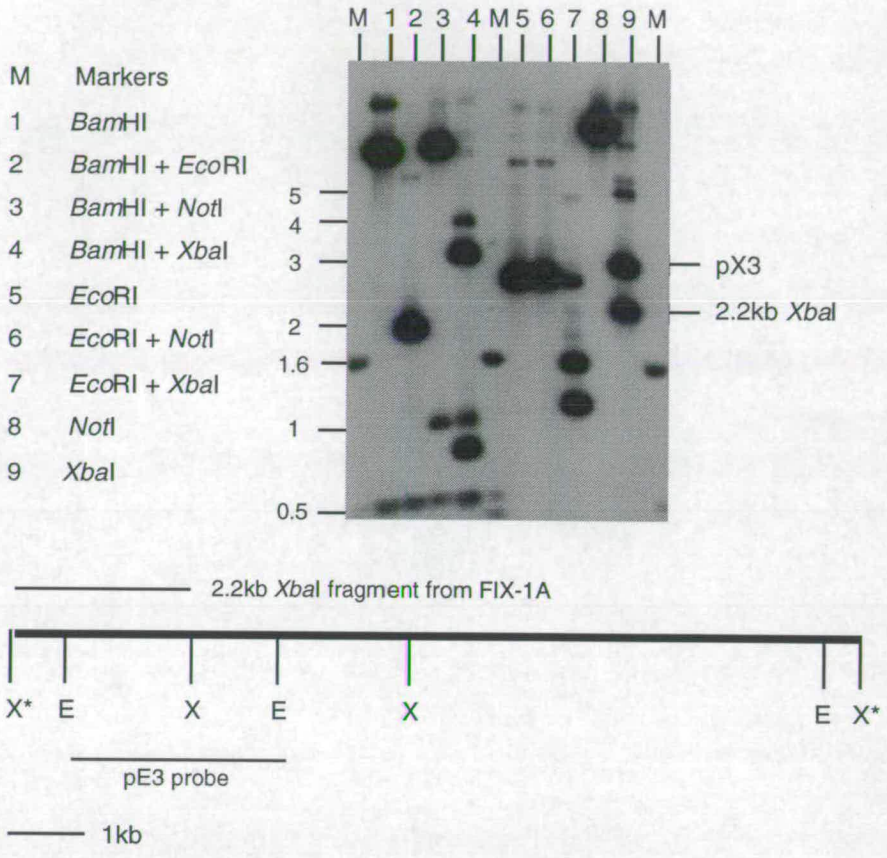


Figure 3.10 Hybridisation of the Reverse Northern membrane with pE3. An overlap with a 2.2kb *Xba*I fragment was observed in addition to the expected overlap with pX3. The position of the *Bam*HI sites was also confirmed with 0.5kb, 1.9kb and ~8kb fragments identified. Marker sizes are in kb. If FIX-1A was a contigous piece of genomic DNA then pE3 was expected to hybridise to an additional ~6.5kb *Bam*HI fragment. Marker sizes are in kb.

3.8 Summary

An *in situ* hybridisation to polytene chromosomes using *didum* cDNA as a probe localises the *didum* gene to the 43BC region. This region has been extensively characterised genetically with a high probability that all complementation groups within this region being identified (Heitzler et al, 1993). Analysis of *Drosophila* lines carrying deficiencies for this region suggested that *didum* might be *43Bb* as the cDNA-34 gene showed band shifts in lines deficient for *43Bc* and *costa*. However, a PCR approach on genomic DNA from homozygote deficient embryos suggested the *didum* gene was in 43C, a result supported by cytological data placing P1 DS00574 in 43C (from Flybase).

Extensive cross hybridisation experiments to resolve the restriction map of FIX-1A cloned genomic DNA revealed that FIX-1A consisted of two independent smaller recombinant bacteriophage. This finding placed further doubt on the deficiency Southern blot result. The recent identification of a new gene, az2, adjacent to the *didum* gene also refutes the original *43Bb-43Bc* hypothesis.

It is most likely that the deficiency Southern blot result was due to a polymorphism in the chromosomes of lines 579 and 580 with the same chromosome being used to generate the deficiencies in these two lines. The 43C region has three genes mapped within it; *43Ca*, *43Cb* and *43Cc*, with *43Cb* and *43Cc* unordered with respect to each other (see figure 3.1). It is tempting to speculate that *didum* and az2 will be either of *43Cb* and *43Cc* because these genes were unable to be separated on the Heitzler complementation map and map close together physically.

Figure 3.11 summarises the physical maps of *didum*, cDNA-34 and az2 genes.

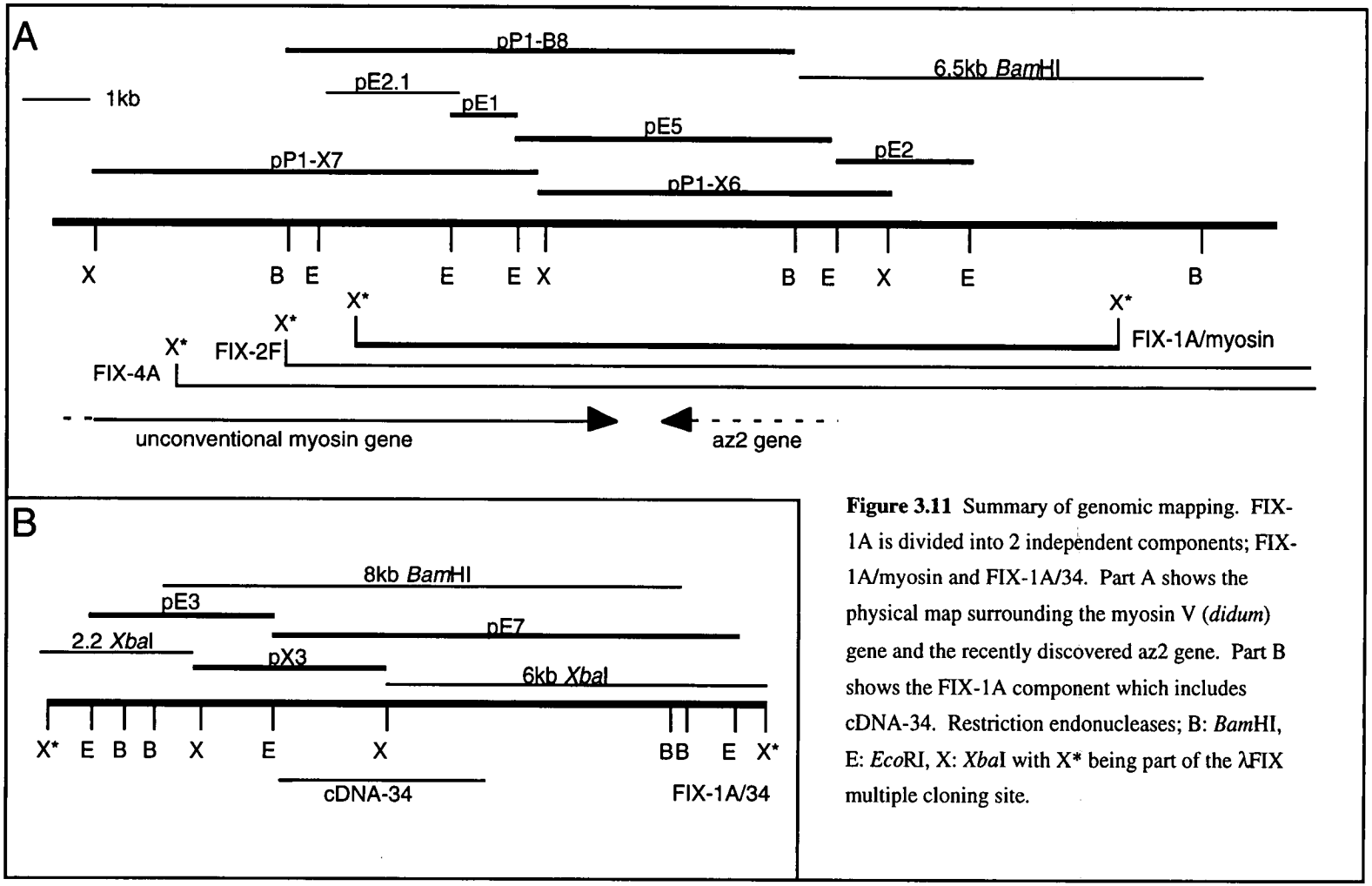


Figure 3.11 Summary of genomic mapping. FIX-1A is divided into 2 independent components; FIX-1A/myosin and FIX-1A/34. Part A shows the physical map surrounding the myosin V (*didum*) gene and the recently discovered az2 gene. Part B shows the FIX-1A component which includes cDNA-34. Restriction endonucleases; B: *Bam*HI, E: *Eco*RI, X: *Xba*I with X* being part of the λFIX multiple cloning site.

Chapter 4

Expression in *Escherichia coli* and Production of Antibodies

4.1 Introduction

An integral part of studying the function of a gene involved in oogenesis is to observe the distribution of the protein product in egg chambers. The use of current recombinant techniques allows the expression of all or part of the cDNA in *Escherichia coli*, the recovery of the recombinant protein via affinity chromatography and the use of this recovered protein to raise antibodies in an appropriate animal. This chapter describes the construction of several recombinant plasmids for expressing parts of the Class V myosin. Initially, an expression plasmid was constructed using a subcloned fragment from the first cDNA isolated, NMC7 (see chapter 2). However, expression was not obtained from this plasmid. An improved expression plasmid was constructed when further cDNAs were isolated and this plasmid proved successful for protein expression.

4.2 Expression Plasmids pBMF1 and pBMF2

The first set of plasmids prepared for expressing recombinant protein utilised the *Eco*RI fragment from pNMC7a (see figure 2.2). This fragment codes for about 280 amino acids but lacks the very last 40 amino acids from the carboxyl terminus.

The pRSET (Invitrogen) vector system was chosen to construct an expression plasmid. This system uses a fusion between the protein of interest and a small peptide at the amino terminus which contains a 6 histidine motif. This histidine motif chelates Ni^{+} ions and allows affinity purification of the recombinant protein by Ni-agarose chromatography. Transcription is driven from a T7 promoter.

An in-frame fusion was made between the *Eco*RI fragment from pNMC7a and the expression vector pRSET-A. The vector was digested with *Eco*RI and dephosphorylated as described in section 7.5.2 while the *Eco*RI fragment was isolated from pNMC7a by *Eco*RI digestion and purification from agarose gel by Qiaex (7.5.4.4). A standard ligation reaction was prepared (7.5.1) and used to transform *E coli* XL1-Blue cells (7.3.1). Plasmids carrying the pNMC7a insert were identified by colony hybridisation. Small scale DNA preparations were made and fragments ligated in the correct orientation were identified using the restriction endonuclease *Pst*I which cuts once in the vector and once in the pNMC7a derived

fragment. The correct orientation was identified by a *Pst*I band of 670bp on an agarose gel while the opposite orientation gave a 200bp band (results not shown).

Both orientations were selected with the correct orientation named pBMF1 and the reverse pBMF2. These plasmids were used to transform the *E coli* strain BL21 DE3, a strain which supplies T7 RNA polymerase inducible by IPTG from a bacteriophage λ lysogen.

A number of experiments were conducted to determine if an approximately 35kDa protein was being expressed (280 amino acids + 40 amino acids from the vector fusion to give 320 amino acid residues, at an average molecular weight of 110 per amino acid residue = 35kDa). Whole cell and soluble cell fractions were examined by discontinuous SDS-polyacrylamide electrophoresis (SDS-PAGE). However expression was not observed (results not shown).

4.3 Construction of pBMF31, pBMF32 and pBMF33

Prior to constructing an improved expression plasmid, it was necessary to combine the 2.2kb cDNA (pBMF10 = ZAP-A etc, see Chapter 2) with pNMC7b to produce a plasmid that contained all of the cDNA obtained to date. This plasmid construction was readily accomplished by inserting the 2.2kb cDNA from pBMF10 into the *Eco*RI site of pNMC7b to create pBMF31. A diagrammatic representation of this construction is given in figure 4.1.

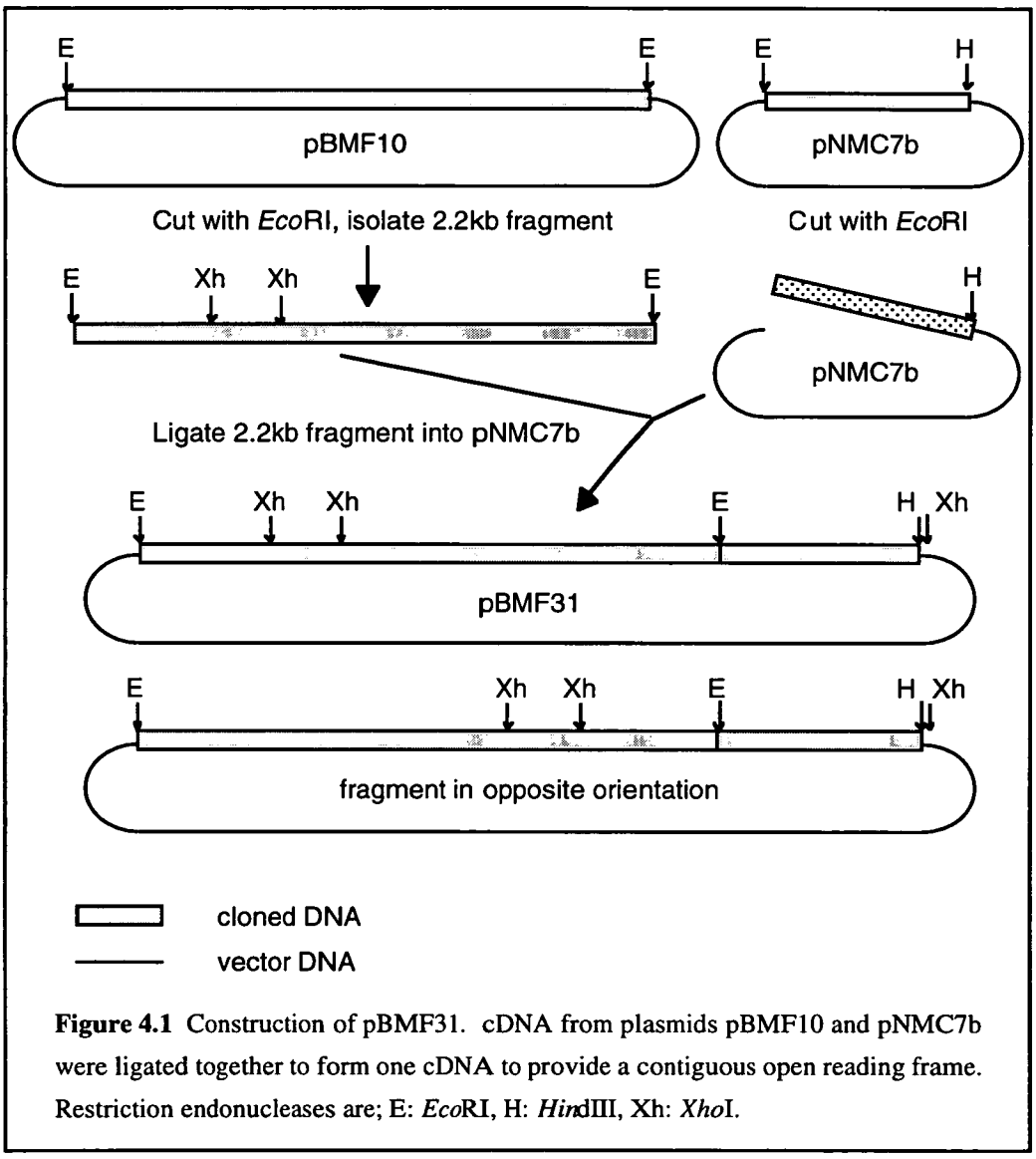


Figure 4.1 Construction of pBMF31. cDNA from plasmids pBMF10 and pNMC7b were ligated together to form one cDNA to provide a contiguous open reading frame. Restriction endonucleases are; E: *Eco*RI, H: *Hin*dIII, Xh: *Xho*I.

A medium scale preparation (7.6.2.2) of pBMF10 was made and 10µg digested with *Eco*RI. The digest was separated on an agarose gel and the 2.2kb cDNA fragment identified and excised with a scalpel. The DNA was recovered by Qiaex with 10% of the final volume run on a gel to check recovery. A medium scale preparation of pNMC7b was made and also digested with *Eco*RI, followed by treatment with CIAP to prevent self ligation. Phenol/chloroform extraction and ethanol precipitation was used to remove the CIAP. A series of ligation reactions were set up in final volumes of 10µl and competent *E coli* XL1 cells were transformed with 5µl of the ligation reaction.

Since the cDNA from pBMF10 is a 2.2kb *Eco*RI fragment, insertion into pNMC7b in both orientations was feasible. In order to determine the desired in-frame orientation, small scale DNA preparations were made from 8 colonies from the transformation and digested with *Xho*I An insertion of the 2.2kb fragment in the desired orientation would produce *Xho*I fragments of 0.3kb, 2.3kb and 3.5kb while the opposite orientation would produce fragments of 0.3kb, 1.3kb and 4.5kb (figure 4.1). Three of the 8 small scale preparations showed the correct orientation with one selected for further work and named pBMF31.

4.3.1 Preparation of a Fragment for Use in an Expression Vector

At the time of preparing the expression construct the DNA sequence at the 5' end of pBMF10 (5' most sequence = nt 1) had not been fully resolved and therefore, the extent of the open reading frame had not been confirmed. A continuous open reading frame had been identified from nt 415 to the stop codon at nt 2448. An *Eco*RV site had been identified at nt 457 and cleaved in-frame. This site, in combination with the *Dra*I site, which coincided with the stop codon, produced a blunt-ended fragment of 1986bp suitable for cloning into an expression vector.

The pBluescript vector contains 3 *Dra*I sites at positions 1912, 1931 and 2623 (Stratagene Inc). A digest of pBMF31 with *Eco*RV and *Dra*I produced a total of 5 bands of sizes 17bp, ~680bp, ~1660bp, ~1720 and 1986bp, with the largest being the required fragment. However, this 1986bp fragment could not easily be resolved from the other larger fragments by gel electrophoresis. These interfering fragments were readily reduced in size by further digestion with the restriction

endonucleases *Bam*HI and *Kpn*I, both of which cut in the multiple cloning site of the pBluescript vector. The 4 way digest was done as follows:

> Approximately 25µg of pBMF31 was double digested with *Eco*RV and *Dra*I in a volume of 100µl in BRL React2 buffer. These restriction endonucleases were heat inactivated at 65°C for 15 minutes, then 1µl of 5M NaCl was added to raise the NaCl concentration of the reaction mix from 50mM to 100mM. The reaction was then digested with *Bam*HI for several hours. Addition of 5µl 2M KCl and 2.5µl 2M Tris-HCl pH 7.2 in a final volume of 200µl produced conditions suitable for digestion with *Kpn*I. The results are shown in figure 4.2.

This *Eco*RV-*Dra*I fragment was then cloned into pRSET-C at the *Pvu*II site to produce and in-frame fusion with the amino terminal histidine tag sequence using cloning procedures described above. The ligation reactions were used to transform *E coli* BL21 DE3. Again, insertion of the *Eco*RV-*Dra*I fragment in either orientation was possible. Digestion with *Eco*RI was used to determine the orientation with the correct orientation being called pBMF33 and the reverse pBMF32.

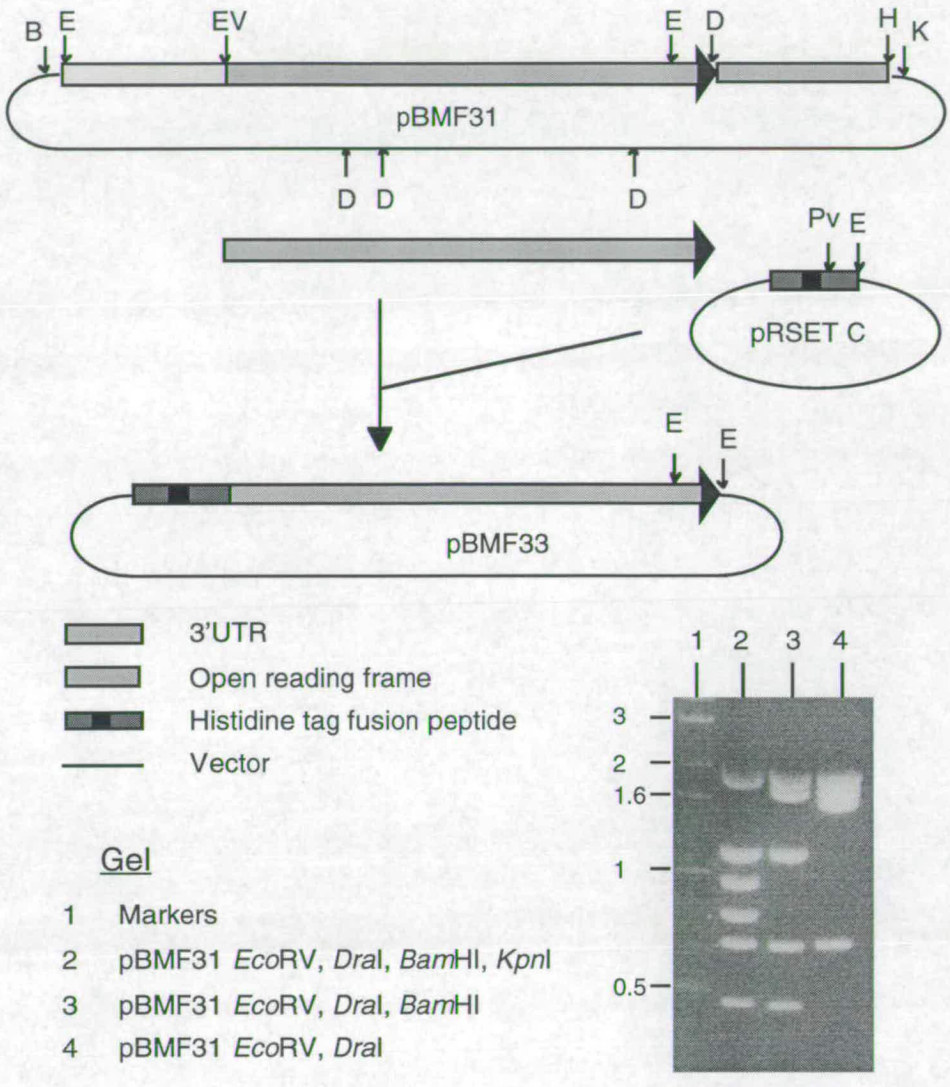
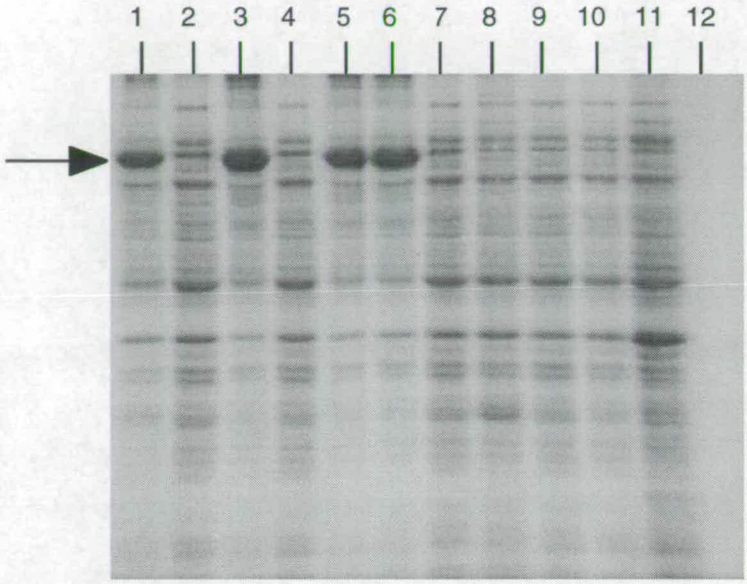


Figure 4.2 Construction of expression plasmid pBMF33 (and pBMF32 - not shown). An *Eco*RV-*Dra*I fragment was isolated from pBMF31 (see gel) and cloned into pRSET-C at the *Pvu*II site. The orientation of the inserted DNA was determined by digestion with *Eco*RI. Restriction endonucleases; B: *Bam*HI, D: *Dra*I, E: *Eco*RI, EV: *Eco*RV, H: *Hin*dIII, K: *Kpn*I. Marker sizes are in kb.

4.4 Expression From Recombinant Plasmid pBMF33

Three clones of pBMF33 were selected for testing for recombinant protein expression along with pBMF32 and pRSET-C as controls. Initially 5ml cultures in LB-broth with 100µg/ml carbenicillin as antibiotic (carbenicillin is a more stable β-lactam than ampicillin) were prepared and grown to an OD_{650} of approximately 1.0-1.3 as judged by eye, apart from the pRSET-C culture which had only reached an OD_{650} of ~0.5. These cultures were stored at 4°C overnight followed by centrifugation of 1.5ml of each (3ml of pRSET-C) to collect the cells. The cells were then resuspended in 1ml of T-broth then added to a 50ml solution of T-broth in a 500ml flask with carbenicillin at 100µg/ml. The cultures were incubated at 37°C for about 3 hours whereupon 10ml samples were drawn and placed on ice, apart from the pRSET-C culture which was incubated for a further 1.5 hours as it continued to show a lower amount of growth. Each culture was induced with a 200mM solution of IPTG to give a final concentration of 0.5mM. Samples of 10ml were taken after 2 hours and 4.5 hours of further incubation and placed on ice. The pRSET-C culture was sampled at 3 hours post induction. All cultures were stopped at 6 hours post induction (about 4.5 hours for pRSET-C).

An aliquot from each sample was used to determine the cell density at OD_{650}. Cells from 6ml of each sample were collected by centrifugation and resuspended in 5ml buffer A (50mM Tris-HCl pH 8.0, 2mM EDTA), then centrifuged once more. The supernatant was removed and the pellets stored frozen at -20°C. The pellets were thawed on ice then resuspended in 1ml of buffer A with a 100µl aliquot removed from each sample and 100µl of 2xSDS-PAGE loading buffer added (7.12) followed by boiling for 3-5 minutes. A 12.5%/4% discontinuous SDS-PAGE was prepared and the samples loaded. To produce an even loading, the amount loaded was adjusted based on the OD_{650} readings, that is, less sample was loaded from cultures which had higher OD_{650} readings. The SDS-PAGE was run overnight at a constant 60V then stained with the Pro-Blue system (7.12). The results, shown in figure 4.3, demonstrate that all 3 clones of pBMF33 produced recombinant protein. Induced samples produce a strong band at approximately 80kDA consistent with the expected size from a fusion protein of 702 amino acids.



1 pBMF33-6 Induced 4.5 hours
2 pBMF33-6 Uninduced
3 pBMF33-3 Induced 4.5 hours
4 pBMF33-3 Uninduced
5 pBMF33-2 Induced 4.5 hours
6 pBMF33-2 Induced 2 hours
7 pBMF33-2 Uninduced
8 pBMF32 Induced 4.5 hours
9 pBMF32 Uninduced
10 pRSET-C Induced 4.5 hours
11 pRSET-C Uninduced
12 Markers

Figure 4.3 SDS-PAGE of the initial expression experiment. The expressed recombinant protein is marked with an arrow. Space did not permit the inclusion of induced 2 hour samples from clones pBMF33-3 and pBMF33-6. The markers were degraded.

4.4.1 Expressed Recombinant Protein is Insoluble

An experiment to obtain affinity purified protein by Ni-agarose chromatography demonstrated that the protein was produced in an insoluble form. Soluble and insoluble fractions were prepared from the remaining culture from clone pBMF33-2 from the initial expression experiment. Cells were collected by centrifugation and resuspended in cold buffer X at half the original volume (10mM Tris-HCl pH 7.9, 1mM imidazole pH 7.9, 500mM NaCl, 10%v/v glycerol, 0.1%v/v Tween 20). The cells were collected once more by centrifugation and resuspended in 5ml buffer X. Lysozyme was added to give a final concentration of 1mg/ml and the suspension incubated on ice for 30 minutes. Lysis was accomplished by

sonication (MSE sonicator) using 30 second bursts with periods of incubation on ice between sonication to prevent the suspension warming.

Soluble and insoluble fractions were separated by centrifugation at 10 000rpm for 15 minutes in a Sorvall RC-5B (SS34 rotor). The soluble fraction (supernatant) was transferred to a fresh tube while the insoluble fraction was washed once with 2.5ml of buffer X with this wash added to the soluble fraction. Samples were then run on a 10%/4% SDS-PAGE at 110V for 5.5 hours with Pro-Blue staining as before. The results, presented in figure 4.4 show that the recombinant protein is present in the insoluble fraction.

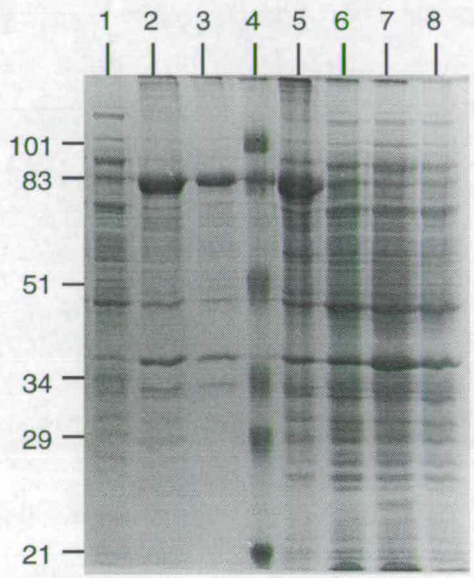


Figure 4.4 SDS-PAGE showing that the recombinant protein is a component of the insoluble fraction of the cell extract. Marker sizes are in kDa.

4.5 Preparation of Protein for Use as Antigen

4.5.1 Inclusion Bodies

The recombinant protein was expressed to a reasonably high level in *E coli* as shown by the SDS-PAGE results. Microscopic examination of the cells expressing from pBMF33 showed evidence of inclusion bodies when compared to cells expressing pBMF32 or the vector only (results not shown). It is possible to purify inclusion bodies for use as antigen Harlow and Lane (1988). The method described by Harlow and Lane (1988) (7.14) was used in an attempt to purify the inclusion bodies. However, this procedure produced a paste that resembled the original cell insoluble fraction and it was not considered suitable or safe to use as an immunogen in rabbits.

4.5.2 Gel Fragments

Ausubel et al (1989) and Harlow and Lane (1988) suggest that polyacrylamide gel fragments make a suitable adjuvant and that subcutaneous injection of crushed excised bands from an SDS-PAGE can be used to raise antibodies. In addition, use of Freund's complete adjuvant was discouraged due to the severe effect this adjuvant has on the animal while alternative commercial adjuvants were deemed too expensive. Thus, it was decided that the most expedient means of producing antigen was to use polyacrylamide gel fragments, a system which also dealt with the solubility problem of the recombinant protein.

A 100ml culture was prepared in T-broth, grown to an OD_{650} of about 1.0 and induced with IPTG at a final concentration of 0.5mM. The cells were harvested by centrifugation in 5 aliquots. Soluble and insoluble fractions from one aliquot were prepared as described above. The insoluble fraction was suspended in 1ml of buffer X with 250µl of 5xSDS-PAGE loading buffer added followed by boiling for 2-5 minutes. Particulate material was removed by a brief centrifugation and the samples were loaded onto a 10%/4% SDS-PAGE with one lane reserved for markers. After electrophoresis, the gel was stained with Pro-Blue and the recombinant protein bands identified and excised with reference to the markers and their strong staining on the gel. The excised bands were cut into small pieces with a scalpel then placed into microfuge tubes and rinsed several times with

physiological saline (Sigma). The gel fragments were then passed through a 21 gauge needle several times to reduce the size further in readiness for injecting into the rabbit.

4.5.3 Rabbit Immunisation Program

Three New Zealand White rabbits were provided and maintained by the MFAA unit in the Biochemistry Department, University of Edinburgh. All injections and test bleeds were carried out by staff members of the MFAA unit.

A 3ml sample of blood was taken from each animal prior to immunisation with recombinant protein to provide a pre-immune sample. Two preparative SDS-PAGE yielded approximately 3ml of crushed gel with half of this material used for the initial immunisation. and divided as evenly as possible amongst the rabbits. Each animal was injected subcutaneously at 4 positions over its back. A test bleed was taken from each animal after 12 days with the blood allowed to clot and the serum collected by mild centrifugation (2000rpm Hereaus model 22R centrifuge) and pooled. The animals were boosted with further gel antigen after 4-5 weeks with a test bleed taken about 10-12 days later.

4.6 Antibody Results

4.6.1 Preliminary Results with Whole Serum

A Western blot was prepared to determine if the rabbits were producing antibodies to the recombinant protein after the first test bleed. A 12 well 10%/4% SDS-PAGE was used with 6 samples run as two duplicate sets to allow probing with pre-immune and test-bleed sera. A Western transfer to Hybond-C was carried out as described in the materials and methods (7.13) using a serum dilution of 1:100. Detection was with the chloro-napthol HRP system (7.13.2) and the results, presented in figure 4.5, show no obvious signs of antibody production from the rabbits but the pre-immune serum gave an indication that there were reasonably low levels of background signal.

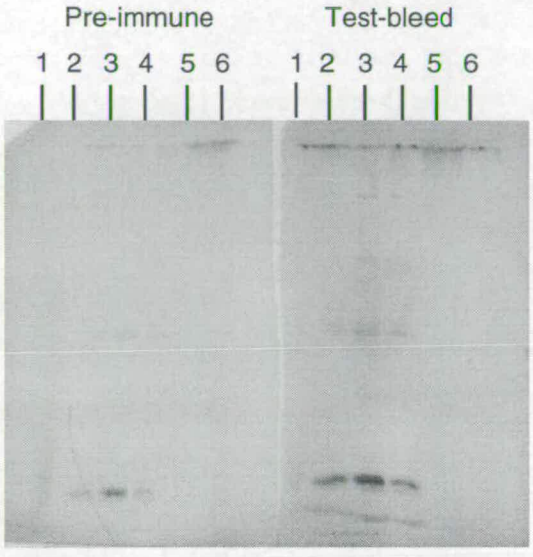


Figure 4.5 Western blot probed with pre-immune serum and first test-bleed serum. Duplicate samples were run on an SDS-PAGE with the gel divided in half and then Western blotted. Lane 6 represents a sample from an initial, unsuccessful, attempt to purify recombinant protein by Ni-agarose affinity chromatography and was included as space was available. The markers had degraded.

The experiment was repeated after further boosts had been applied to the rabbits. A Western blot using similar conditions to those described above was prepared. Extracts from cells expressing pBMF32 and pBMF33 were run on the SDS-PAGE with the pBMF33 sample producing a discernible band of the expected size of 80kDa and not seen in the pBMF32 sample (this result is not shown but a similar result is shown in figure 4.6). This result indicated antibodies were being produced against the recombinant protein. A further experiment was conducted to determine if specific proteins could be detected in samples from *Drosophila*, with extracts from ovaries, adult males and adult females prepared. Ovaries from 20 females were collected and homogenised in 100µl of PBS followed by the addition of 100µl of 2xSDS loading buffer, then boiled for 3-5 minutes. OregonR adult flies were etherised with 25 of each sex sorted and placed into microfuge tubes. The flies were homogenised in 250µl of PBS, then 250µl of 2xSDS-PAGE loading buffer was added and the homogenates were boiled for 3-5 minutes. All three samples were centrifuged in a microfuge to pellet debris. Two sets of 6 samples (details given in figure 4.6) were loaded on to a 10%/4% SDS-PAGE with

one half of the gel used for Western transfer to Hybond-C and the other half stained with Pro-Blue to check the amount of protein loaded.

The results, in figure 4.6, show that the recombinant protein was detected in the cell extract from the *E coli* expression strain (pBMF33), and that high molecular weight signals were detected in adult fly lanes. This result was considered inconclusive at the time as the cDNA sequence had not been fully resolved and the expected wild type protein size was thought to be in the order of 75-85kDa based on the earlier estimated transcript size of about 3.5kb (see Chapter 2).

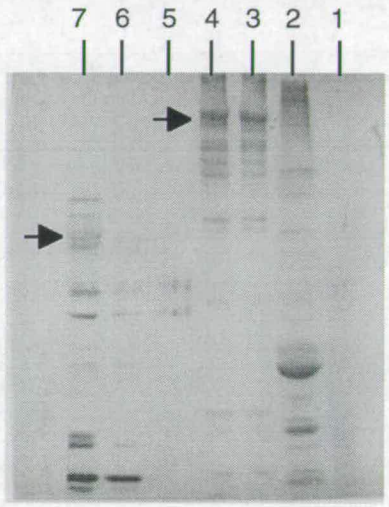


Figure 4.6 Western blot with *Drosophila* samples included. This membrane also includes samples from *E coli*; pBMF32 (negative control) and pBMF33 where the recombinant protein is marked with an arrow and appears as a large diffuse band. High molecular weight proteins were also detected in the adult fly extracts and are also marked with an arrow. Detection was with 1-chloro-4-napthol-HRP system. The dye labelled markers failed to show.

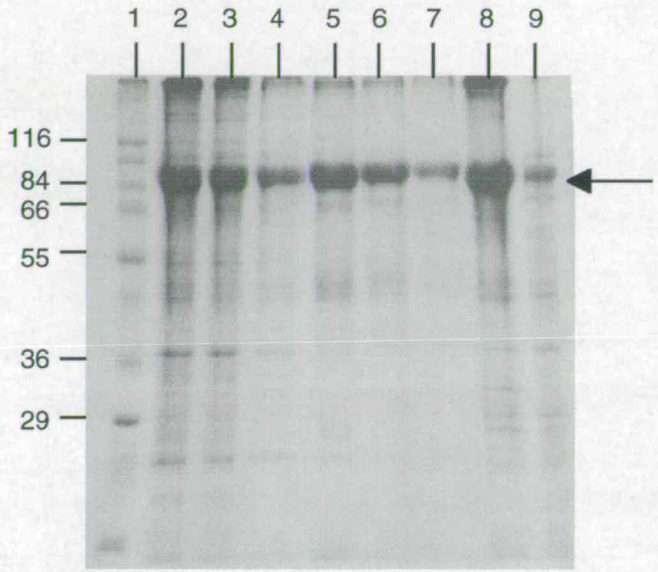
4.6.2 Affinity Purification of Recombinant Protein

The histidine tag fusion protein facilitated by the pRSET expression system allows recombinant protein to be purified by Ni-agarose affinity chromatography under denaturing conditions for situations where the recombinant protein is insoluble. Either urea or guanidinium-HCl can be used as the chaotropic agent, with urea

chosen for use in these experiments as it allowed samples to be analysed directly by SDS-PAGE. Protein was eluted from the Ni-agarose using imidazole.

4.6.2.1 Expression of Recombinant Protein

A 200ml culture in T-broth expressing pBMF33 was prepared and induced as described above. The cells were recovered and divided into 5 aliquots. Soluble and insoluble fractions were prepared from one aliquot as before but used buffer X1 (10mM Tris-HCl pH 8.0, 500mM NaCl, 2mM imidazole pH 7.9, 0.1%v/v Tween-20). The insoluble fraction of the cell extract was dissolved overnight in 20ml buffer X1U (buffer X1 + 6M urea) with gentle stirring. Material that remained insoluble was removed by centrifugation. A 2ml Ni-agarose column was prepared in a 30ml syringe plugged with a glass bead. The column was equilibrated with 20ml (10 volumes) of buffer X1U then loaded with 10ml of the solubilised fraction. This fraction was passed over the column once more to ensure saturation af the Ni^{+} ions. The column was washed with 10ml of buffer Y1U (as for buffer X1U but with 20mM imidazole pH 7.9). Bound protein was eluted with buffer Z1U (as for buffer X1U but with 200mM imidazole pH 7.9) with samples collected as aliquots in microfuge tubes. Samples were analysed on a 10%/4% SDS-PAGE gel with the results shown in figure 4.7. It was clear that recombinant protein was being obtained at a reasonable degree of purity from the column, however, for it to be of further use the urea needed to be removed.



1 Markers - Sigma wide range
2 Ni-agarose column - loaded material
3 Ni-agarose column - fallthrough fraction
4 Ni-agarose column - wash
5 Ni-agarose column - elution fraction 1
6 Ni-agarose column - elution fraction 2
7 Ni-agarose column - elution fraction 3
8 pBMF33 Insoluble fraction
9 pBMF33 new preparation

Figure 4.7 SDS-PAGE of Ni-agarose affinity purified recombinant protein. Marker molecular weights are in kDa. Recombinant protein is marked with an arrow.

4.6.2.2 Dialysis and Concentration

The fractions collected from the Ni-agarose column were pooled and placed into dialysis tubing prepared as described in Sambrook et al (1989). The urea and imidazole were removed from the sample by dialysis against a series of buffers containing decreasing amounts of urea . The base buffer, W, was 10mM Tris-HCl pH 7.5, 100mM NaCl, 0.1%v/v Tween-20 and 5%v/v glycerol. Lots of 500ml of buffer W were prepared with 4M, 2M, 1M and 0.5M urea. The recombinant protein was dialysed successively against each buffer beginning with W+4M for a minimum of 4 hours with stirring. A final dialysis was carried out with 3 changes of buffer W without urea. The recombinant protein was centrifuged at 10000 rpm for 15 minutes (Sorvall RC-5B) to remove any precipitated material, then concentrated on a centricon 50 ultrafilter (Amicon) using the manufacturer's instructions to a final volume of 1.5ml.

4.6.3 Coupling of Recombinant Protein to Affi-Gel 10

The amount of recombinant protein recovered from the affinity purification procedure was estimated to be 1.5 - 2mg on the basis of band densities on SDS-PAGE gels (a faint band detectable by Pro-Blue staining represents about 1µg of protein). Half of this protein was coupled to Affi-Gel 10 (BioRad, gift from A Jarman), an activated acrylic matrix, to make a column for affinity purifying antibodies. The Affi-Gel 10 was prepared by placing about 1.5ml in a plugged syringe column and washing with 10-20 column volumes of sterile distilled water. The gel was then removed to a 5ml bijou bottle and the recombinant protein added. The protein-gel slurry was incubated at room temperature on an orbital shaker with sufficiently vigorous shaking to prevent the gel settling. Unbound sites on the Affi-Gel 10 were blocked by the addition of 100µl 1M ethanolamine pH 8.0, incubated with shaking for one hour at room temperature followed by incubation overnight at 4°C. The gel was then transferred to a small glass column (BioRad) and washed with 20 volumes of PBS.

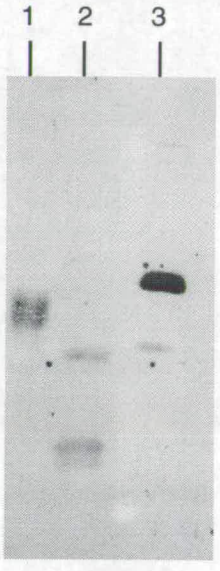
4.6.3.1 Affinity Purification of Antibody

The procedure described in Harlow and Lane (1988) was used for affinity purification of antibodies. Serum, 3ml, was thawed and centrifuged to remove precipitated material that would otherwise block the column. The serum was diluted to 30ml with PBS and was passed over the column 3 times. The column was washed with 20ml PBS followed by 25ml PBS + 500mM NaCl. A low pH elution was carried out with 15ml of 100mM glycine pH 2.5. The pH of the column was then adjusted with 25ml 10mM Tris-HCl pH 8.4 and further antibody was eluted with 15ml triethylamine pH 11.5. The column was then washed with 20ml of PBS with 0.1%w/v Na-azide added to the final few millilitres of PBS to prevent microbial growth.

Each eluted antibody fraction was concentrated separately to about 1ml using centricon 50s (Amicon) with 10x PBS added to make the solution 1x PBS to stabilise the antibodies. An immunoblot was performed by spotting samples directly onto Hybond-C and using the Western blot procedures for probing and detection. The result showed antibody present in both high and low pH fractions (result not shown) These fractions were combined, then concentrated and diafiltered with 5 to 8 volumes of PBS in a centricon 50.

4.6.3.2 Western Blot with Affinity Purified Antibodies

A Western blot was carried out using the affinity purified antibodies with ovary, adult fly extract, purified recombinant protein, and a pBMF32 culture extract as control. Detection was with the DAB system, with the result shown in figure 4.8. A strong band was observed in the ovary lane which was estimated to be 110-120kDA in size whereas a full length unconventional myosin would be expected to produce a 180-200kDA band. However, the markers failed to transfer so the sizing of bands on this membrane is uncertain. A large, diffuse band was observed in the recombinant protein lane consistent with that expected. The pBMF32 control lane showed weak background reactivity.



1 Ni-agarose purified recombinant protein
2 pBMF32 extract (negative control)
3 Ovary extract

Figure 4.8 Western blot with *Drosophila* extracts. The antibody used was affinity purified from an Affigel-recombinant protein column. dection was with the DAB system. The Sigma markers degraded.

4.7 Immunohistochemistry

Experiments were undertaken using both whole serum and affinity purified antibodies on whole mount ovaries in order to determine the expression pattern of the class V myosin . The method used is described in the materials and methods, section 7.16. Initially an HRP detection system was used but it was found that post-fixed ovaries exhibited a natural peroxidase activity and thus strong backgrounds occurred in all samples. Experiments were undertaken with a fluorescent (FITC-conjugated) secondary antibody but again strong background signals masked any genuine signal that may have been present. As these techniques clearly require further refinement, the preliminary results (showing high background signals) have not been presented.

4.8 Summary

Initial experiments designed to obtain expression from the pNMC7a *Eco*RI fragment proved unsuccessful. The most probable reason for this failure was that the expressed protein lacked the last 40 amino acids from the carboxyl tail of the wild type protein and thus was unstable and rapidly degraded in the *E coli* cell. When further cDNAs were isolated an improved expression plasmid was constructed utilising much more open reading frame and including all the carboxyl tail. This plasmid proved successful for producing recombinant protein although the protein was insoluble. However, the system used for expression allowed the recombinant protein to be purified using denaturing conditions.

The most expedient means to raise antibodies in rabbits was to use polyacrylamide gel fragments containing the recombinant protein. The polyacrylamide acted as an effective adjuvant and SDS-PAGE circumvented the solubility problem encountered with the recombinant protein. The antibodies raised were shown to be reactive to the recombinant protein but it is unclear whether or not they recognise the wild type *didum* protein from *Drosophila* tissue extracts on Western blots. An early Western blot (figure 4.6) showed high molecular weight bands from the adult fly extracts, while a later Western blot (figure 4.8), using affinity purified antibody, produced a strong signal in the ovary extract lane with the size of this protein estimated to be 110-120kDa. The later result shows a protein much smaller than that expected for a full length unconventional myosin. Further experiments are required with the affinity purified antibody to resolve these differences.

Use of the antibodies for immunohistochemical studies demonstrated that procedural problems need to be overcome. However, there is considerable scope for refinement of these experiments as only a preliminary examination was carried out.

The successful expression of recombinant protein and the development of a procedure for its purification provide the basic tools required for studying the localisation of the *didum* protein in wild type and mutant flies.

Chapter 5

Additional Genes Identified During This Work

5.1 Introduction

The work presented in this thesis is primarily concerned with the function of an class V myosin in oogenesis. However, during the course of this investigation, several other genes were identified. This Chapter describes how these genes were identified and presents the complete sequence of one cDNA and the initial identification of another gene.

5.2 cDNA-34

The Reverse Northern approach used towards the identification of genomic fragments containing 5' exons of the myosin V gene is described in detail in Chapter 2. Initial sequence analysis showed that no similarities could be found between the new cDNA (now termed cDNA-34) and the existing *Drosophila* myosin V DNA sequence or the mouse *dilute* unconventional myosin peptide sequence. On this basis it was concluded that cDNA-34 represented an unrelated cDNA and, therefore, a new gene.

An initial restriction map indicated that the insert cDNA from this recombinant plasmid was about 2.4kb using the restriction endonucleases *Eco*RI and *Hin*dIII. A more extensive restriction map showed few of the restriction endonuclease sites from the pBluescript multiple cloning site cut within the insert. Those that did were *Cla*I, *Pst*I and *Xba*I with the initial restriction map generated shown in figure 5.1.

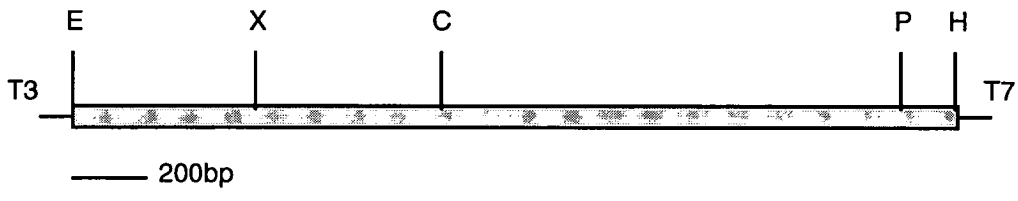


Figure 5.1 Schematic representation of the initial restriction map obtained for cDNA-34. T3 and T7 indicate the respective positions of T3 and T7 primers. Restriction endonucleases; C: *Cla*I, E: *Eco*RI, H: *Hin*dIII, P: *Pst*I, X: *Xba*I.

5.2.1 Sequencing Strategy

A similar subcloning method to that used on the 2.2kb myosin cDNA was employed to determine the DNA sequence of cDNA-34. A set of deletion constructs were prepared using *Cla*I, *Pst*I and *Xba*I. At the same time the fragments being deleted from the original plasmid were recovered and subcloned into pBluescript SK- as before. A list of the subcloned recombinant plasmids generated is given in Table 5.1. Although the original size of the insert had been estimated at 2.4kb, it was not possible to obtain a complete overlap of the sequences generated from these plasmids. The sequence data also revealed the presence of a second *Cla*I site 118bp from the first site identified in the restriction map. In addition, 3 *Nco*I sites within the cDNA were identified from the sequence data and these sites were used to generate 2 additional *Nco*I subcloned recombinant plasmids in pGEM5 (Promega) and for a *Nco*I-*Xho*I deletion plasmid. As the overhangs produced by *Xho*I (in the multiple cloning site) and *Nco*I (within the cDNA) were incompatible for re-ligation, end-filling was required. This procedure was accomplished by treatment with dNTPs at 200μM each and 1 unit of Klenow enzyme.

Table 5.1 List of Plasmids Used for Sequencing cDNA-34

Plasmid Name	Description	Vector	Primers
cDNA-34	Original cDNA, 2.67kb insert	pBS SK-	T7, T3
pBMF35	*Cla*I deletion	pBS SK-	T7
pBMF36	*Cla*I subcloned fragment	pBS SK-	T7, T3
pBMF37	*Pst*I deletion	pBS SK-	T3
pBMF38	*Pst*I subcloned 0.15kb	pBS SK-	T7, T3
pBMF39	*Pst*I subcloned 0.3kb	pBS SK-	T7, T3
pBMF40	*Pst*I subcloned 1.85kb	pBS SK-	T7, T3
pBMF41	*Xba*I deletion	pBS SK-	T3
pBMF42	*Xba*I subcloned 0.5kb	pBS SK-	T7, T3
pBMF44	*Nco*I subcloned 1.2kb	pGEM5	T7, SP6
pBMF45	*Nco*I subcloned 0.2kb	pGEM5	T7, SP6
pBMF46	*Nco*I-*Xho*I deletion	pBS SK-	T7

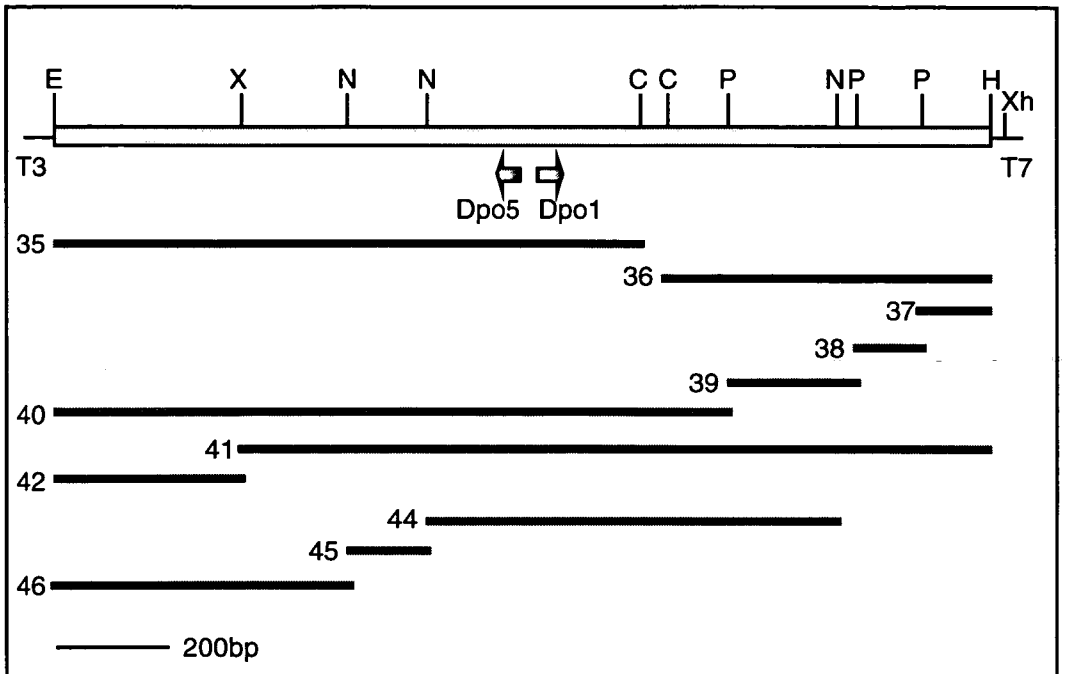


Figure 5.2 Schematic diagram of cDNA-34 showing all subcloned plasmids generated for DNA sequencing (see also Table 5.1). Also shown are 2 oligonucleotide primers, Dpo1 and Dpo5 used for sequencing. Restriction endonucleases; C: *Cla*I, E: *Eco*RI, H: *Hin*dIII, N: *Nco*I, P: *Pst*I. X: *Xba*I, Xh: *Xho*I.

5.2.2 Sequence Analysis

The full DNA sequence of cDNA-34 is 2674bp and is shown in Appendix III. This sequence has been submitted to the GenEMBL database with accession number Y07720. Analysis of the sequence with the GCG MAP program revealed an open reading frame of 524 amino acids and is shown schematically in figure 5.3. A 3' untranslated region (3'UTR) of 654nt and 5'UTR of 433nt are also observed.

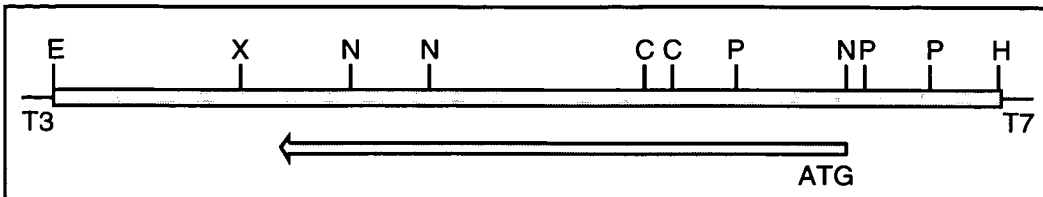


Figure 5.3 Schematic representation of the open reading frame deduced from cDNA-34. Restriction endonucleases; C: *Cla*I, E: *Eco*RI, H: *Hin*dII, N: *Nco*I, P: *Pst*I, X: *Xba*I.

5.2.3.1 The ATG and Poly Adenylation Sites

Cavener (1986) reports the consensus sequence preceding the ATG start codon for *Drosophila melanogaster* genes as being C/A A A A/C ATG. The four nucleotides preceding the predicted start codon for cDNA-34 conform reasonably well to this consensus, being C A C C.

The polyadenylation consensus signal is gven as AATAAA (Proudfoot, 1991). This sequence can be identified at nucleotides 2320-2324 which is 306nt into the 3'UTR. Two occurrences of the sequence AATTAA occur at nucleotides 2632- 2637 and 2641-2646, very near the end of the cDNA.

5.2.4 An Ion-Dependent Inorganic Phosphate Co-transporter

The deduced protein sequence was used to search the GenEMBL database using the GCG TFASTA program and identified sequence similarity to a mouse sodium-dependent inorganic phosphate co-transporter (BNPI, Ni et al [1994]). The BNPI protein spans the cell membrane and acts as a inorganic phosphate transporter across the membrane driven by a sodium ion gradient.

```
        301                                                        350
Humpo   eYItSSLvQQ v....SSsRQ SLPIKAilKS LPvWAIsigs FtffWshnim
Bunpo   eYItSSLiQQ g....SStRQ SLPIKAmiKS LPlWAIsfcc FaylWtysrL
Moupo   dYIiSSLmqq a....SSgRQ SLPIKAmlKS LPlWAIilns FafiWsnslL
Ratpo   kYIedaiges aklmnpvtkf ntPwrrffts mpvyAIivan FcrsWtfylL
Dropo   kfIekslgas iq....gskg ptPwKAiats rpvWlnvvaq wggiWglftL

        351                                                        400
Humpo   tlYTPmfIns mLHVNIkENG fLSSLPYLfA wICGnlAGql sDFfLtRnIl
Bunpo   ivYTPTlIns mLHVdIRENG LLSSLPYLfA wICGviAGht aDFLmsRnml
Moupo   vtYTPTfIst vLHVNvRENG LLSSLPYLlA yICGilAGqm sDFfLtRkIf
Ratpo   lisqPayfee vfgfeIskvG LvSaLPhLvm tIIvpigGqi ADFLrsRhIm
Dropo   mthaPtyfrl ihHwNIRatG fLSgLPhLmr mlfayvfsif ADyLLrtdkm

        401                                                        450
Humpo   SviaVRKLFT aaGfllPaiF gvCLpYLSst FYSiviFLiL AgatgSFclg
Bunpo   SlTaiRKLFT aiGllPivF smCLlYLSsg FYStItFLiL AnassSFclg
Moupo   SivtVRKLFT tlGsfcPviF imCLlYLSyn FYStviFLtL AnstlSFsyc
Ratpo   StTnVRKLmn cgGfgmeatl llvvgY.Shs kgvaIsFLvL Avgfsgfais
Dropo   SrTnVRKLat fiCggtkgli vlalayfgyn ataaIvlvtv Atmlhgavss

        451                                                        500
Humpo   GvfINgLDIA PRYfGFiKac StLtGMiGGl IaSTltGLiL .kQDPEsaWf
Bunpo   GaLINaLDlA PRYyvfiKgV ttLiGMtGGm tSSTvaGLfL .sQDPEssWf
Moupo   GqLINaLDIA PRYyGFlKaV taLiGMfGGl ISSTlaGLiL .nQDPEyaWh
Ratpo   GfnvNhLDIA PRYasilmgi SngvGtlsGm vcpiivGamt .khktreeWq
Dropo   GpLasmyDLs PnYaGivlgV SgmiGgmpGf ISpfivGqlt hnnqtidaWk

        501                                                        550
Humpo   KtFiLMaaIN VtgliFYliv AtaEIQDWAK EkqhTRL*.. ..........
Bunpo   KiFLLMSiIN VisviFYliF AkaEIQDWAK EkqhTRL*.. ..........
Moupo   KisflMagIN VtclvFYflF AkgEIQDWAK EiktTRL*.. ..........
Ratpo   yvFLiaSlvh yggvIFYgvF AsgEkQpWAe peemseekcg fvghdqlags
Dropo   nvFLLtSlml tgsgIlYvlF sesklQpWns gchqlpdsgl kelqnlgrdq

        551                                                592
Humpo   .......... .......... .......... .......... ..
Bunpo   .......... .......... .......... .......... ..
Moupo   .......... .......... .......... .......... ..
Ratpo   desemedeve ppgappappp sygathstvq pprppppvrd y*
Dropo   ddeeekkplk sdhdketpiv aeqetktksd cdgk*..... ..
```

Figure 5.4 continued

Several similar genes have been described from mouse (Chong et al, 1995), rabbit (Werner et al, 1991) and human (Miyamoto et al, 1995). An alignment of these sequences using the GCG PILEUP program is shown in figure 5.4. The percentile results of a pairwise comparison between these protein sequences using the GCG GAP program is shown in table 5.2.

Table 5.2 Comparison of sodium-dependent inorganic phosphate co-transporters using GCG GAP program. Results are presented as percentage similarity followed by percentage identity.

	Drosophila	rat	mouse	rabbit
human	50 / 25	54 / 29	80 / 65	84 / 70
Drosophila		57 / 32	53 / 28	50 / 27
rat			55 / 30	57 / 32
mouse				81 / 64

It is clear from these results that the rat, rabbit and human sequences are closely related and the mouse and *Drosophila* sequences are more divergent. There are two features of the sequence which can be used to characterise these proteins as sodium-dependent inorganic phosphate co-transporters; a series of hydrophobic domains for spanning the cell membrane and a Na^+ binding domain.

A comparison of the rat sequence of Ni et al (1994) and the *Drosophila* sequence using the hydrophobicity/hydrophilicity plot from the GCG PEPPLOT program shows a very similar profile between the two protein sequences (figure 5.5). The predicted membrane spanning sequences from the rat sequence have been identified by Ni et al (1994) and are shown underlined in figure 5.6, which shows an alignment with the *Drosophila* sequence.

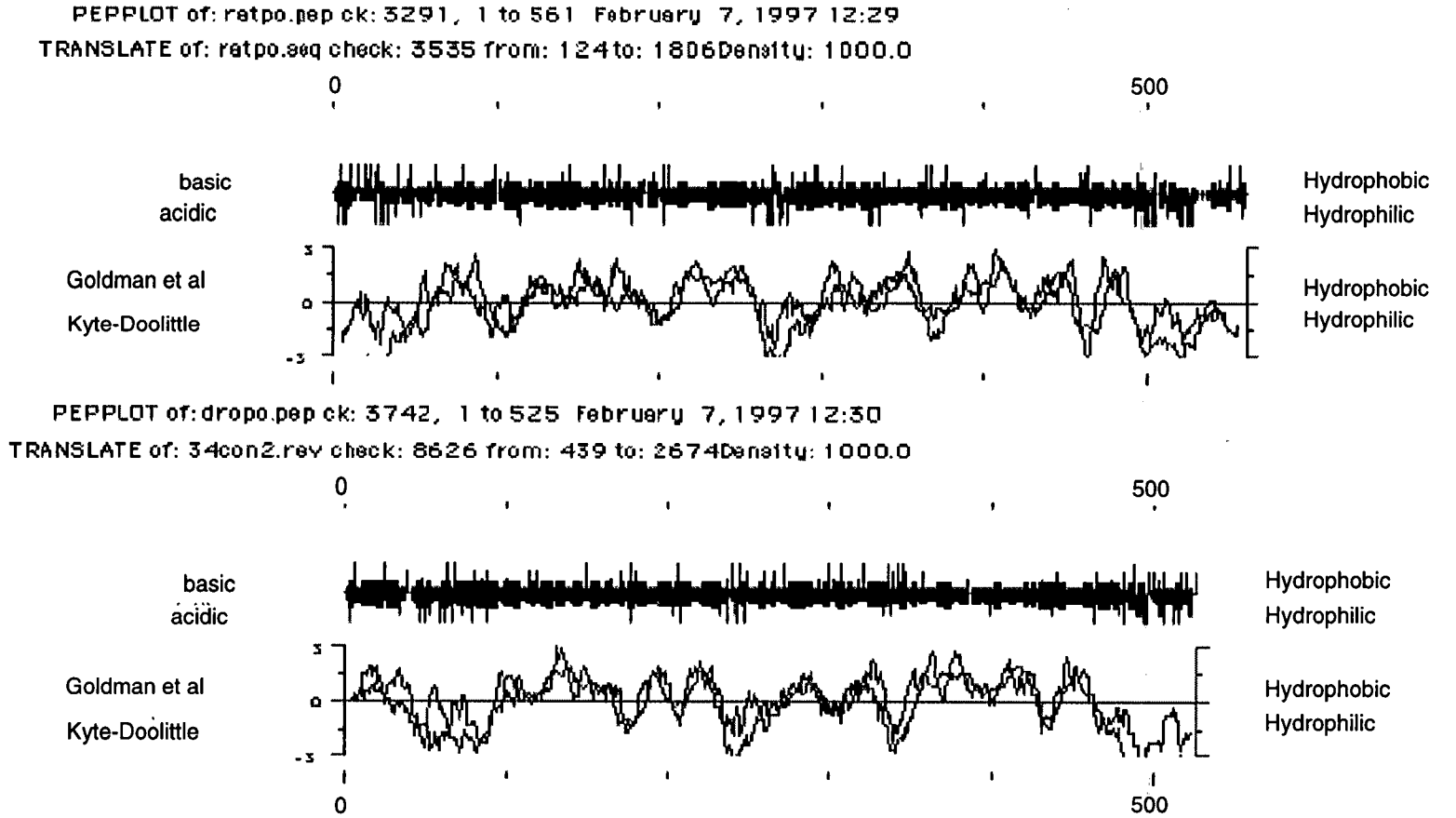


Figure 5.5 Results of GCG PEPPLOT showing hydrophobicity and hydophilicity of the rat (Ni et al, 1994) and *Drosophila* peptide sequences. A similar profile can be observed for both sequences.

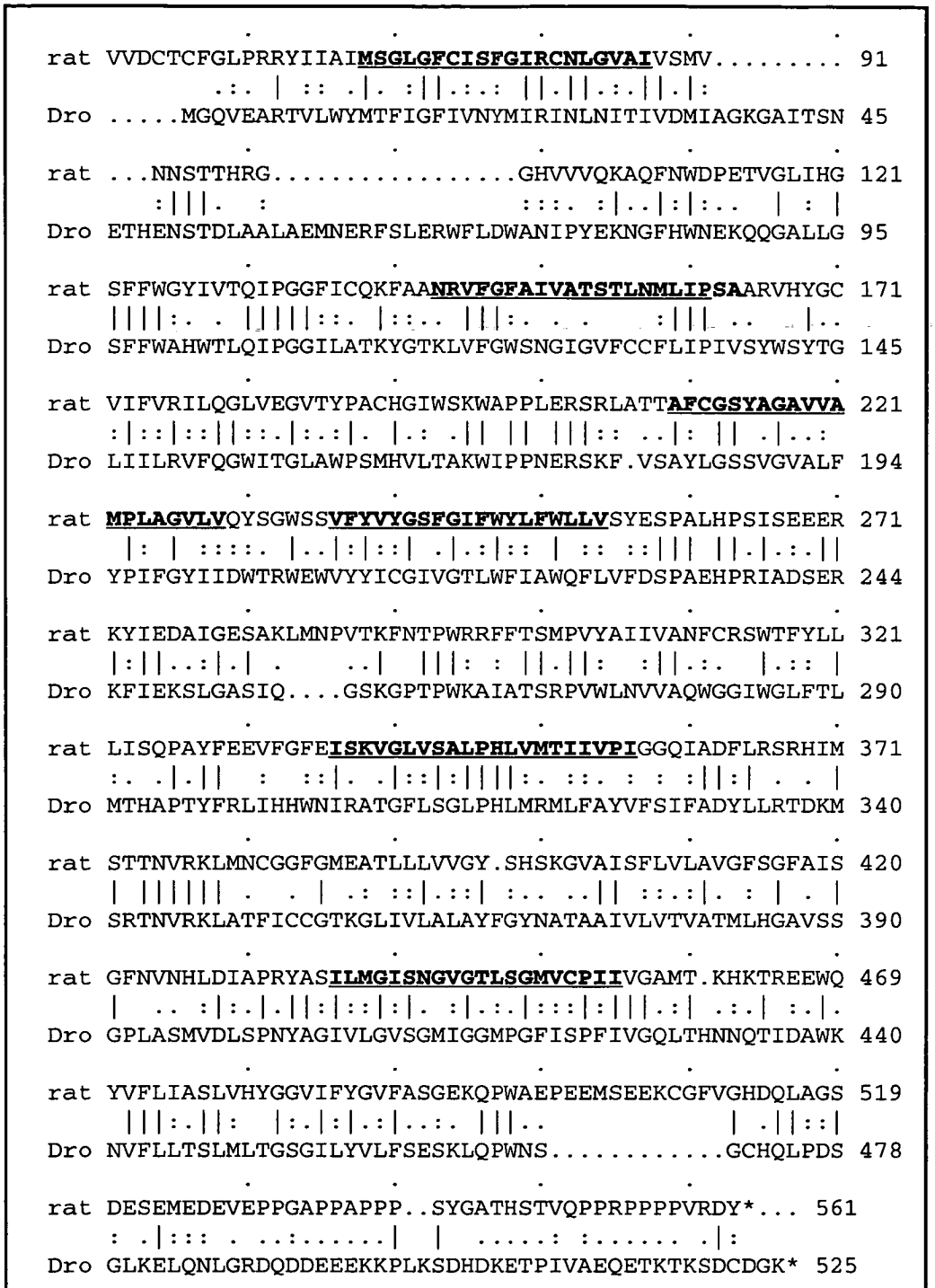


Figure 5.6 GCG GAP comparison of the rat sequence and the new *Drosophila* sequence with the putative hydrophobic membrane spanning regions of the rat sequence highlighted.

The Na^+ binding domain has been highlighted in figure 5.4. These amino acids have been highlighted on the basis of alignments from Ni et al (1994) who used a variety of sodium-dependent transporter systems. The sequences used, with the consensus amino acids capitalised, included:

```
  rat  Na/phosphate  382 cgGfg......aisgfnvnhLdiaPRya 434
rabbit Na/phosphate  340 aiGLl......clggAlinaLdlaPRyy 393
rabbit Na/glucose    378 lrGLm......irkkAsekeLmiaGRlf 429
human  Na/glucose    378 lrGLm......vrkrAsekeLmiaGRlf 429
```

and the *Drosophila* sequence is:

```
                     354 ficcgtkgl..vssGpLasmvDLsPnYa 403
```

On first examination it would seem that the *Drosophila* sequence shows poor conformity to the consensus. However, there are many amino acids which are conserved in this region among the sequences shown in figure 5.4 and there are conservative changes in the *Drosophila* sequence for several important positions, eg, L -› V at position 396 and R -› N at position 401. The leading part of the consensus sequence shows a conserved glycine (G) which does not align in the *Drosophila* sequence but there are two glycine residues within the region. The lack of alignment may be due to a greater divergence in the *Drosophila* sequence as all the other sequences compared have been from mammalian sources.

5.3 PCR Identifies Sequences from an ATPase

The myosin head domain shows a reasonable degree of conservation amongst all myosins. Two motifs within the head domain are highly conserved, the ATP-binding domain and the actin binding domain. In the class V myosins, the ATP-binding domain is located about 160 amino acids from the amino-terminus of the protein and covers 10 highly conserved amino acid residues (using the mouse *dilute* myosin V as the paradigm). The actin-binding domain is about 650 amino acids from the amino-terminus in the mouse *dilute* sequence with the most conserved part containing 6 amino acids. A degenerate oligonucleotide primer was designed based on the sequence of the ATP-binding domain in an attempt to obtain cDNA sequence by PCR. The ATP-binding site was chosen over the actin-binding domain as it contained a higher number of conserved residues and it was closer to the 5' end of the gene. Figure 5.7 shows a diagrammatic representation of this experiment.

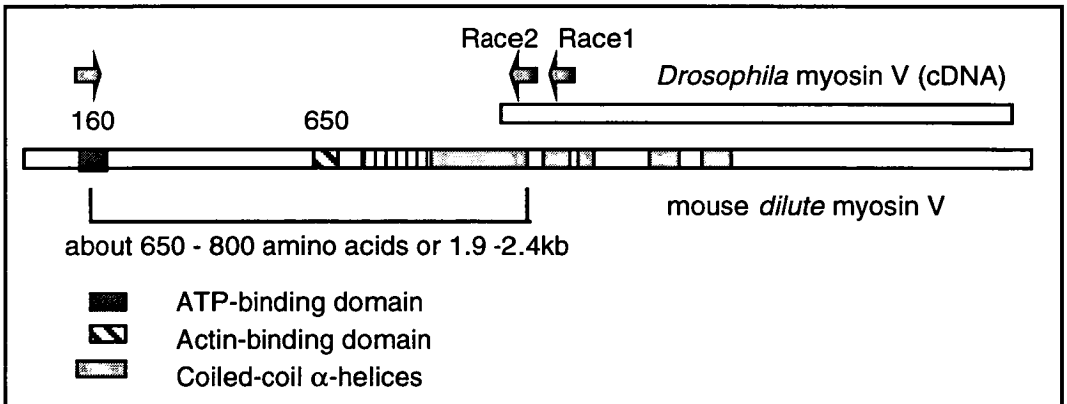


Figure 5.7 Diagram of the ATP-binding and actin-binding domains in mouse *dilute* myosin V (not to scale). Numbers represent the relative postion of each domain from the amino terminus. The approximate size of a PCR product is represented by the bar.

5.3.1 Primer Design

The ATP-binding domains from a number of unconventional myosins were aligned and an oligonucleotide primer designed based on this alignment. Sequences chosen were two class I myosins, Myo-Ia and Myo-Ib (Morgan et al, 1994), and a class VI myosin (Kellerman and Miller, 1992) from *Drosophila*, and the chicken p190 class V myosin (Espreafico et al, 1992). This alignment and the oligonucleotide primer are

shown in figure 5.8. Codon usage in *Drosophila* was taken into account in the primer design. Ashburner (1989) lists the codon usage for a large number of genes. This information has been summarised in percentage from for the amino acids in the ATP-binding domain in Table 5.3. Degenerate bases were built into the primer on the basis of the more frequently used codons.

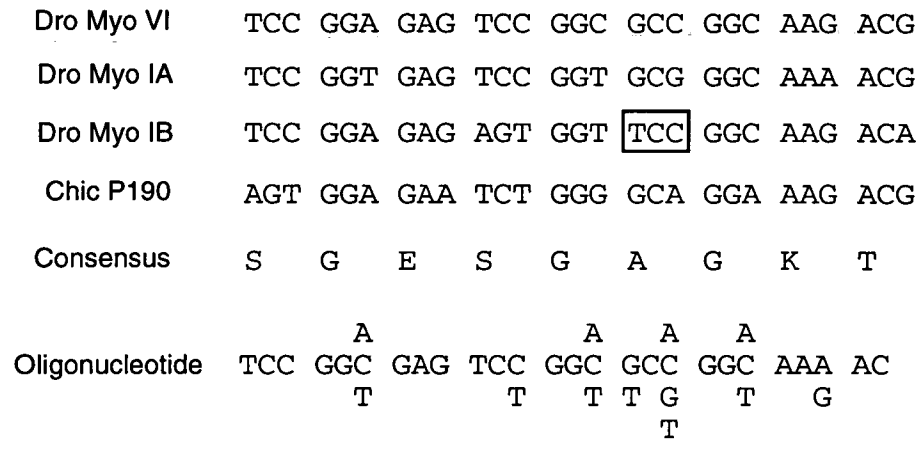


Figure 5.8 Alignment of ATP-binding domain from 3 *Drosophila* unconventional myosins and the chicken p190 class V myosin. The boxed codon in the Myo 1B sequence represents a serine residue. The degenrate oligonucleotide designed from this alignment is shown.

Table 5.3 Codon usage in *Drosophila* presented as percentages from Asburner (1989).

Amino Acid	Codon	Percent
A	GCT	20
Alanine	GCC	50
	GCA	13
	GGG	16
G	GGT	24
Glycine	GGC	43
	GGA	28
	GGG	5
K	AAA	
Lysine	AAG	
S	AGC	12
Serine	AGT	30
	TTT	4
	TTC	18
	TTA	24
	TTG	12
T	ACN	
Threonine		

5.3.2 PCR and Cloning

A standard PCR was performed using the NB40 early embryonic library as template with combinations of ATP-binding domain primer, Race1 and Race2 used. The NB40 library was chosen as it was considered to be the library resource most likely to contain a full length cDNA. The PCR conditions used were:

94°C for 2 minutes
80°C hold, Taq Polymerase added
40 cycles of: 94°C 0.5 minute
53°C 0.5 minute
72°C 3 minutes
72°C 7 minutes

A product of approximately 2kb was obtained as shown in figure 5.9. This PCR product was purified from an agarose gel (7.5.4.3) and prepared for cloning into pBluescript KS+. The PCR product was treated with a number of enzymes to ensure blunt ends were obtained and to add the phosphate groups necessary for ligation. This treatment consisted of:

DNA		20µl
T4 polynucleotide kinase buffer		5µl
2mM ATP		5µl
10mM dNTPs		1µl
T4 DNA polymerase (NEB)		1 unit
T4 polynucleotide kinase	(NEB)	1 unit
Klenow (NBL)		1 unit
Sterile distilled water		to 50µl
37°C for 30 minutes		

The treated PCR product was phenol/chloroform extracted and ligated to a de-phosphorylated *Eco*RV treated pBluescript KS+ plasmid in a 10µl volume. Half of this ligation was used to transform *E coli* XL1-Blue cells using standard methods (7.3.1). Only one transformed colony carrying a recombinant plasmid was obtained from several transformation experiments. This plasmid is referred to as pATP-B

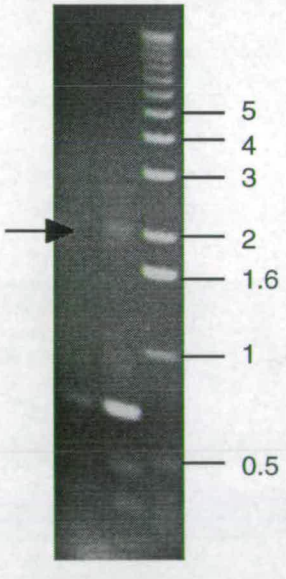


Figure 5.9 Result of PCR with ATP-binding domain and Race2 primers on the NB40 embryonic library. The arrow marks the ~2kb product. Marker sizes shown are in kb.

5.3.3 Mapping and Sequence Analysis

A restriction mapping experiment was performed on a medium scale plasmid preparation (7.6.2.2) of pATP-B with a diagrammatic representation shown in figure 5.10. Only restriction endonucleases *Pst*I, *Xho*I and *Pvu*II cut within the inserted DNA from a selection of *Bam*HI, *Cla*I, *Eco*RI, *Hin*dII, *Kpn*I, *Pst*I, *Pvu*II, *Sac*I, *Xba*I and *Xho*I that were tested (results not shown).

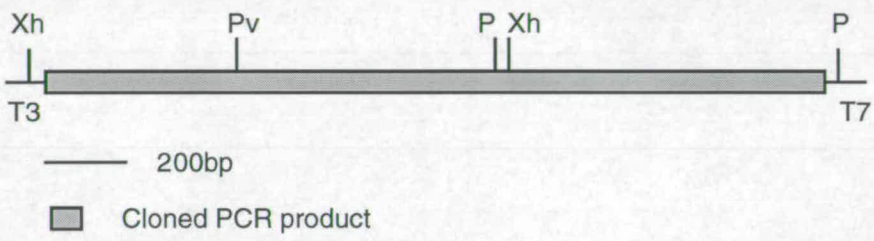


Figure 5.10 Schematic representation of the retriction map of pATP-B in pBluescript KS+. Restriction endonucleases; P: *Pst*I. Pv: *Pvu*II, Xh: *Xho*I.

In addition to restriction mapping, DNA sequencing was carried out using automated cycle sequencing (7.7.2.2) with T3 and T7 primers to determine the sequence at the ends of the inserted DNA. This sequence was used to search the database using the GCG FASTA program with similarity to a cDNA from *Caenorhabditis elegans* called *mei-1* revealed (Clark-Maguire and Mains, 1994). A comparison between the sequence from the T7 end of pATP-B and the *mei-1* deduced protein sequence using TFASTA showed 49% identity over 131 amino acids in frame 1 with this result shown in figure 5.11. A shift to frame 3 at a position where frame 1 identity ceased continued to show sequence similarity with 34.5% identity over 55 amino acids. Amino acid sequence showing identity to the ATP-binding domain can be identified near the beginning of the cloned sequence, consistent with PCR amplification with the ATP-binding primer. Interestingly, the sequence derived from the T3 primer (from the opposite end of the cloned PCR product) also showed similarity to the ATP-binding consensus sequence.

```
SCORES      Frame: (1) Init1:    176 Initn:    267 Opt:
358
             48.9% identity in 131 aa overlap

            210       220       230       240       250       260
cemei1 KQVLHEAVTLPLLVPEFFQGLRSPWKAMVLAGPPGTGKTLIARAIASESSSTFFTVSSTD
                                    :| :|:||||||::|||:::::||::::::
pATP-B              GXNIIGPPTLNSIDGLDSGESGSGKTLIAKSIASQAKAKFFSINPSS
                           10        20        30        40

            270       280       290       300       310       320
cemei1 LSSKWRGDSEKIVRLLFELARFYAPSIIFIDEIDTLGGQRGNSGEHEASRRVKSEFLVQM
       |:||| ||:||:|: ||::| ::|:||||||:|:| ::| :::|:|:: |:|:|||:::
pATP-B LTSKWVGDAEKLVKTLFAVAAAHQPAIIFIDEVDSLLSKR-SANENESTLRLKNEFLIHL
        50        60        70        80        90       100

            330       340       350       360       370       380
cemei1 DGSQNKFDSRRVFVLAATNIPWELDEALRRRFEKRIFIPLPDIDARKKLIEKSMEGTPKS
       ||:::: :: ||:|::||| | |||||:|||| |:::|||:
pATP-B DGAASN-EEIRVLVIGATNRPQELDEAVRRRFVGRLYVPLPTRKHATKSLKSXFIRXSTT
        110       120       130       140       150       160


SCORES      Frame: (3) Init1:     62 Initn:     75 Opt:
92
             34.5% identity in 55 aa overlap

     330       340       350       360       370       380
cemei1 SRRVFVLAATNIPWELDEALRRRFEKRIFIPLPDIDARKKLIEKSMEGTPKS-DEINYDD
                                    :|:: :||:|:||| :: : :: : ::  :
pATP-B GNPSAGNRSHQPTAGAGXGCAPSICRASVRSLTHEEARHKIIEKLIHQVKHNLHVMQLIE
            120       130       140       150       160       170

      390       400       410       420       430       440
cemei1 LAARTEGFSGADVVSLCRTAAINVLRRYDTKSLRGGELTAAMESLKAELVRNIDFEAALQ
       ||: |:|:||:|| :||| |::: :
pATP-B LAELTDGYSGTDVDTLCRYASMAPFAPXHLTNGVIQTITCLPFLLDD
            180       190       200       210
```

Figure 5.11 TFASTA of pATP-B showing significant similarity to the *mei-1* sequence from *Caenorhabditis elegans* (Clark-Maguire and Mains, 1994).

5.4 Summary

Two approaches used towards isolating further cDNA sequence from the myosin V gene have resulted in the identification of two previously uncharacterised genes. A 2674bp cDNA (called cDNA-34) was plasmid rescued from the λZAP library as a result of screening with FIX-1A genomic DNA. This cDNA was sequenced in full and encodes a putative sodium-dependent inorganic phosphate co-transporter. While the sequence identity between the *Drosophila* sequence and other sodium-dependent inorganic phosphate co-transporters is not high, the *Drosophila* sequence possesses characteristics common to these proteins; a series of hydrophobic membrane spanning regions and a putative Na^+ binding domain. An experiment to determine if *Xenopus* oocytes show an increase in phosphate uptake when injected with *in vitro* transcript from this cDNA should confirm the function of cDNA-34.

A PCR approach using a primer based on the ATP-binding domain resulted in the identification of sequence with significant similarity to an ATPase involved in meiotic spindle formation in *C elegans*. No further work has been carried out with this sequence.

Chapter 6

Concluding Discussion

6.1 Cloning of an Unconventional Myosin

The main focus of this project was the cloning of an unconventional myosin. It became clear that the gene had the potential to encode an unconventional myosin when a Northern analysis was undertaken and an estimate of the transcript size had been made. The ovarian cDNA λZAP library had yielded a number of recombinants of 2.2kb but the problem was how to progress towards the 5' end of the gene in view of this library having been inadequately prepared resulting in all *Eco*RI sites being restricted.

It was decided to screen all cDNA libraries that were available in the laboratory. However, this approach was unsuccessful as the cDNA was not found in any of the cDNA libraries examined. The genomic λFIX library was also screened with this library yielding a recombinant phage that contained less coding sequence than was available from the existing cloned cDNAs. The breakthrough came when it was found that the cDNA hybridised to the recombinant P1 DS00574. Isolation of the subcloned pP1-X7 plasmid and the finding that it contained sequence coding for amino terminal residues from the myosin V gene allowed the RT-PCR strategy to be developed to obtain most of the coding sequence. Although the 5'RACE product has not been confirmed as being part of the myosin V sequence, it is hoped that this situation will be resolved in the near future. The gene has been named *didum*.

6.2 A Class V Myosin

This gene has been referred to as coding for a Class V myosin or a Class V unconventional myosin. Conventional myosins are the filament forming myosins found in skeletal muscle, smooth muscle and most non-muscle cells (reviewed in Cheney and Mooseker [1992] and Mooseker and Cheney [1995]). These myosins are also known as type II on the basis of dimer formation through the association of long coiled-coil domains in the carboxyl terminals. A monomeric myosin, or myosin I, was originally identified in *Acanthamoeba* by Pollard and Korn (1973). As more non-muscle myosin types were discovered the term "unconventional myosin" was introduced. A comparison of the many myosin types based on a phylogenetic comparison of the conserved head domain has resulted in the division into, currently, 11 classifications as shown in figure 6.1 (Cheney et al [1993a], Mooseker and Cheney

[1995]). A brief description of each unconventional type will be given followed by a more detailed discussion of the class V myosins.

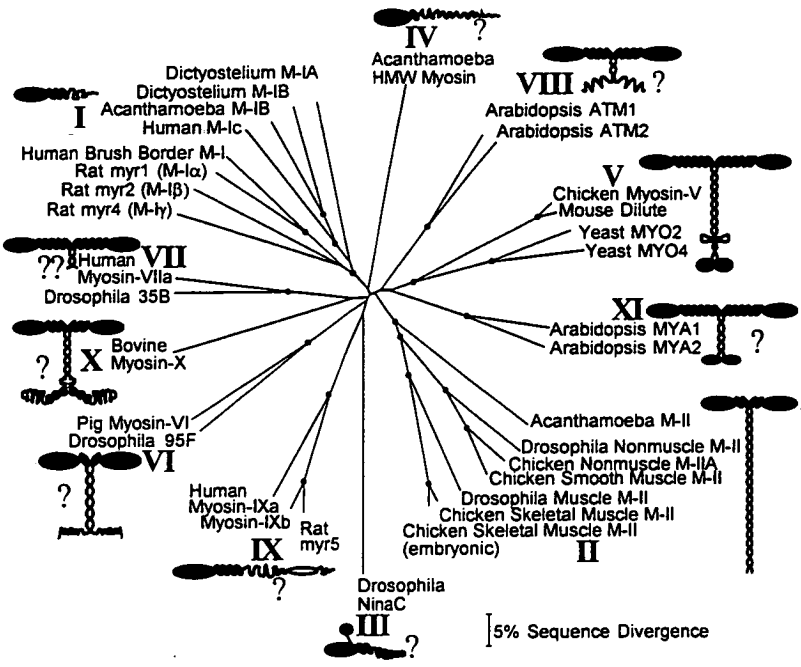


Figure 6.1 An unrooted phylogenetic tree of the myosin superfamily taken from Mooseker and Cheney (1995). A hypothetical model is shown for each class based on sequence prediction.

6.2.1 Class I

The class I myosins have been the most intensively studied unconventional myosins with examples from organisms as diverse as yeast and mammals described. Phylogenetic analysis of the head domains has resulted in these myosins being subdivided into 4 subclasses.

6.2.1.1 Myosin I subclass 1

As mentioned above, the first unconventional myosin was identified in *Acanthamoeba* and ameoboid organisms continue to be a main area of investigation into this class with multiple class I myosins being identified from *Acanthamoeba* and *Dictyostelium*.

There are 5 examples in *Dictyostelium* which have either a short basic residue rich tail region (myoA, myoE) or a longer tail region with 3 distinct domains; TH1, TH2 and TH3 (TH = tail homology, myoB, myoC, myoD). The TH1 domain is rich in basic residues while the TH2 domain is rich in glycine, proline and either alanine or glutamine. This TH2 domain appears to be a ATP-independent actin binding domain (Jung and Hammer, 1994). The TH3 domain contains an SH3 domain, domains well known to facilitate protein-protein interactions.

The ATP-independent actin binding domain (TH2 domain) allow this class of myosin to form cross-linked actin arrays as demonstrated in *Acanthamoeba* (Pollard et al, 1991). Maximal actin activity requires phosphorylation at a site containing either a threonine or serine within the actin binding domain of the head domain (Bement and Mooseker, 1995).

Studies in *Dictyostelium* and *Acanthamoeba* suggest that this class of myosin has major functions in actin based motility, for example pseudopod extension and phagocytosis. Gene knockout studies in *Dictyostelium* have shown that there is considerable functional overlap between these myosins (Peterson et al, 1995).

6.2.1.2 Myosin I subclass 2

This subclass is characterised by the chicken brush border myosin-1. Other homologues of this myosin have been cloned from various mammalian species and all appear to be expressed predominantly in enterocytes. However, there are structurally similar proteins which are expressed in a wider variety of tissues such as myr1 from rat (Ruppert et al, 1993). These proteins bind calmodulin and have 3-6 IQ motifs. The tail domain is rich in basic residues and, in the case of the chicken brush border myosin-1, has been shown to be the target of a kinase and protein kinase C.

6.2.1.3 Myosin I subclass 3

Members of this class are quite similar to the brush border myosins and contain 3 IQ motifs. The *Drosophila* myosin 1b gene (Morgan et al, 1995) falls into this subclassification. This myosin is only expressed in the gut and in the follicle cells during oogenesis.

6.2.1.4 Myosin I subclass 4

The second class I myosin from *Drosophila*, myosin Ia (Morgan et al, 1994), falls within this grouping as do myr4 from the rat (Bähler et al, 1994) and myosin 1a from *C elegans*. The *Drosophila* and rat sequences have 2 IQ motifs with the rat sequences shown to bind calmodulin. The tail domains are rich in basic residues. The *Drosophila* myosin 1a protein is expressed solely in the gut (Morgan et al, 1995). Expression begins late in embryogenesis when gut development starts with an initial pattern similar to myosin 1b. Later development shows differing subcellular locations of these two myosins in the *Drosophila* gut.

6.2.2 Class III

The only members of this class are coded for by the *ninaC* gene in *Drosophila* with two proteins which vary in the size of their tail domain produced by alternative splicing (Montell and Rubin, 1988). The *ninaC* proteins are unusual in that they contain a putative kinase domain at the amino terminus. These proteins are found exclusively in the photoreceptors cells of the adult eye and are essential for the correct functioning of these cells. Mutants show degeneration of the retina over time and with exposure to light (Porter et al, 1992). The larger 174kDa protein seems to carry out the major functions, but both forms seem important in maintaining calmodulin levels within the photoreceptor cell.

6.2.3 Class IV

This class was identified in *Acanthamoeba* by Horowitz and Hammer (1990) in a screen for novel myosins. In addition to the myosin head domain, this protein contains a single IQ motif and a large ~87kDa tail region which incorporates an SH3 domain.

6.2.4 Class VI

Two examples of this class have been described from *Drosophila* (95F, Kellerman and Miller, 1992) and pig (Hasson and Mooseker, 1994). Interestingly, these two sequences show a higher degree of identity over the tail domains than over the head domains suggesting a conserved and vital role for the tail domain.

Some understanding of the function of this class comes from studies on *Drosophila* 95F. It is expressed during early embryogenesis and has a crucial role in the nuclei divisions within the syncytial blastoderm. The 95F myosin is required for the actin based movement of cytoplasmic granules (Mermall et al, 1994, Mermall and Miller, 1995). Use of an anti-95F antibody to disrupt this process results in improper nuclear divisions.

Antibodies to pig myosin-VI reveal that it is expressed in a wide range of tissues including brain, heart, liver, kidney, lung and intestine (Hasson and Mooseker, 1994).

6.2.5 Class VII

At present, no full length sequences have been described for members of this class, but partial sequences have been isolated from *Drosophila* (myosin 35BC, Chen et al, 1991), pig and human. Antibodies react with a large protein of about 240kDa suggesting that the transcript for this protein is also large (Hasson et al, 1995).

Sequences from the *shaker-1* locus in the mouse share similarity to the partial cDNAs described so far (Gibson et al, 1995). The *shaker-1* mutation results in deafness in mice and the locus is believed to be the orthologue of the Usher syndrome 1b which results in deaf-blindness in humans (Steel and Brown, 1994). Familial analysis has revealed mutations in the human myosin VIIb supporting the hypothesis that this syndrome is caused by a dysfunctional unconventional myosin (Weil et al, 1995)

6.2.6 Class VIII

This class represents a distinct group of unconventional myosins described in plants. The ATM1 myosin was identified in *Arabidopsis thaliana* by a PCR based strategy (Knight and Kendrick-Jones, 1993) while a second related gene, ATM2, has also been reported (Kinkema et al, 1994). These sequences have 3 and 4 IQ motifs respectively and regions in the tail domain predicted to form coiled-coil α-helices.

6.2.7 Class IX

A complete sequence from the rat (myr5, Reinhard et al, 1995) and a partial sequence from human (Bement et al, 1994) identify this class of unconventional myosins. They both feature a large insertion of 140 amino acids within the head domain that is thought

to influence the actin binding domain although no aberrant biochemical properties have been reported as yet (Reinhard et al, 1995). These myosins feature 4 IQ motifs and the large tail domain has four distinct subdomains, none of which appear capable of forming coiled-coil domains. The first subdomain is rich in proline and the second has cysteine residues and a structure similar to the zinc and phospholipid binding domains of protein kinase C. Assays using a bacterially expressed fusion protein of this domain demonstrate zinc binding but not phospholipid binding (Reinhard et al, 1995). The third subdomain shows homology to the GTPase-activating proteins from the rho/rac family (Lamarche and Hall, 1994). The final subdomain shows divergence, possibly as a result of alternative splicing, between the myr5 and myosin-IXb sequences, being acidic in myr5 and proline rich in myosin-IXb. The finding of signal transducing domains in this myosin perhaps establishes a link between signal transduction and cytoskeletal rearrangements.

6.2.8 Class X

The main feature of this class of myosin are the PH (pleckstrin homology) domains in the tail domain (Oliver et al [1996], see also Mooseker and Cheney, [1995]). These PH domains are found in proteins associated with signal transduction pathways. Very little is known about these myosins although the bovine sequence shows 3 IQ motifs in the neck region.

6.2.9 Class XI

This class comes from studies in *Arabidopsis* with a PCR based approach identifying two full length sequences, MYA1 and MYA2, and four partial sequences (Kinkema and Schiefelbein [1994] and Kinkema et al [1994]). The MYA1 sequence has been included in the alignment of the class V myosins in Chapter 2 as it shares features of this class. However, the discovery of several related proteins has resulted in their re-classification (Mooseker and Cheney, [1996], Kinkema et al {1994]). In addition to the myosin head domain, the MYA1 and MYA2 proteins have 6 IQ motifs and regions predicted to form coiled-coil α-helices.

6.3 The Class V Unconventional Myosins

After the class I myosins, this class is probably the best characterised of the unconventional myosins. To date, examples have been described in mouse (*dilute*, Mercer et al, 1991), chicken (p190, Espreafico et al, 1992), yeast (*MYO2*, Johnston et al [1991] and *MYO4*, Haarer et al [1994]) and rat (myr6, Zhou et al, 1996). It has also been established that the sequence described by Huang et al (1990) as a glutamate decarboxylase is a second class V myosin within the mouse.

The *dilute* gene was the first to be described as a result of investigations into this common mutant locus in the mouse. The *dilute* locus has 2 distinct phenotypes, a lightening of the coat colour and a neurogenic defect resulting in lethality (Mercer et al, 1991). The lightening of the coat colour is a result of abnormal melanocyte morphology while the cause of the neurogenic defect is unknown. The *dilute* gene is expressed in a wide variety of tissues including brain, kidney, spleen, thymus, testis and muscle but not in the liver. Interestingly, RT-PCR identified the second class V myosin in the liver.

The best biochemically characterised class V myosin is p190 from the chicken brain. This myosin was purified as protein from brain extracts (Espindola et al, 1992) and cDNAs were obtained from an expression library (Espreafico et al, 1992). The IQ motifs were expressed in bacteria and the recombinant protein shown to bind calmodulin although this binding does not appear to have a strict requirement for calcium. Antibody staining in cells cultured from rat hippocampus showed very strong staining in the perinuclear region of the cell body. Staining was also observed along dendritic and axonal processes with the myosin associating with F-actin in these processes.

Further efforts have shown the stoichiometric relationship between calmodulin and this myosin V to be 4:1 (Cheney et al, 1993b). Two additional light chains of 17kDa and 23kDa were identified with this myosin from the purification studies. Visualisation of the myosin by electron microscopy has given clues to the structure of this molecule. A 2-headed molecule was observed, suggesting a dimer, with the heads spaced further apart than those seen in the conventional myosin II. This greater spacing is probably due to the myosin V having 6 IQ motifs versus 2 in myosin II. A stalk domain and a

globular domain consistent with the coiled-coil α-helices and the carboxyl globular domain predicted from the sequence data were also observed (Cheney et al, 1993b).

The chicken p190 protein has been shown to bind to actin filaments and to translocate towards the barbed end by *in vitro* experiments (Cheney et al, 1993b). The ATPase activity was shown to be enhanced in the presence of Ca^{2+} ions but the translocation of actin filaments was strongly reduced by these ions. The effect on translocation by Ca^{2+} was reduced when exogenous calmodulin was added suggesting that the effect of Ca^{2+} is on the binding and turnover of the calmodulin light chains.

Two examples of class V unconventional myosins, *MYO2* and *MYO4*, have been identified in *Saccharomyces cerevisiae*. The *MYO2* gene is essential and is required for bud formation (Johnston et al, 1991). A mutant for this gene showed accumulation of vesicles suggesting that the role of the protein was in transporting material to the bud site. Budding occurs in several stages; bud site selection, bud site assembly, cytoskeleton organisation and polarised growth. Mutants for *MYO2* fail to deposit the chitin ring which is required for bud site formation but these mutants were able to complete the cell division when shifted to the restrictive temperature at a late stage suggesting no role for this myosin in cytokinesis or nuclear migration (Johnston et al, 1991). The myo2p myosin has been shown to be regulated by calmodulin but in contrast with results seen for the chicken p190 myosin, this regulation is independent of Ca^{2+} ions (Brockèrhoff et al, 1994).

The *MYO4* gene was isolated by random cloning and sequence comparison (Haarer et al, 1994). Deletion of the gene showed that it was not essential. Recently the *MYO4* gene was identified as *SHE1*, one of 5 genes involved in the switching of the mating type. The *HO* endonuclease switches the mother cell mating type during cell division. It is postulated that the *SHE* genes are responsible for transporting an *HO* endonuclease repressor into the daughter cell (Jansen et al, 1995).

6.4 Is There a Second Myosin V in *Drosophila*?

It can be seen that there are two distinct myosin V genes in the budding yeast. Results from this laboratory and the recent description of myr6 from the rat show that there is a second myosin V in the mouse in addition to the *dilute* gene. What are the possibilities of a second myosin V in *Drosophila*?

A Reverse Northern approach was used to determine whether genomic DNA cloned into FIX-1A contained any 5' sequence from the *didum* myosin V gene. This experiment was undertaken before any map of FIX-1A had been determined. A clear result was obtained showing hybridisation to bands different to those hybridising to the myosin V cDNA (see figures 2.12 and 2.13 in Chapter 2). Thus, this result suggested that FIX-1A contained 5' sequence from the myosin V gene. However, use of the cloned genomic DNA from FIX-1A to screen the ovarian λZAP library resulted in the isolation of an unrelated cDNA - cDNA-34, encoding a putative ion-dependent inorganic phosphate cotransporter.

How then should the Reverse Northern result be interpreted? It only became apparent that the original FIX-1A isolate contained two independent recombinant phage when this DNA was being used to match the *didum* and cDNA-34 cloned DNA to genes in the 43BC region. The gene encoding cDNA-34 does not map to the 43BC region and the bands seen on the Reverse Northern are now known to associate with the same genomic region as cDNA-34 does. Since the probe used to obtain the Reverse Northern result was generated using reverse transcription from the *didum* myosin V transcript using a gene specific primer, this probe should be specific to the myosin head domain. Therefore, it seems reasonable to predict that the genomic sequence upstream of the cDNA-34 gene contains sequences related to a myosin head domain, possibly from a second class V myosin. Further investigation into this hypothesis can be undertaken by determining the DNA sequence from the recombinant plasmid pE7 and searching for myosin V homology.

As stated in Chapter 2, a 5'RACE product of ~400bp was obtained using gene specific primers. At this stage the RACE product sequence cannot be merged into the existing bulk cDNA. Perhaps this product is from a second myosin V gene. It is likely that a second myosin V will show greater sequence identity to the *didum* sequence over the myosin head domain than over the tail domain. Thus the second myosin could also amplify from primers designed to the *didum* gene.

A developmental profile (A McCormack, personal communication) has shown that the *didum* myosin V gene is expressed predominantly in the gonads of both adult males and females. Some expression is also seen in the early embryo. This embryonic signal might be considered to be a result of maternal transcript carried into the embryo as strong expression of this gene is seen at late stages of oogenesis. However, for this gene to match one of the 3 candidate genes within the 43C region (*43Ca*, *43Cb* and

43Cc), embryonic expression must occur as all 3 genes produce an embryonic lethal phenotype. It is possible that this gene is expressed elsewhere in the fly but at levels undetectable by Northern analysis.

6.5 Role in Oogenesis

The expression pattern of the *didum* gene in oogenesis gives a few clues as to its role. The gene is expressed very early in oogenesis so perhaps there is a function in oocyte differentiation. It is not known what signals control the determination and differentiation of the oocyte. Two genes that are essential for this process are *Bicaudal-D* and *egalitarian*. Now that a mutant myosin is close at hand it should be possible to undertake some epistatic genetic analyses to determine where the *didum* myosin fits within these pathways. Analysis of the *in situ* hybridisation pattern of the *didum* myosin within *Bicaudal-D* and *egalitarian* mutant backgrounds shows that it is disrupted (A McCormack, personal communication).

The second interesting feature of the localisation is the accumulation at the anterior margin of the developing oocyte. However, it should be noted that many transcripts localised to the oocyye undergo a temporal accumulation at the anterior margin, eg, *oskar* (Kim-Ha et al, 1993) and *bicoid* (St Johnston et al, 1989). It is possible that this myosin is involved in specific transport processes across the oocyte. Most investigations of transportation systems within the oocyte have focused on the microtubule network with all localised RNAs shown to be dependent on this network. Indeed, *didum* transcript's localisation can also be affected by depolymerising the microtubule network (A McCormack, personal communication).

While there is considerable information about the role of microtubules within oogenesis, less is known about the actin cytoskeleton. Actin is known to part of the ring canals (reviewed in Cooley and Theurkauf, 1994) as are the products of the *hu-li tao shao* and *kelch* genes. The *hu-li tao shao* product appears to recruit actin to the ring canals while the *kelch* product is probably required for ring canal stability. An actin cytoskeleton is used to anchor the nurse cell nuclei at the time the nurse cells rapidly transfer their cytoplasm to the oocyte. Three genes, *chickadee*, *singed* and *quail* have been identified that affect this process with all 3 genes producing actin binding proteins. These genes have been discussed in Chapter 1, but briefly *chickadee* encodes a homologue of profilin, a protein which controls the assembly of actin filaments. The *singed* gene encodes a fascin homologue, while *quail* encodes a Villin

homologue with both these proteins involved in stabilising actin bundles. There are probably many other proteins that interact with the actin cytoskeleton that have yet to be discovered.

As stated above, localisation of the posterior determinant *oskar* is dependent on microtubules. A surprising finding is that this localisation is also dependent on tropomyosin II (Erdélyi et al, 1995). In a screen for mutations with defects in germ cell formation 9 alleles of the *tropomyosin II* gene were recovered. Creation of germline clones carrying lethal alleles of *tropomyosin II* resulted in the recovery of sterile adults as well as embryos showing abdominal defects. This result shows that the *tropomyosin II* gene is required for posterior localisation of *oskar* and the formation of pole plasm. Examination of the actin cytoskeleton using rhodamine-conjugated phalloidin showed that it was normal in ovaries mutant for *tropomyosin II*, therefore, limiting the effect of this mutation to *oskar* localisation.

This observation, that *oskar* localisation is dependent on tropomyosin II as well as the microtubule network, suggests an interrelationship between the actin cytoskeleton and the microtubule network. There is increasing evidence that this is the case. Studies on the giant axon from the squid show that organelles use both actin and microtubule networks (Kuznetsov et al, 1992). It has been suggested that microtubules are used for transport over longer distances within the cell while actin filaments are used over shorter distances (reviewed by Langford, 1995). In the case of tropomyosin II in oogenesis, one hypothesis is that actomyosin transport is required to move the *oskar* transcripts to the microtubule network at the anterior of the developing oocyte and the microtubule network is then responsible for transporting *oskar* transcripts to the posterior (Erdélyi et al, 1995). This hypothesis is supported by the observation that *oskar* transcripts accumulate at the anterior in the *tropomyosin II* mutants.

6.6 Myosins in Oogenesis

Several myosins are known to function in oogenesis. The *myosin-Ib* gene is expressed in the follicle cells during oogenesis, initially in all follicle cells, but by stage 10, only in the columnar cells surrounding the oocyte (Morgan et al, 1995). Antibody detection shows that this myosin has a basolateral location followed later by an apical location within the follicle cells. This apical location may be associated with the microvilli that extend from the follicle cells into the perivitelline space. The microvilli

have an actin core and are possibly associated with secretion from the follicle cells of yolk proteins, the vitelline membrane and the chorion (Morgan et al, 1995).

The 95F class VI myosin has essential functions during embryogenesis (Mermall et al, 1994, Mermall and Miller, 1995). Work in this laboratory has also shown that this myosin functions in oogenesis (W-M Deng, personal communication). Disruption of 95F myosin function by use of an antisense construct resulted in defects in follicle cell migration. Follicle cell movements are also dependent on nonmuscle myosin II, a conventional myosin (Edwards and Kiehart, 1996). Nonmuscle myosin II forms bipolar filaments and is thought to drive cellular contractile events. Mutations in the regulatory light chain gene, *spaghetti squash*, of this myosin complex affects the centripetal migration of the follicle cells that surround the anterior of the developing oocyte (Edwards and Kiehart, 1996). Strong staining of this myosin is also seen in the border cells.

6.7 Antibodies

While limited progress was made with the antibodies generated to the *didum* protein, the basic foundation has been established. An expression plasmid has been constructed, recombinant protein produced and a procedure for purifying the recombinant protein has been determined. Antibodies have been raised and it has been possible to affinity purify these antibodies. It should be possible to raise further antibodies using purified recombinant protein and refine the procedures required to observe the wild type protein location in whole mount ovaries.

6.8 Future Work on the *didum* Gene

An immediate goal for future work is to identify the remaining coding sequence and 5' untranslated sequence. This goal should be readily achievable using the oligonucleotide primer set available. It would also ideal to physically map and clone upstream 5' sequence. A physical map is useful for mapping P-element mutations that would be generated as part of a mutagenesis study. Genetic mapping has placed the *didum* gene in the 43C region and indications are that the candidate gene results in embryonic lethality. To study the role of this gene in oogenesis will require the generation of chimeric germ line clones. The technique developed by Chou and Perrimon (1992) using site-specific homologous recombination with a yeast recombinase driven by a heat shock promoter can be used. Originally this method

could only be done with X-linked mutants because the mutant ovo^{D1}, which allows the selection for germ line recombinants, was on the X chromosome. Now lines exist with ovo^{D1} on P-elements inserted into each autosome arm (Chou et al, 1993). Using other mutants it can be set up such that only flies with a mutant germline clone actually lay eggs - so thousands of flies do not have to be screened to find abnormal eggs. It would also be possible to mis-express the *didum* gene with P-element vectors and to mis-localise it to the posterior using the *oskar* 3'UTR to gain insights into its function.

Any viable mutations obtained from a P-element mutagenesis can be used in conjunction with other mutated genes to establish genetic pathways and hierarchies. The localisation of other mRNAs, eg *bicoid*, *oskar*, and *gurken* can be examined in a *didum* mutant background, thus identifying the genes which depend upon the correct function of the myosin for normal expression.

Perhaps one of the most interesting questions surrounding not only this myosin but all unconventional myosins is: "what molecules other than actin do these myosins interact with?" A yeast two-hybrid screen would be an ideal experimental system to initiate to begin to answer these questions.

6.9 The cDNA-34 Gene

Interest in membrane spanning inorganic phosphate cotransporters has focused on the kidney of mammalian systems where reabsorption of phosphate is an essential part of maintaining the organism's phosphate homeostasis (Werner et al, 1991). This area of research has resulted in the cloning of three renal sodium-dependent inorganic phosphate cotransporters from rabbit (Werner et al, 1991), mouse (Chong et al, 1995) and human (Miyamoto et al, 1995). These cDNAs appear closely related with the predicted proteins showing at least 80% similarity. In the case of the rabbit gene, expression appears confined to the kidney and to a lesser extent, the liver (Werner et al, 1991).

A different approach was taken by Ni et al (1994) leading to the discovery of the brain sodium-dependent inorganic phosphate cotransporter. These workers used a screen to identify differentially expressed genes after exposure to an excitatory amino acid N-methyl-D-aspartate (NMDA). This screen resulted in the isolation of 11 cDNAs, one of which was the sodium-dependent inorganic phosphate cotransporter and designated BNPI.

The deduced amino acid sequence from BNPI shows a lower level of sequence identity to the predicted proteins from the renal systems, although it contains at least six predicted membrane spanning regions and was shown to mediate phosphate uptake in the *Xenopus* oocyte assay. Thus BNPI establishes a separate family or subfamily of phosphate cotransporters. The new *Drosophila* sequence shows a similar level of sequence identity to the renal phosphate cotransporters as does the rat sequence. Sequence identity to the rat BNPI predicted protein is slightly higher and many conservative substitutions can be observed. The *Drosophila* sequence shares a very similar hydrophobicity/hydrophilicity profile to that of the BNPI protein suggesting that it is a membrane spanning protein.

Analysis by Northern blot shows that BNPI is expressed only in the brain, while *in situ* hybridisation studies show expression is in a neuronal rich subset of brain structures such as the cerebral cortex, hippocampus and cerebellum (Ni et al, 1994). Speculation is that the BNPI cotransporter may be involved in regulating inorganic phosphate levels within the neuron as part of protein phosphorylation and ATP metabolic cycles. The cDNA was isolated as part of a differential screen in cerebral granule cells in response to NMDA transmitter stimulation. Northern analysis has shown that BNPI levels increased 2-3 fold in response to NMDA treatment (Ni et al, 1994).

The *Drosophila* cDNA-34 gene was isolated from an ovarian cDNA library. Analysis by *in situ* hybridisation shows expression in the follicle cells with the RNA located at the cell cortex (A McCormack, personal communication). It is possible that this putative phosphate cotransporter is required to maintain high levels of intracellular phosphate during oogenesis to support the growth requirements of the developing oocyte. It is also conceivable that this protein could serve as a regulator of intracellular phosphate levels in a similar manner to that speculated for the rat BNPI protein. There are many signalling events involved in oogenesis for axes determination (see Chapter 1) and protein phosphorylation is a part of these signalling systems.

6.10 A Putative ATPase

Use of a PCR based technique to isolate *didum* sequence from the ATP-binding site resulted in the cloning of a PCR product from an unrelated gene. Sequence generated from this PCR product revealed similarity at both the DNA and protein levels to the *mei-1* gene from *Caenorhabditis elegans* (Clark-Maguire and Mains, 1994). The *mei-1*

gene was isolated from a series of mutants which disrupt the transition from meiosis to mitosis in the *C elegans* oocyte. Loss-of-function mutations in this gene disrupt the formation of the meiotic spindle whereas a dominant gain-of-function allele leads to abnormal mitotic cleavages after normal meiosis takes place. This genetic analysis suggests that *mei-1* is required only for meiosis.

Meiosis occurs in the developing oocyte in *Drosophila* at stage 13. It is possible that the cloned PCR product is the results from transcripts of the *Drosophila* homologue of *mei-1* being carried through oogenesis into the embryo. Another possibility is that this cloned PCR represents an ATPase from a family of related genes. The cloned PCR product can be used as a probe to isolate cDNAs from ovarian libraries to allow this *Drosophila* gene to be investigated further.

Chapter 7

Materials and Methods

7.1 General

7.1.1 Chemicals and Radioisotopes

All chemicals were of analytical grade unless otherwise specified and were purchased from Sigma Chemical Co, Boehringer Mannheim, Fison or BDH Chemicals Ltd. Radioisotopes Deoxycytidine 5'(α-^{32}P) triphosphate and Deoxyadenosine 5'(α-^{35}S) triphosphate were purchased from Amersham.

7.1.2 Buffers and Solutions

All buffers and solutions were prepared with double distilled water and autoclaved or filter sterilised unless otherwise stated. All buffers were pH adjusted and used at room temperature unless otherwise stated. Solutions used for RNA work were prepared using sterile water treated with DEPC (Diethyl pyrocarbonate) Common buffers used were:

TE: 10mM Tris-HCl, 1mM EDTA, pH 8.0

TBE: (Borate buffer for agarose gel electrophoresis) 89mM Tris-HCl, 89mM boric acid, 2.5mM EDTA, pH 8.3

TAE: (Acetate buffer for preparative gel electrophoresis) 40mM Tris-acetate, 1mM EDTA pH 7.8

MOPS (for RNA agarose gels) 20mM Na-MOPS pH 7.0, 50mM Na-acetate, 10mM EDTA

Ringer's (saline for dissection): 6.5g NaCl, 0.14g KCl, 0.2g $NaHCO_3$, 0.12g $CaCl_2$, 0.01g NaH_2PO_4 per litre, sterilised by autoclaving.

PBS (phosphate buffered saline): 8g NaCl, 0.2g KCl, 1.44g Na_2HPO_4, 0.24g KH_2PO_4 dissolved in 800ml water, pH adjusted to 7.4, made to 1 litre, sterilised by autoclaving.

SM (bacteriophage buffer): 100mM NaCl, 8.1mM $MgSO_4$, 50mM Tris-HCl pH 7.5, 0.5%w/v gelatin.

7.2 The Culturing of Bacteria and Bacteriophage

7.2.1 Growth Media

Luria Broth (LB-broth): 1% Bactotryptone, 0.5% yeast extract, 1% NaCl, pH 7.0 (Luria and Burrous, 1957).

Terrific Broth (T-broth): Bactotryptone 12g, Bacto yeast extract 24g, glycerol 4ml, dissolved in 900ml of water and sterilised by autoclaving. A 100ml sterile solution of 12.51g K_2HPO_4 and 2.31g KH_2PO_4 is then added.

2xYT: Bactotryptone 16g, Bacto yeast extract 10g, NaCl 5g, per litre, pH adjusted to 7.0 with NaOH, sterilised by autoclaving.

7.2.2 Bacterial Strains and Plasmids

Table 7.1 Bacterial Strains and Markers

E coli Strain	Markers	Comments
Y1090	F-, *Δ(lac)U169 lon- 100 araΔ139, rspL(Str'), supF, mcrA, trpC22:Tn10* (pMC9:Tet^r Amp^r)	For plating λgt libraries
NM422	F', *lacI^{q}Δ(lacZ)M15 proA$^+$B$^+$ /supE thi Δ(lac$^-$proAB)/ Δ(hsdMS$^-$ mcrB)5* ($r_K^-m_K^-$ McrBC$^-$)	For plating λFIX library
Bl21 DE3	F$^-$, *ompT [lon] hsdS$_B$*	carries prophage DE3, used for protein expression
XL1 Blue	*A1, gyrA96, thi, hsdR17, supE44*, (F', *proAB, lacI^{q}ZΔM15*)	general use for recombinant DNA work, λZAP libraries

Table 7.2 Plasmids Used in this Work.

Plasmid	Description	Features	Source
pBluescript	General cloning vector	Contains Δ15 region of *LacZ*, ampicillin resistance	Stratagene
pGEM5	General cloning vector	As for pBluescript	Promega
pRSET	Expression vector	Bacteriophage T7 Polymerase promoter system	Invtrogen

7.2.2.1 Description of pBluescript Cloning Vectors

The pBluescript series of vectors have a number of useful features which are worth describing here. They contain the *lacZ* α-peptide which contains a multiple cloning region and produces functional β-galactosidase when complemented in a strain containing the *lacZ*DM15 gene. Functional β-galactosidase results in blue colonies when plated on media containing 5-bromo-4-chloro-3-indolyl-β-galactopyranoside.(X-Gal), whereas disruption of the *lacZ* α-peptide by insertion into the multiple cloning region results in no complementation and white colonies on agar plates containing BCIG. Induction of the *lac* promoter was facilitated by the inclusion of the lactose analog isopropyl-β-D-thio-galactopyranoside (IPTG) in the agar plates. The pBluescript vectors contain the origin of replication of the filamentous bacteriophage f1, which allows production of single stranded DNA for sequencing. In addition, the universal M13 sequencing primers (see also section 7.7.2) can be used to sequence DNA cloned in the multiple cloning region and transcripts can be produced from inserts using T3 and T7 promoter sites.

7.2.3 Excision of pBluescript Phagemid from λZAP Bacteriophage

The λZAP bacteriophage vector has the useful feature of allowing the cDNA to be excised in the form of a pBluescript recombinant plasmid eliminating the need for subcloning. A λZAP plaque was selected and suspended in 0.5ml of SM solution with 20μl chloroform added followed by incubation overnight at 4°C. A culture of

XL1-Blue *E coli* was grown to an OD_{650} of ~1.0 with 0.2%w/v maltose and 10mM magnesium included in the LB-broth. A 100µl aliquot of eluted phage was added to 200ml of XL1-Blue cells with 1µl of helper phage (VCMS13) also added, then incubated at room temperature for 15 minutes followed by 15 minutes incubation at 37°C to allow adsorption and penetration of the phage DNA. To this mix was added 2ml of 2xYT followed by incubation at 37°C for 2-3 hours with shaking. The culture was then heated to 70°C for 20 minutes to kill the *E coli* cells followed by centrifugation at 4000*g* for 15 minutes. The supernatant was collected as a phagemid stock with 20µl chloroform added to prevent bacterial growth. Bacterial colonies carrying the pBluescript recombinant plasmid were prepared by adding 1µl and 10µl of phagemid stock to 200ml XL1-Blue cells with incubation at 37°C for 15 minutes. This culture was plated on to LB-agar plated containing 100µg/ml ampicillin and allowed to incubate overnight at 37°C.

7.2.4 Antibiotics

Antibiotics were used in growth media where appropriate at the following final concentrations:

Ampicillin:	100µg/ml for pBluescript and pRSET-C.
Carbenicillin	100µg/ml
Kanamycin	50µg/ml

7.2.5 Preparation of Bacteriophage Plating Cells

Plating cells were prepared by inoculating a single colony into 20ml LB-broth supplemented with 10mM $MgSO_4$ and 0.2%w/v maltose. The culture was grown to an OD_{650} of approximately 1.0. Cells, 100µl for 90mm plates, 1.5ml for 25cm x 25cm plates, were mixed with the bacteriophage and incubated at room temperature for 10 minutes followed by incubation at 37°C for 15 minutes. Top agarose (0.7% agarose in LB-broth) at 42-50°C was added to the cell/phage mix and poured immediately onto a fresh LB-agar plate. Once set, the plates were covered, inverted and incubated overnight at 37°C.

7.2.6 Libraries Used in This Work

7.2.6.1 The λFIX Genomic Library

This library was purchased from Stratagene Inc. It uses the λFIX-II bacteriophage vector which contains *Xho*I restriction endonuclease sites in the multiple cloning site. The library was constructed by ligating a size fractionated partial *Sau*3AI digestion of *Drosophila* CantonS genomic DNA to partially filled *Xho*I sites in the vector. Insert sizes vary from 9-23kb and each end is flanked by a T3 or T7 promoter.

7.2.6.2 The NB40 Early Embryonic Library

This library was constructed using the pNB40 plasmid vector (Brown and Kafatos, 1988). First strand cDNA was synthesised using a oligonucleotide primer with 12 T residues and a unique 14 nt 5' end, then tailed with 10-15 G residues using terminal transferase. The cDNA was annealed with two vector fragments, one containing a poly C overhang and the other with sequence complementary to the 14 nt part of the primer. The annealing reaction was diluted and ligated to promote recircularisation, then used to transform *E coli*.

7.2.6.3 The Ovarian λgt22 Library

This library was constructed by Stroumbakiss et al (1994). Directional cloning was carried out to give unique *Eco*RI and *Sal*I restriction sites at the cDNA 5' end and a unique *Not*I site flanking the Poly(A) tail. Initial complexity was 1.5 x 10^6 independent recombinants.

7.2.6.4 The Ovarian λZAP library

This library was constructed by Stratgene Inc for the Jan laboratory (Jongens et al, 1992) and uses the λZAP vector which allows recombinant cDNAs to be excised as pBluescript plasmids.

7.3 Transformation of *Escherichia coli*

The *E coli* strains XL1-Blue or BL21 DE3 were made competent using either of the following methods. Fresh competent cells were prepared for each transformation.

7.3.1 Cold $CaCl_2$ Method

The method of Lederberg and Cohen (1974) was used as follows: Mid-log phase cultures were pelleted by centrifugation, then resuspended in one-half volume cold 0.1M $MgCl_2$ and left on ice for 15 minutes. The cells were pelleted by centrifugation, then resuspended in one-half volume cold 0.1M $CaCl_2$ and incubated on ice for 30 minutes. The cells were again pelleted by centrifugation, then resuspended in 1/20th volume 50mM $CaCl_2$, 15% glycerol and used immediately. Transformation was carried out by mixing approximately 50ng of DNA with 100µl of cells followed by incubation on ice for 30 minutes. The transformed cells were heat shocked at 42°C for 2 minutes in a water bath. LB-broth (900µl) was added and the cells allowed to grow at 37°C for 1-1.5 hours on a shaker. The culture was then spread onto antibiotic plates to select for transformants.

Transformation efficiency was measured by transforming 10µl of a 100ng/ml solution of the plasmid pBluescript KS-.

7.3.2 Electroporation

7.3.2.1 Preparation of Cells

A 5ml overnight culture of *E coli* was used to inoculate 300-500ml of fresh LB-broth at a 1:100 ratio. The culture was incubated at 37°C with shaking until growth reached mid-late log phase, then chilled on ice for 15 minutes. The cells were collected by centrifugation at 4000rpm (Sorvall RC-5B, GSA rotor). The cell pellet was resuspended in a volume of cold sterile distilled water equal to that of the LB-broth initially used (a low ionic buffer, eg. 1mM Hepes pH 7 can be substituted for water at this point). The cells were again collected by centrifugation, then resuspended in about 20ml of cold sterile water as before and transferred to chilled 30ml glass Corex tubes. The cells were washed twice more with about 20ml of cold sterile water and

finally resuspended in 2ml of water or in 2ml of 10%v/v glycerol if aliquots were to be stored. The cells were now ready for electroporation.

A microfuge tube was pre-cooled on ice before 40µl of cells were added followed by 1-2µl of DNA solution. A Gene-Pulser (BioRad) was set at 2500V, 25µF and 200 ohms. The DNA-cell mix was added to a precooled electroporation cuvette, placed into the pre-cooled safety slide, pushed between the contacts and pulsed immediately. A pulse time of 5-6msec was a typical result and indicated successful transformation. The cuvette was removed from the safety slide and immediately 1ml of LB-broth supplemented with 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$ and 20mM Glucose was added and mixed with a pasteur pipette. The culture was transferred to a microfuge tube and incubated at 37°C with shaking for 1 hour to allow expression of the antibiotic resistance gene, then plated on LB-agar plates containing the appropriate antibiotic.

7 4 *Drosophila melanogaster*

7.4.1 Strains used in This Work

The wild type strain used throughout this work was OregonR (Lindsley and Zimm, 1992). Table 7.3 lists the developmental stages of OregonR at 25°C. Other strains of flies used in this work are noted and referenced in the relevant results section.

Table 7.3 Life Cycle of *Drosophila melanogaster*

Hours	Days	Developmental Stage
0	0	Fertilisation and fusion of pronuclei
1.5	0	Preblastoderm stage. Migration of cleavage nuclei and pole cell formation
3	0	Blastoderm stage. Migrated nuclei form cells
3.5	0	Gastrulation begins
6 - 8	0	Segmentation begins
18	0	Larval differentiation nears completion
24	1	Hatching from egg. Onset of first larval instar
48	2	First moult. Second larval instar begins
72	3	Second moult. third larval instar begins
120	5	Puparium formation with white puparium
122	5	Puparium darkens
124	5	Puparium moult
132	5	Pupation. Eversion of imaginal discs
216 - 240	9 -10	Eclosion of adult from pupal case

7.4.2 Maintenance of Fly Stocks

Stocks were maintained at 18°C on Staffan cornmeal food consisting of cornflour (250g), sugar (500g), yeast pellets (175g) and agar (100g) dissolved in distilled water to a final volume of 10 litres. The food was boiled and allowed to cool to approximately 60°C before being poured into bottles or vials. A fungicide, Nipagin, was added to a final concentration of 4.5 μg/L. To prevent mite contamination, strips

of Whatman 3M paper soaked in 3%v/v benzyl benzoate (in ethanol) and air dried were placed on top of the cornmeal food.

Adh food consisted of 100g dried flake yeast, 100g brown sugar, 16g agar and Nipagin at 4.5mg/ml final concentration in one litre.

7.4.3 Collection of Sexed Adult Flies

Adult flies of various ages were anaesthetised with diethyl ether and sexed according to external morphology; presence of sex combs on the male forelegs or abdominal pigmentation.

7.4.4 Collection of Virgin Females

All existing adult flies were tipped from the bottle which was then returned to the 25°C incubator. Newly eclosed flies were examined and sexed early in the morning and late in the day and placed into separate vials.

7.5 DNA Manipulation

7.5.1 Enzymatic Reactions

Restriction endonucleases were supplied by Bestheda Research Laboratories (BRL), New England Biolabs (NEB), Boehringer Mannheim, Promega or Northumbria Biological Laboratories (NBL). Digestion was carried out as recommended by the manufacturer using buffers supplied with the enzymes in volumes ranging from 10µl to 100µl, with typically 5 units of endonuclease per µg of DNA.

T4 Polynucleotide kinase, DNA Polymerase I (Klenow fragment), T4 Polymerase, calf intestinal alkaline phosphatase (CIAP) and T4 DNA ligase were supplied by Bestheda Research Laboratories (BRL) or Boehringer Mannheim and used as specified. Mung bean nuclease was supplied by BRL and used as specified.

DNA Polymerase for use in PCR reactions was supplied by BRL, Promega or Dynazyme

RNase was obtained from Sigma Chemical and prepared as a solution of 10mg/ml as described in Sambrook et al (1989) which was boiled for 20 minutes to remove contaminating DNase activity.

7.5.2 Dephosphorylation of Plasmid Vectors

Dephosphorylation with calf intestinal alkaline phosphatase (CIAP) was carried out to facilitate cloning of DNA fragments into the vectors by preventing vector religation. The plasmid was restricted with the desired restriction endonuclease, then CIAP and 0.5x CIAP buffer was added. Incubation was continued at 37°C for a further 30 minutes then at 55°C for 30 minutes. The plasmid DNA was extracted 2 or 3 times with phenol chloroform and the DNA precipitated with ethanol as described below. The plasmid was redissolved in a volume of TE to give a concentration of about 100ng/µl.

7.5.3 Phenol Extraction and Alcohol Precipitation

Phenol-chloroform extraction is a common procedure for the removal of proteinacious material from DNA solutions. Phenol was equilibrated with Tris-HCl pH 8 as described by Sambrook et al (1989) then mixed in a ratio of 25:24:1 phenol : chloroform : iso-amyl alcohol respectively and stored at -20°C in 1.5ml portions.

DNA was extracted by the addition an equal volume of phenol-chloroform and mixed by vortexing, or if shearing of DNA was to be avoided by repeated, gentle, inversions. The phases were separated by centrifugation and the organic phase was carefully removed by pipettor, then a second extraction was performed, with further extractions performed if required. A final extraction with an equal volume of 49:1 chloroform : iso-amyl alcohol was performed to ensure complete removal of phenol. Precipitation was carried out by adding 1/10th volume 3M Na-acetate pH 5.2 with either 2 volumes ethanol or 1 volume isopropanol, at -20°C for at least one hour. The DNA was pelleted by centrifugation in a microfuge for 20 minutes. Additional salts were removed by rinsing of the pellet with 70%v/v ethanol with a brief centrifugation and careful removal of the rinse solution. The DNA pellet was dried under vacuum and then redissolved in a solution of choice for further manipulation.

7.5.4 Agarose Gel Electrophoresis

7.5.4.1 Molecular Weight Standards

All agarose gels were run with one of the following molecular weight (MW) standards:
BRL 1 kilobase ladder (1 kb ladder)
200- 500 ng of marker was used in a standard gel lane.

7.5.4.2 Standard Gel Electrophoresis

Horizontal slab gels were used for routine separation of DNA fragments in a submarine apparatus (Sambrook et al 1989). Gel-agarose concentrations varied from 0.4%w/v to 1.2%w/v in TBE buffer depending on the size of the fragments to be separated. Agarose from Boehringer Mannheim was used.

7.5.4.3 Preparative Gel Electrophoresis

Submarine gels were prepared with BRL ultrapure agarose in TAE buffer and run as for standard gels, see above.

7.5.4.4 Recovery of DNA fragments

DNA fragments were extracted from preparative agarose gels using kits obtained from Qiagen. In most cases Qiaex was used but at times the Qiaquick column method was employed. Qiaex uses a NaI based solution (QX1) and heating to 55°C to dissolve the gel slice with the DNA then being bound to a silica based particle system. Impurities are washed away with QX1 and the ethanol based PE wash buffer, the silica particles dried briefly, then the DNA eluted with TE solution, with a 55°C elution step used for DNA molecules larger than 5kb. The Qiaquick system is very similar but uses a microfuge column in place of the loose silica particles in the Qiaex kit. It was felt that Qiaex produced better, more consistent yields than the Qiaquick system.

7.5.5 Radiolabelling of DNA

DNA was labelled with α^{32}P dCTP or digoxigenin as follows:

7.4.5.1 End labelling of 1 kilobase ladder

For Southern blots it is useful to have reference markers as for normal ethidium bromide stained gels. This is most easily accomplished by end filling the 'sticky ends' of the standard 1 kilobase ladder (BRL, Life Technologies Inc). Approximately 200ng of 1 kilobase ladder was radiolabelled in a 40µl reaction mix with 1unit of Klenow enzyme (Boehringer Mannheim), 50mM each of dATP, dGTP and dTTP, and 1µl of α^{32}P dCTP. The reaction mix was incubated at 37°C for 30 minutes. Unincorporated radiolabel was removed by use of a Pharmacia Sephadex Nick column.

7.5.5.2 Random Prime Labelling

To generate probes for Southern and Northern hybridisations, DNA was radiolabelled using a random priming kit based on the method of Feinberg and Vogelstein (1983).

Approximately 50-200ng of DNA was denatured first by boiling then immediately plunged into ice. The DNA was labelled in a reaction mix containing, 10µl OLB buffer, 5µl of α^{32}P dCTP, 1µl of 20mg/ml BSA and 5 units of Klenow enzyme in a final volume of 50µl by incubating at 37°C for 30 minutes. OLB consists of a ratio of 2:5:3 of components A, B and C which are described below.

Solution A:	950µl of solution 'O'
	18µl of β-mercaptoethanol
	25µl each of 20mM dATP, dGTP and dTTP
Solution B:	2000mM HEPES pH 6.6
Solution C:	Hexanucleotides 90 OD/ml (Boehringer Mannheim or Pharmacia)
Solution O:	125mM $MgCl_2$
	1250mM Tris-HCl pH 8

Later stages of the project used the Pharmacia Ready-To-Go kit which provided hexanucleotides, dATP. dGTP, dTTP, Klenow polymerase and buffer salts as a lyophilised mix that was redissolved in 20µl of sterile distilled water. The DNA was

denatured in a 25µl volume as before, added to the reaction mix tube along with 5µl (50mCi) of α-^{32}PdCTP and then incubated at 37°C for 1 hour.

Unincorporated radiolabel was removed using a Pharmacia Nick Column. These columns use the Sephadex G-50 for gel filtration allowing the larger labelled DNA molecules to pass through the column while retaining the unincorporated nucleotides. The column was prepared by tipping off the storage solution, rinsing with one column volume of TE followed by one column volume of TE run through the column. The labelling reaction was pipetteted directly on to the top support surface of the gel bed and washed in with 400µl of TE. The material collected at this stage contained very little labelled DNA and was discarded while a second wash of 400µl eluted the bulk of the labelled DNA.

Efficiency of incorporation was not determined other than by rough estimation from hand held monitors.

7.5.5.3 Labelling of DNA with Digoxigenin

DNA, 500-1000ng in a volume of 15µl was denatured by boiling in a waterbath for 5 minutes followed by plunging into ice. A Boehringer Mannheim labelling kit was used with a labelling reaction consisting of denatured DNA, 2µl hexanucleotides, 2µl digoxigenin labelling mix and 1 unit of Klenow enzyme. Labelling was carried out for at least 60 minutes at 37°C. Unincorporated label was removed by precipitation of the labelled DNA with 1µl 0.5M EDTA, 1.5µl 10M LiCl and 2.5 volumes of ethanol at -70°C for 15 minutes. The labelled DNA was collected by centrifugation, washed with 70%v/v ethanol and air dried. The labelled DNA was redissolved in the appropriate buffer and denatured by boiling and plunging into ice prior to use.

7.6 DNA Preparation

7.6.1 Genomic DNA Preparation

Adult flies were collected by etherising, sorted as required and placed in a vial for several hours to void yeast. The flies were then transferred to a microfuge tube and immersed in liquid nitrogen briefly to render them motionless. The flies were homogenised (Kontab homogeniser) in Flybuffer (5%w/v sucrose, 80mM NaCl,

100mM Tris-HCl pH 8.0, 50 mM EDTA pH 8.0, 0.05%w/v SDS) added at a ratio of 0.5ml per 50 flies. The homogeniser rod was rinsed with an equal volume of Flybuffer, then the homogenate was frozen for 15 minutes at -70°C. The homogenate was thawed on ice and RNase added to a final concentration of 10µg/ml, then transferred to a 70°C waterbath and incubated for 30 minutes. Potassium acetate (3M stock) was added to a final concentration of 160mM with incubation on ice for 30 minutes. Debris and insoluble material was separated by centrifugation in a microfuge for 5-10 minutes and the supernatant then transferred to a fresh microfuge tube. The solution was phenol/chloroform extracted twice and chloroform extracted once, then precipitated with 0.75 volumes of isopropanol. The precipitate was collected by centrifugation for 15 minutes in a microfuge and washed with 70%v/v ethanol, then air dried. The pellet was dissolved in a suitable volume (dependent on the number of flies used) of TE overnight at 4°C.

7.6.2 Plasmid Preparation

7.6.2.1 Small Scale

The method of Birnboim and Doly (1979) was used to prepare 2-5µg of plasmid DNA. Colonies were selected and inoculated into 5ml LB-broth plus antibiotic then incubated overnight at 37°C. 1.5ml of culture was centrifuged in a microfuge for 30 seconds with the cell pellet resuspended in 100µl of Solution I (40mM Tris-HCl pH 8.0, 10mM EDTA pH 8.0, 50mM Glucose) followed by incubation at room temperature for 5 minutes. This step was followed by the addition of 200µl of Solution II (1% SDS, 200mM NaOH) with mixing by gentle rolling of the tube to prevent shearing of genomic DNA, then incubated on ice for 10 minutes after which 150µl of Solution III (3M Sodium acetate pH 4.8) was added with gentle mixing as before and incubation on ice for 10 minutes. Cell debris was pelleted by centrifugation in a microfuge for 5 minutes with the supernatant being transferred to a fresh tube. The DNA was precipitated by the addition of 1ml ethanol at room temperature, mixed, then centrifuged in a microfuge for at least 5 minutes. The supernatant was removed and the pellet washed with 80%v/v ethanol, re-centrifuged, then dried in a speedvac desiccator. The DNA was dissolved in 50µl of TE buffer.

7.6.2.2 Medium Scale, LiCl method

This procedure was used for preparation of 50-200µg of plasmid DNA from 100-200 ml of culture. It was developed from a small-scale procedure as described by Ausubel

et al (1989), incorporating suggestions from G Albertson (personal communication). 100-200 ml of culture was grown overnight in L-broth plus antibiotic. The cells were pelleted by centrifugation at 5000 rpm for 10 minutes, then resuspended in 3ml Solution I (as for miniprep procedure), and left to stand at room temperature for 5 minutes. The cells were lysed by the addition of 6ml of Solution II (as for small-scale prep), mixed gently and allowed to stand at room temperature for 10 minutes. This was followed by the addition of 4.5ml of 3M potassium acetate pH 4.8 with gentle mixing and incubation on ice for 45 minutes. Cell debris was separated by centrifugation at 8000 rpm for 10 minutes. The supernatant was transferred to a fresh tube and centrifuged at 15 000 rpm for 30 minutes to pellet the last traces of cell debris. The supernatant was precipitated by the addition of 1 volume of isopropanol at room temperature for at least 10 minutes. The precipitate was collected by centrifugation at 15 000 rpm for 30 minutes, dried, then redissolved in 0.5ml TE and transferred to a microfuge tube. 20µl of 10mg/ml DNase free RNase was added and the tube was incubated at 60°C for 15 minutes. Next 375µl of 7M LiCl was added, incubated on ice for 20 minutes, then microfuged for 15 minutes. The supernatant was transferred to a fresh tube then phenol/chloroform extracted followed by a chloroform extraction. The DNA was precipitated by the addition of 1 volume of isopropanol for 30 minutes at room temperature or overnight at -20°C. The DNA was pelleted by microfuging for 15 minutes, washed with 80%v/v ethanol, dried, then redissolved in 200-500µl TE.

7.6.2.3 Use of DNA Preparation Kits for Plasmid Preparation

The advent of kit technologies allows DNA to be readily prepared from a variety of sources. Kits from Qiagen and Promega (Wizard™) were used to prepare high purity plasmid DNA. Both systems began with an alkaline lysis procedure as described above then utilised the manufacturer's proprietary matrix systems to preferentially bind DNA. Impurities were removed with supplied wash buffers and the DNA eluted with a low salt (TE) solution or water.

7.6.2.4 Preparation of Single-Stranded DNA

The pBluescript plasmids contain the origin of replication for the f1 filamentous bacteriophage allowing production of single-stranded DNA for sequencing purposes when the plasmid-containing bacterium is infected with a suitable helper phage. The plasmid must be contained in a host suitable for helper phage infection, in this case the strain used was XL1-Blue and the helper phage used was VCSM13.

A 10-20ml culture was inoculated from an overnight culture of the strain carrying the desired plasmid and then incubated for 30 minutes with vigorous shaking at 37°C. The culture was then infected by the addition of VCSM13 helper phage at a ratio of 10-20 phage particles per cell. Incubation was continued for a further 6 to 8 hours with vigorous shaking. The bacterial cells were separated by centrifugation at 12 000rpm for 10 minutes, - the phage particles remained in the supernatant. Phage particles were precipitated by the addition of 0.25 volume of 20% PEG/2.5M NH_4Cl, mixed and incubated on ice for 30 minutes and then collected by centrifugation at 12 000rpm for 15 minutes. The phage particles were resuspended in 400µl TE buffer then the protein coat was stripped from the DNA by the addition of 400µl of chloroform:isoamyl alcohol (49:1) with vortexing for 1 minute followed by centrifugation at 12 000rpm for 5 minutes. The upper phase was transferred to a fresh tube and extracted with 400µl of phenol/chloroform with vortexing for 1 minute and centrifuging for 5 minutes. This step was repeated one or two times more until no material was visible at the interface. The upper aqueous phase was transferred to a fresh tube and extracted with 400µl of chloroform as above. The aqueous phase was transferred to a fresh tube and precipitated by the addition of 0.5 volume of 7.5M ammonium acetate and 2 volumes of ethanol and incubated at -20°C for at least 30 minutes. The single-stranded DNA was pelleted by centrifuging for 20 minutes in a microfuge, then rinsed with 80%v/v ethanol. The single-stranded DNA was redissolved in 20-50µl of TE buffer, and quantified by estimation from agarose gel electrophoresis.

7.6.2.5 Preparation of Bacteriophage lambda DNA

An inoculum of 100µl of 10^{10} pfu/ml of bacteriophage was added to 2ml of late log phase of the appropriate *E coli* strain grown in L-broth supplemented with 10mM $MgSO_4$ and 0.2%w/v maltose. The infected bacterial culture was then added to 20ml LB-broth also containing 10mM $MgSO_4$ and 0.2%w/v maltose. Incubation was carried out at 37°C with vigorous shaking for 5 to 7 hours whereupon lysis of the bacteria was observed. The bacteria were completely lysed by the addition of 300µl of chloroform with shaking continued for a further 10 minutes. The cell debris was removed by centrifugation for 10 minutes at 7000rpm (Sorvall RC-5B, SS34 rotor). Solutions of DNase and RNase, 20µl of 20mg/ml and 50µl of 10mg/ml respectively, were added to the clarified supernatant with incubation at 37°C for 1 hour to digest the bacterial nucleic acids.

Next, 4ml of TES buffer (2.5%w/v SDS, 250mM Tris-HCl pH 7.5, 25mM EDTA) was added and the mixture was incubated at 70°C for 30 minutes then allowed to cool to room temperature for 10 minutes. This step was followed by the addition of 5ml of 4.8M K-acetate (pH 4.8), mixed, then incubated on ice for 15 minutes. The potassium-SDS precipitate was removed by centrifugation for 15 minutes at 12 000rpm (Sorvall RC-5B). Isopropanol at 0.7 volumes was added to the supernatant with incubation at room temperature for 10 minutes. The phage particles were collected by centrifugation at 15 000rpm for 30 minutes. The supernatant was decanted off and the pellet was dried then dissolved in 500µl of TE. This solution was then extracted twice with phenol/chloroform and once with chloroform followed by precipitation with 0.7 volumes of isopropanol. The precipitate was washed with 70%v/v ethanol, dried and redissolved in 100µl TE.

7.7 DNA Sequencing

DNA was sequenced by two methods. Initially manual sequencing was performed until a departmental automated sequencer became available, whereupon template DNA was prepared and supplied to the sequencing facility. All sequencing was carried out using the cloning vectors pBluescript (Stratagene) or pGEM5 (Promega Corp). The universal forward and reverse primers were used for all sequencing reactions.

7.7.1 Manual Sequencing

The dideoxy chain termination method of Sanger et al (1977) was used in the form of a Sequenase^{R} version 2.0 kit (United States Biochemical Corporation) or a TaqTrack kit (Promega). Double stranded DNA was sequenced as described in the kit support manual, which labelled the sequenced DNA with α^{35}S dATP. The sequencing reactions were run on a 6% polyacrylamide/7Murea gel in TBE buffer. Gels were electrophoresed at a constant power setting of 40W for 40cm long gels or 65W for 60cm gels. The gel was fixed in a solution of 10%v/v methanol/10%v/v acetic acid for 30 minutes. The gel was transferred onto Whatman 3MM paper and dried under vacuum for at least 1 hour at 85°C. The dried gel was exposed to X-ray film (GRI) at 4°C for 1-3 days depending on the amount of radioactive signal present.

7.7.2 Automated Sequencing

The Applied Biosystems 373A sequencer also uses the dideoxy chain termination method of Sanger et al (1977). However, DNA bands are detected by fluorescent dyes which are either incorporated in the primers or into the dideoxy nucleotides. Templates were supplied for sequencing initially on a 373A sequencer then for a 377A machine with results provided in the form of a computer file.

7.7.2.1 Dye-primer Sequencing

A primer based on the universal M13 -20 sequence was synthesised by Oswel Ltd with fluorescent dyes incorporated at the 5' terminus of the primer. Nucleotide and dideoxynucleotide stock solutions were prepared as shown in Tables 7.4 and 7.5 using Pharmacia chemicals.

Table 7.4 Nucleotide Mixes for Dye-Primer Sequencing (volumes in µl)

dNTPs at 5mM	A	C	G	T
dATP	10	40	80	80
dCTP	40	10	80	80
dGTP	40	40	20	80
dTTP	40	40	80	20
water	270	270	540	540
Total	400	400	800	800

Table 7.5 Dideoxynucleotide Mixes for Dye-Primer Sequencing (volumes in µl)

ddNTP (5mM)	volume	water	Total
ddATP	240	160	400
ddCTP	120	280	400
ddGTP	40	260	760
ddTTP	400	400	800

The dNTP and ddNTP stocks were used at a ratio of 1:1 although it should be noted that ratios of dNTPs to ddNTPs can be varied to suit different templates as follows 1.5:1, 1:1.5, 1:2, 2:1, 3:1. CSB buffer is 400mM Tris-HCl pH 8.9, 100mM $(NH_4)_2SO_4$, 25mM $MgCl_2$.

A master mix was prepared for each base as follows:

For A and C

1µl	5 x CSB buffer	x	No of templates
1µl	dNTP/ddNTP	x	No of templates
1µl	dye-primer at 3.2pmol/ml	x	No of templates
1µl	Taq polymerase mix	x	No of templates

After thermal cycling all four individual reactions for a particular template were transferred to a 0.5ml microfuge tube containing 80µl 95%v/v ethanol and 1.5µl Na-acetate pH 5.2. These mixes were incubated on ice for at least 10 minutes and the sequenced DNA then collected by centrifugation for 15 minutes at 4°C. The ethanol phase was removed by pipettor and the pellet washed with 80%v/v ethanol, then dried. The dried pellet was supplied to the automated sequencer operator for loading. Prior to loading the pellet was dissolved in 5µl of formamide/50mM EDTA and heated to 90°C for 2 minutes followed by immediate plunging into ice.

7.7.2.2 Cycle Sequencing

To determine DNA sequence with a gene specific or non dye-labelled primer, the Perkin-Elmer Prism™ Ready reaction DyeDeoxy™ Terminator system was used. This system incorporates the fluorescent dyes read by the automated sequencer into the dyedeoxy terminators. Double stranded templates were used in a reaction which consisted of 9.5µl terminator mix, 1µl of primer at 3.2pmol/µl, approximately 0.5µg of template DNA in a total volume of 20µl. The reaction mixture was overlaid with mineral oil and cycle sequencing was performed using the following conditions:

96°C	30 seconds
50°C	15 seconds
60°C	4 minutes

25 cycles

Unincorporated dye-terminators were removed by phenol/chloroform extraction as follows. The sequencing reaction volume was increased to 100µl by the addition of 80µl of water and the mineral oil overlay was removed by extracting with 100µl

chloroform. The aqueous phase was extracted twice with 100µl phenol:water:chloroform at a ratio of 68:18:14 then precipitated by the addition of 15µl Na-acetate pH 5.2 and 300µl of 95%v/v ethanol. This mix was centrifuged for 15 minutes, the ethanol phase removed, then washed with 70%v/v ethanol and dried. The reaction was then supplied to the automated sequencer operator as before.

Further improvements in automated sequencing technology has resulted in a modified dye-terminator method. A sequencing reaction mix of 8µl terminator mix, 1µl of primer at 3.2pmol/µl and template at 250-500ng for double stranded plasmid or 30-90ng for PCR products was made to 20µl final volume and cycle sequenced as before. The sequenced DNA was recovered by adding the reaction mix to 2µl of Na-acetate pH 5.2 and 50µl ethanol followed by incubation on ice for at least 10 minutes. The DNA was collected by centrifugation for 20 minutes and the pellet washed with 250µl 70%v/v ethanol, dried, then supplied to the sequencer operator.

7.7.3 Sequence Analysis

Sequence analysis was carried out using University of Wisconsin Computer Genetics Group (GCG) software (Devereux et al, 1984) using VMS VAX and Unix operating systems. Sequence editing was performed using the SEQED program, and sequence alignment was performed using the GELSTART set of programs. Manual sequences were input by hand while sequences generated by the automated sequencer were edited with GeneJockey II (BioSoft) or Factura (Applied Biosystems Inc) software before being transferred to GCG readable files.

7.8 The Polymerase Chain Reaction

Polymerase Chain Reactions (PCR) were carried out using a Hybaid thermal cycler. Details of cycles used are given in the appropriate Chapters. The standard reaction volume was 50µl in sterile 0.5ml tubes with 30-50µl of mineral oil added to prevent evaporation. Pharmacia deoxynucleotide triphosphates (dNTPs) were used for all reactions. Actual concentrations of dNTPs and primers are given in the appropriate Chapter. The standard PCR buffer used was 50mM KCl, 20mM Tris-HCl pH 8.3, 2mM $MgCl_2$ and 1mg/ml gelatin.

7.9 Oligonucleotides

Oligonucleotides were obtained from Oswel or Perkin Elmer and were supplied dissolved in ~1ml of sterile distilled water or ~1ml 20% acetonitrile in water respectively. Synthesis scales were either 40nmol or 200nmol using cyanoethyl phosphoramidite chemistry (Gait, 1984).

7.10 RNA Preparation

Total RNA was prepared using one of two methods; earlier preparations were done with the RNeasy system from Qiagen while later preparations were made using the Trizol solution from BRL. Ovary and testis RNA was prepared from gonadal tissue dissected from etherised adult flies. During the dissection procedure, the tissue was collected in a microfuge tube held in dry ice. When sufficient tissue had been collected the lysis buffer (Qiagen) or Trizol solution was added as described in the manufacturer's protocol and the tissue homogenised with a Kontab hand-held homogeniser. The RNeasy procedure utilised a silica-gel based microfuge column to bind the RNA with wash steps to remove impurities. The RNA was eluted from the column in 2 lots of 50µl using DEPC treated water supplied with the kit and heated to 65°C.

The Trizol system, a solution of phenol and guanidium isothiocyanate, is based on the method of Chomczynski and Sacchi (1987). Tissue, up to 100mg, was homogenised in 0.5ml Trizol solution with a further 0.5ml of Trizol then added. The homogenate was incubated at room temperature for 5 minutes then 200µl of chloroform was added, the solution shaken and incubated at room temp for 3 minutes whereupon the aqueous and organic phases were separated by centrifugation at 12000*g* at 4°C. The aqueous phase was removed and the RNA was precipitated by the addition of 0.7 volumes of isopropanol at room temperature for 10 minutes followed by centrifugation at 12 000*g* for 15 minutes at 4°C. The supernatant was removed and the RNA pellet washed with 1ml of 70%v/v ethanol with centrifugation at 7500*g* for 5 minutes. The ethanol wash was removed and the pellet air dried then dissolved in DEPC sterile water.

Yield and quality were estimated by standard TBE/agarose gel electrophoresis although prior to electrophoresis the gel apparatus was soaked in a solution of SDS/NaOH and thoroughly rinsed with distilled water to remove any contaminating RNases.

7.11 Hybridisation Techniques

The following solutions were used during hybridisation procedures:

SSC: 20x: 3M NaCl, 0.3M tri-sodium citrate

SSPE 20x 3.6M NaCl, 0.2M sodium phosphate, 0.02M EDTA pH7.7

Denaturing solution: 1.5M NaCl, 0.5M NaOH

Neutralising solution: 1.5M NaCl, 0.5M Tris-HCl pH 7.2, 0.001M EDTA

Denhardt's solution : (100x): 2%w/v bovine serum albumin, 2%w/v Ficoll™, 2%w/v polyvinylpyrolidone

Prehybridization Solution: 5x SSPE, 5x Denhardt's, 0.5%w/v SDS

7.11.1 Southern Blotting

The protocol for transferring DNA to Hybond-N and Hybond-N+ membranes (Amersham) was that supplied with the membranes, which is outlined below.

7.11.1.1 Alkali Transfer

DNA was separated in agarose gels using standard methods previously described. After electrophoresis, the gel were soaked in a 0.25M HCl solution for approximately 30 minutes to depurinate and improve transfer of large fragments. The gel was rinsed with water. Transfer of DNA to the membrane (Hybond-N+ only) was performed using 0.4M NaOH for at least 3 hours on a capillary blot (Southern, 1975). After transfer, the membrane was rinsed in 2xSSC. The membrane was then prehybridised at 65°C for a minimum of 1 hour and hybridised with denatured probe.

7.11.1.2 SSC Transfer

The gel was electrophoresed and treated with HCl as above, rinsed in distilled water, then soaked in denaturing solution until the running dyes had returned to their original colours. The gel was then soaked in 2 changes of neutralising solution for 15 minutes each. A capillary blot was set up as above and transfer was carried out with 20xSSC to Hybond-N or Hybond-N+. After transfer, the membrane was rinsed in 2xSSC then the DNA was fixed by exposure to 254nm and 365nm uv light for 10 minutes. Prehybridisation and hybridisation were carried out as above.

7.11.2 Northern Blots

All solutions were prepared with DEPC water. RNA was separated on 0.7%w/v to 1.0%w/v denaturing agarose gels in a MOPS buffering system. Agarose was dissolved in 10ml 10xMOPS and 73ml of water and cooled to about 55°C, then 17ml of 37%v/v formaldehyde was added, the solution mixed, and poured immediately into a gel tray. Gels were run in a 1xMOPS buffer.

RNA samples were incubated at 65°C for 5 minutes in the following buffering system: RNA 5µl, formamide 12.5µl, 10xMOPS 2.5µl, formaldehyde 4µl, then chilled immediately on ice. Prior to loading, 2.5µl loading solution (50%v/v glycerol, 0.1mg/ml bromophenol blue) was added.

After electrophoresis the RNA was transferred to Hybond-N or Hybond-N+ using the SSC transfer protocol.

7.11.3 Colony and Plaque Lifts

Bacteriophage were plated as described in section 7.2.5 whereas recombinant plasmids were screened by patch plating colonies onto agar plates. The DNA was transferred to the Hybond-N membrane by overlaying the membrane on to the agar surface for a minimum of 1 minute as described in the Amersham protocol. The membranes were placed plaque/colony side up for 7 minutes in a dish of denaturing solution. The membranes were transferred in a dish of neutralising solution for 3 minutes followed by a second dish of neutralising solution for a further 3 minutes, then rinsed in a dish of 2xSSC. The DNA was fixed by exposing to uv light for 10 minutes as for Southern blots. The membranes were then ready for prehybridisation and hybridisation.

7.11.4 Prehybridisation, Hybridisation and Washing

Prehybridisation was carried out using the Amersham protocol. Membranes were prehybridised at 60°C for 1 hour (42°C for 4 hours for Northerns) in 25ml (50ml for larger membranes) of prehybridising solution containing 1mg/ml of non-specific sonicated salmon sperm DNA (Sigma) prepared as described in Sambrook et al (1989). The probe, prepared as described in section 7.5.5.2, was denatured in a

heated block at 110°C for 5 minutes, then plunged into ice. Hybridisation was carried out by adding probe directly to the prehybridisation solution. If a hybridisation solution was being reused, the prehybridisation solution was removed and stored and the probe solution added. Hybridisation was carried out at 65°C for at least 12 hours (42°C for for Northerns).

A wash procedure described in the Amersham protocol was used for all hybridisations. The probe solution was carefully removed. Approximately 100ml of 2xSSC/0.1% SDS was added and gently shaken for 10 minutes at room temperature. A second, identical wash was carried out. Approximately 100ml of 1xSSC/0.1% SDS was added with gentle shaking and incubation at 65°C for 15 minutes. Membranes were then wrapped in Saranwrap™ and exposed to X-ray film (GRI) at 70°C in cassettes containing one or two intensifying screens for times varying from a few hours to several days.

7.11.5 Stripping of Membranes

To remove a probe from a membrane to allow use of a fresh probe, the membrane was boiled in a solution of 0.1%w/v SDS for 5 minutes. Membranes were checked for absence of probe by exposing to X-ray film, as described above, overnight.

7.12 SDS-PAGE Electrophoresis of Proteins

Proteins were analysed by discontinuous SDS polyacrylamide gel electrophoresis (SDS PAGE) as described in Sambrook et al (1989) using an Atto system. Gels of 12cm x 12cm x 1mm thick were prepared in the casting apparatus at final acrylamide concentrations ranging from 7.5% to 12.5%. Once poured, the separating gel was overlaid with a buffer saturated butanol solution to allow an even interface on setting. Stacking gels were at a final acrylamide concentration of 4%.

Gels were run at a constant 100-150V for 4-5 hours or overnight at 30-60V. Once run, the glass plates were carefully separated, the stacking gel removed, and the separating gel transferred to a staining box. Gels were initially fixed in a solution of 12%w/v trichloroacetic acid for 1 hour, then prestained with 100ml of a 1:5 dilution of ProBlue (Integrated Separation Systems) solution A. Staining was carried out with the ProBlue solution (20ml methanol, 16ml Solution A, 64ml water, 1.6ml Solution B) for a minimum of 4 hours but typically overnight. Gels were destained with a 10% v/v methanol solution. When sufficiently destained, gels were photographed on a light box using 35mm Ilford PanF ISO 18° negative film through a red filter, then dried between 2 sheets of BioRad cellophane prewetted with distilled water at 75°C on a Hoefer gel-dryer for 2-3 hours.

Table 7.6 Protein Molecular Weight Markers

Protein	BioRad Prestained*	Sigma Prestained	Sigma Wide range
Rabbit muscle myosin			205
β-galactosidase (*E coli*)		123	116
Phosphorylase-B (rabbit)	104		97
Fructose-6-phosphate kinase (rabbit)		89	84
Bovine serum albumin	82		66
Pyruvate kinase		67	
Glutaraldehyde dehydrogenase (bovine)			55
Ovalbumen (chicken)	48.3	50	45
Lactate dehydrogenase (rabbit)		37.5	
Glyceraldehyde-6-phosphate dehydrogenase (rabbit)			36
Triose phosphate kinase (rabbit)		34	
Carbonic anhydrase (bovine)	33.4		29
trypsinogen			24
Trypsin inhibitor (soybean)	28.3		20
Lyoszyme	19.4		
Lactalbumin (milk)			14.2
Aprotinin (bovine)			6.5

* These molecular weights vary depending on batch

7.13 Western Blotting

7.12.1 Transfer from SDS-PAGE

Proteins were transferred from SDS-PAGE gels to Hybond-C (Amersham) with a BioRad trans-blot-cell apparatus. Transfer buffer was prepared as follows: Tris base 3g/l, glycine 14.4g/l, SDS 1g/l, methanol 200ml/l with about 3l required for the apparatus. Proteins were transferred at a constant 60V for 4 hours or at a constant

15mA overnight. After transfer, the protein could be visualised on the membrane by staining with Ponceau-S (recipe)

7.13.2 Immunodetection.

The membrane was blocked with 3%w/v skim milk powder in Tris-buffered saline (TBS: 10mM Tris-HCl pH 8.0, 100mM NaCl, 0.05%v/v Tween-20) for at least one hour at room temperature. Primary antibody, at an appropriate dilution, in TBS was added to the membrane with gentle agitation and incubation overnight at room temperature. Na-azide at a final concentration of 0.01%w/v was used to prevent microbial contamination.

Excess primary antibody was removed by successive washes with TBS at 30 minute intervals. At least 5 washes were used to ensure all Na-azide was removed as this chemical inhibits the HRP detection system. The secondary antibody was then added (typically HRP-conjugated anti-rabbit IgG) in TBS at a dilution recommended by the supplier, normally 1:500 to 1:1000. Incubation was for a minimum of one hour at room temperature.

Early Western blot experiments used chloro-napthol detection; four 30mg 1-chloro-4-napthol (Sigma) tablets were dissolved in 20ml of methanol then added to 100ml TBS followed by 100µl of 30%w/v H_2O_2. This solution was added to the membrane and allowed to incubate for approximately 30 minutes at room temperature. The membrane was rinsed several times in TBS then photographed, as the signal tended to fade rapidly. Later Western blots used the DAB detection system, a Sigma Chemical Co kit which supplied DAB and peroxidase as tablets sufficient to prepare 1ml of solution. Several sets of tablets were used for larger membranes with the solution diluted about 1:5 with PBS to ensure even coverage.

7.14 Purification of Inclusion Bodies

The method described by Harlow and Lane (1988) was used for the purification of inclusion bodies from *E coli*. Cells from a 50ml culture expressing the recombinant protein were collected by centrifugation at 5000rpm for 5 minutes (Sorvall RC-5B, SS-34 rotor) with the pellet resuspended in 5ml of buffer IB (100mM NaCl, 1mM EDTA, 50mM Tris-HCl pH8.0). Lysozyme was added to give a final concentration of 1mg/ml with the cell suspension incubated at room temperature for 20 minutes. The

cell suspension was centrifuged for 5 minutes at 5000*g* (6500 rpm SS-34) with the supernatant discarded (a sample was kept for SDS-PAGE analysis). The pellet was suspended in 5ml of buffer IB2 (buffer IB plus 0.1%w/v Na-deoxycholote), then incubated on ice for 10 minutes with occasional mixing. The solution was adjusted to 8mM $MgCl_2$ and 10µg/ml DNaseI, then incubated at 4°C until a noticeable reduction in the viscosity of the solution occurs. The inclusion bodies were collected by centrifugation at no more than 10 000*g* (~9200rpm SS34) for 10 minutes then washed by re-suspension in buffer IB3 (buffer IB plus 0.1%v/v NP-40) followed by collection by centrifugation. Further washes were carried out with buffer IB.

7.15 *in situ* Hybridisation to Polytene Chromosomes

7.15.1 Preparation of Chromosomes

Larvae at the 3rd instar stage were grown on cornmeal food supplemented with dried yeast. The larvae were collected from the sides of vials or bottles with a moist modeller's paintbrush then the salivary glands were dissected in Ringer's solution under the dissecting microscope. The salivary glands were transferred to a drop of 1:2:3 fixative (1 part lactic acid, 2 parts water, 3 parts acetic acid) on a clean microscope slide, then a clean siliconised coverslip was carefully placed on top. The salivary gland cells were disrupted by tapping the coverslip with the back end of the paintbrush in a circular motion and spread by gently streaking across the coverslip. An edge of the coverslip should be held during tapping and streaking to prevent it sliding and damaging the chromosomes. The slide was then sandwiched between layers of blotting paper and the chromosomes squashed by applying considerable thumb pressure. The chromosomes were examined for quality at 250x or 400x magnification using phase contrast.

The slides were marked to show which side the chromosomes were on. Coverslips were removed by dipping the slide into liquid N_2 then quickly flicking the coverslip from the slide with a scalpel blade. The chromosomes were dehydrated in 70%v/v ethanol for 5 minutes at room temperature followed by 95% ethanol for 5 minutes then air dried.

7.15.2 Denaturation

Prior to denaturation the chromosomes were treated in 2xSSC at 65°C for 30 minutes followed by treatment in 2xSSC at room temperature for 10 minutes. Chromosomes were denatured by treating in 70mM freshly prepared NaOH for 2-3 minutes followed by rinsing in 2xSSC. Dehydration through an ethanol series as described above was then carried out.

7.15.3 Hybridisation

A digoxigenin labelled DNA probe, prepared as described in section 7.5.5.3, was denatured by boiling in a water bath then plunging into ice. A 20µl volume of probe was added to each slide and a coverslip carefully placed over each slide. The slides were then placed into a box on a bed of pasteur pipettes which in turn rested on a wad of damp paper tissue. The box was sealed and hybridisation was allowed to take place overnight at either 42°C or 55°C. The coverslips were removed by placing the slides in 2xSSC and allowing them to float off. A 1 hour wash in 2xSSC at the equivalent hybridisation temperature was then performed.

7.15.4 Washing and Detection

After 2xSSC treatment, the slides were given 2 washes in PBS for 5 minutes each, followed by 1 wash in PBS-TX (PBS + 0.1%v/v Triton X-100) for 5 minutes followed by a further wash in PBS.

A 1:80 dilution of anti-digoxigenin-HRP conjugated antibody was prepared in PBS and 50µl applied to each slide with a coverslip added. The slides were incubated in the moist box at room temperature for 1-2 hours.

A series of PBS washes as above was carried out to remove excess antibody. Signal was detected using a DAB system (Sigma Chemical Co) - a DAB pellet and a peroxide/urea pellet were dissolved in 1ml of water with about 100µl applied to each slide and allowed to incubate at room temperature for 10-20 minutes in the moist box. Excess DAB solution was washed away by rinsing in PBS, then the slides were examined under phase contrast for appearance of signal (dark brown - black bands).

A giemsa (BDH) counterstain was used to allow visualisation of the chromosome bands.

7.16 Immunohistochemical Detection of Protein *in situ*

The method presented here is an amalgam of methods described by Siegel et al (1993) and Hedgé and Stephensen (1993).

Ovaries were dissected under the dissecting microscope in Ringer's solution and transferred immediately to a microfuge tube containing paraformaldehyde fixative. The fixative was prepared by dissolving 4%w/v paraformaldehyde in PBS at 65°C then adding 5% /v DMSO. The ovaries were fixed for 1 hour on a rotating wheel at room temperature. The fixative was carefully removed and 1ml of PBT (PBS + 0.1%v/v Triton X-100) was added. Incubation was continued for a further 5 minutes then the ovaries were extracted with PBS + 1%v/v Triton X-100 for a minimum of 2 hours (or overnight) on a rotating wheel. The ovaries were then washed with 3 changes of PBT for 5 minutes per wash. Blocking was accomplished by incubation with 10%v/v goat serum in PBT for 1 hour on the rotating wheel.

Antibody in PBT at the appropriate dilution was added and incubated overnight at room temperature on the rotating wheel. The antibody solution was removed and residual antibody was washed away with 3 changes of PBT with 30 minutes incubation per change. The secondary antibody was pre-absorbed with fixed ovaries treated in the same manner as above apart from incubation with primary antibodies. The secondary antibody was then applied to the ovaries and incubated from 2 hours to overnight. Experiments using fluorescent secondary antibody were wrapped in foil to prevent quenching by natural light. Excess secondary antibody was removed with 3 PBT washes at 30 minutes each as before.

Detection was then carried out. Ovaries treated with fluorescent secondary antibody were mounted on slides in a solution of propyl gallate in glycerol and examined under the fluorescent microscope. Those ovaries treated with a peroxidase secondary antibody were placed in a cavity slide to allow monitoring of the reaction and detected with a DAB solution (Sigma Chemical Co). The DAB reaction was stopped by rinsing with several changes of PBT.

7.17 Detection of Transcript by *in situ* Hybridisation to Whole Mount Ovaries

A method modified from that of Tautz And Pfiefle (1989) was used. All incubations were carried out on a rotating wheel unless otherwise stated. Ovaries were dissected from adult female flies which had been tipped to fresh food supplemented with dried yeast the day before. Dissected tissue was placed in a microfuge tube and maintained at a low temperature in dry ice. When sufficient tissue was collected it was fixed in 1ml of a freshly prepared 4%w/v paraformaldehyde in PBS for one hour at room temperature on a rotating wheel. The fixative was removed and the ovaries washed in 3 changes of PBT (PBS +Tween-20) with 5 minute incubations between changes. Proteinase K was added from a 20mg/ml stock to give a final concentration of 50mg/ml with incubation for 30 minutes at 37°C. The ovaries were then washed in 2 changes of PBT followed by a 20 minute treatment in fixative followed by 3 further washes in PBT with 10 minute intervals.

To prepare the ovaries for hybridisation, they were washed in a 1:1 solution of PBT:HS for 20 minutes (HS: 50%v/v formamide, 5xSSC, 50mg/ml heparin, 0.1%v/v Tween-20, 100µg/ml sonicated salmon sperm DNA) followed by incubation in 1xHS for 60 minutes. Prehybridisation was carried out in 1xHS for a minumum of 40 minutes at 45°C in a heating block. Hybridisation was carried out at 45°C overnight by replacing the HS with probe labelled with digoxigenin as described in section 7.5.5.3.

The probe was removed by 20 minute washes, first in 1xHS, then by HS:PBT washes at ratios of 4:1, 3:2, 2:3, 1:4 and finally in 2 washes of PBT. Signal was detected by incubation at room temperature with anti-digoxigenin antibody (Boehringer Mannheim) at a dilution of 1:5000 for I hour. The antibody solution was preabsorbed with ovaries treated in an identical manner as above but without being exposed to a probe. Excess antibody was removed with 4 washes of PBT at 20 minute intervals followed by three 5 minute washes in NMTLT buffer (100mM NaCl, 50mM $MgCl_2$, 100mM Tris-HCl pH 9.5, 1mM levamisole [freshly prepared], 0.1%v/v Tween-20). The ovaries were transferred to a cavity slide and 4.5ml NBT (0.5g nitro-blue-tetrazolium salt in 10ml dimethyformamide) and 3.5ml X-phosphate were added with thorough mixing. The signal was developed in the dark for 10 to 30 minutes and the reaction stopped by rinsing with PBT. Signal could then be observed by microscopic examination.

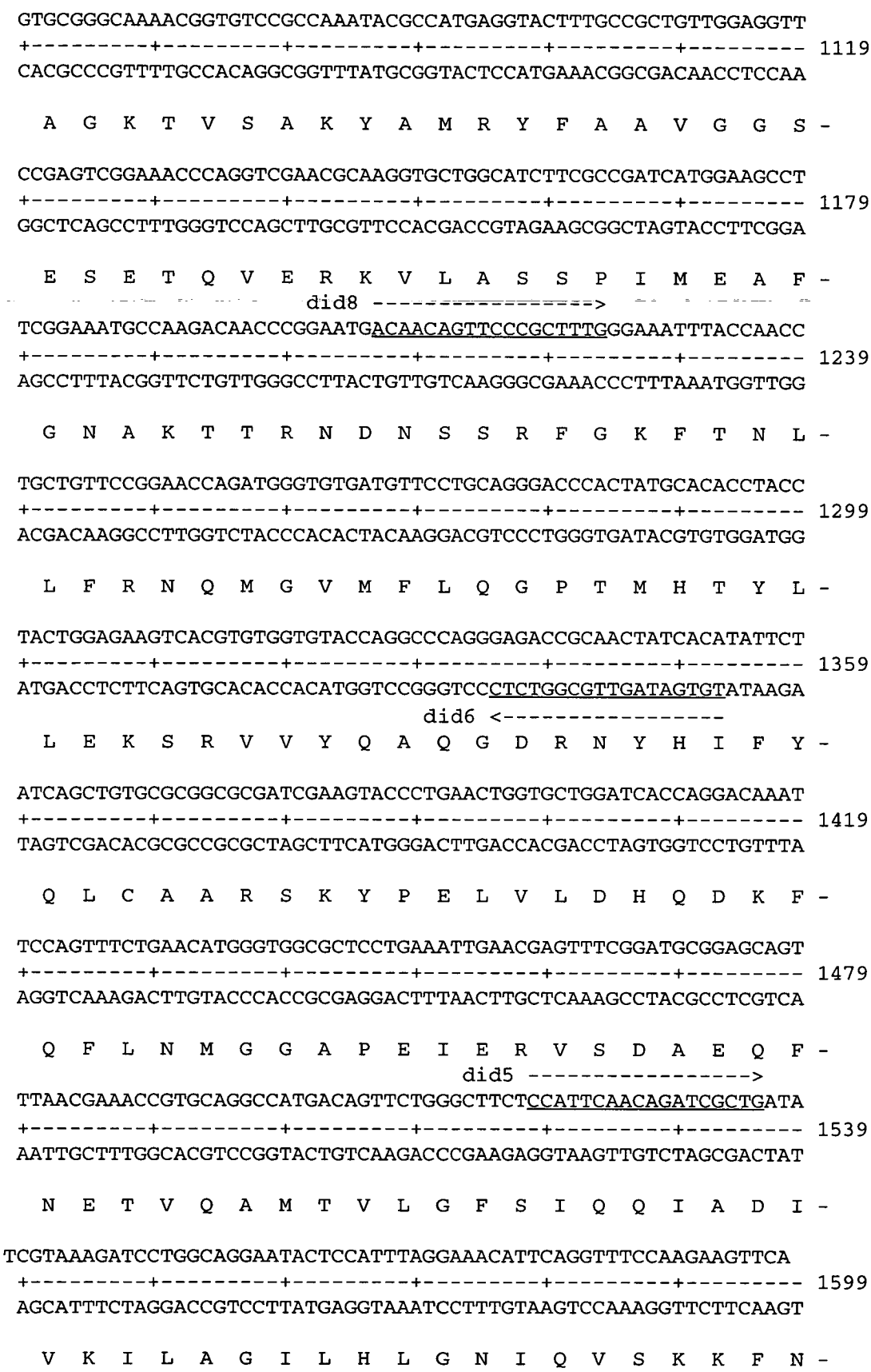
Appendices

Appendix I cDNA Sequence of the *didum* Gene

The first 666 nucleotides are uncorrected sequence from 5'RACE and RT-PCR experiments to determine the 5' end of the gene. The sequence determined from cDNA, genomic DNA and RT-PCR sequencing is shown from nucleotide 700 with the predicted protein sequence shown below. Primers used in this work have also been shown. Sequence from the cloned cDNA (pBMF10 and NMC7) begins at the *Eco*RI site marked at position 3489.

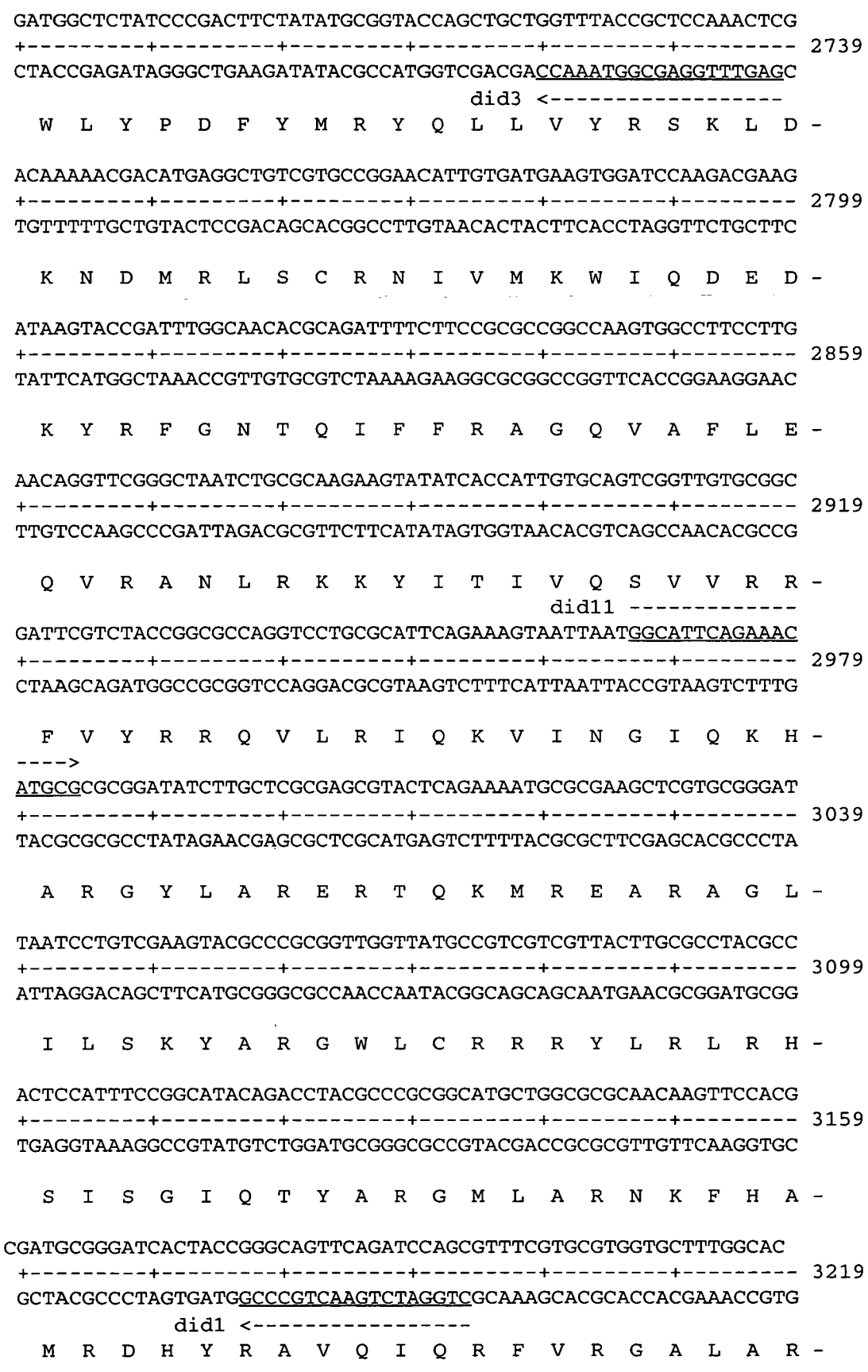
```
CyTTtyycmmtTTTwAAAargGGGGGGGGGGGgrGkGGaAmATGGAyCATGGTGTcTaAc
---------+---------+---------+---------+---------+---------+ 60
GrAAarrgkkaAAAwTTTtycCCCCCCCCCCcyCmCCtTkTACCTrGTACCACAgAtTg

           did7 ------------------->
AtATGTGTAAGTTAtTGGGATATAAACCTAATAGCGTAATTaACTTGAttaATaatggga
---------+---------+---------+---------+---------+---------+ 120
TaTACACATTCAATaACCCTATATTTGGATTATCGCATTAAtTGAACTaatTAttaccct

ttAGTtTTTtAgCTAtTTaTagtCTaATTAACACAATCCCGGGGcgtTCtaTATAGTTay
---------+---------+---------+---------+---------+---------+ 180
aaTCAaAAAaTcGATaAAtAtcaGAtTAATTGTGTTAGGGCCCCgcaAGatATATCAAtr

GTwTAaTGtATATTTAtaTtAttTaTGCCTCTaAcTGGaACGTACCyTGAGCATATATGc
---------+---------+---------+---------+---------+---------+ 240
CAwATtACaTATAAATatAaTaaAtACGGAGAtTgACCtTGCATGGrACTCGTATATACg

TGTGACCcGAAAGwtGGTGAACttATACTTGATCAGGTTGAAGTgCAGGGGamACCcTGA
---------+---------+---------+---------+---------+---------+ 300
ACACTGGgCTTTCwaCCACTTGaaTATGAACTAGTCCAACTTCAcGTCCCCtkTGGgACT

TgGAAgAccgaaAcAGTTtCtGAcGTtCAaATgCgATtgTCagAATTGAgTaTaGgGGcg
---------+---------+---------+---------+---------+---------+ 360
AcCTTcTggcttTgTCAAaGaCTgCAaGTtTAcGcTAacAGtcTTAACTcAtAtCcCCgc

AaAgAcCAATcgaACCATCTagtAgcTggTTCcTTCCgAAgTttCCCtCAgGAtAgCTGG
---------+---------+---------+---------+---------+---------+ 420
TtTcTgGTTAgctTGGTAGAtcaTcgAccAAGgAAGGcTTcAaaGGGaGTcCTaTcGACC

TGCATTTTAATATTATATAAAATAAtCTTATCTGGTaAAGCsAATGATTAgAgGcCtTAg
---------+---------+---------+---------+---------+---------+ 480
ACGTAAAATTATAATATATTTTATTaGAATAGACCAtTTCGsTTACTAATcTcCgGaATc

GGTCCAAACgAtCTTAACCTATTCTCaAACTTTAAATGGGTwAgAACCTTAACTTTCtTG
---------+---------+---------+---------+---------+---------+ 540
CCAGGTTTGcTaGAATTGGATAAGAGtTTGAAATTTACCCAwTcTTGGAATTGAAAGaAC

AtaTGAAaTTCmAGGkTaTtAaaTaATkTtCCCAytGGGCCaCtTTTGGTAACCaAAayT
---------+---------+---------+---------+---------+---------+ 600
TatACTTtAAGkTCCmAtAaTttAtTAmAaGGGTraCCCGGtGaAAACCATTGGtTTtrA
```

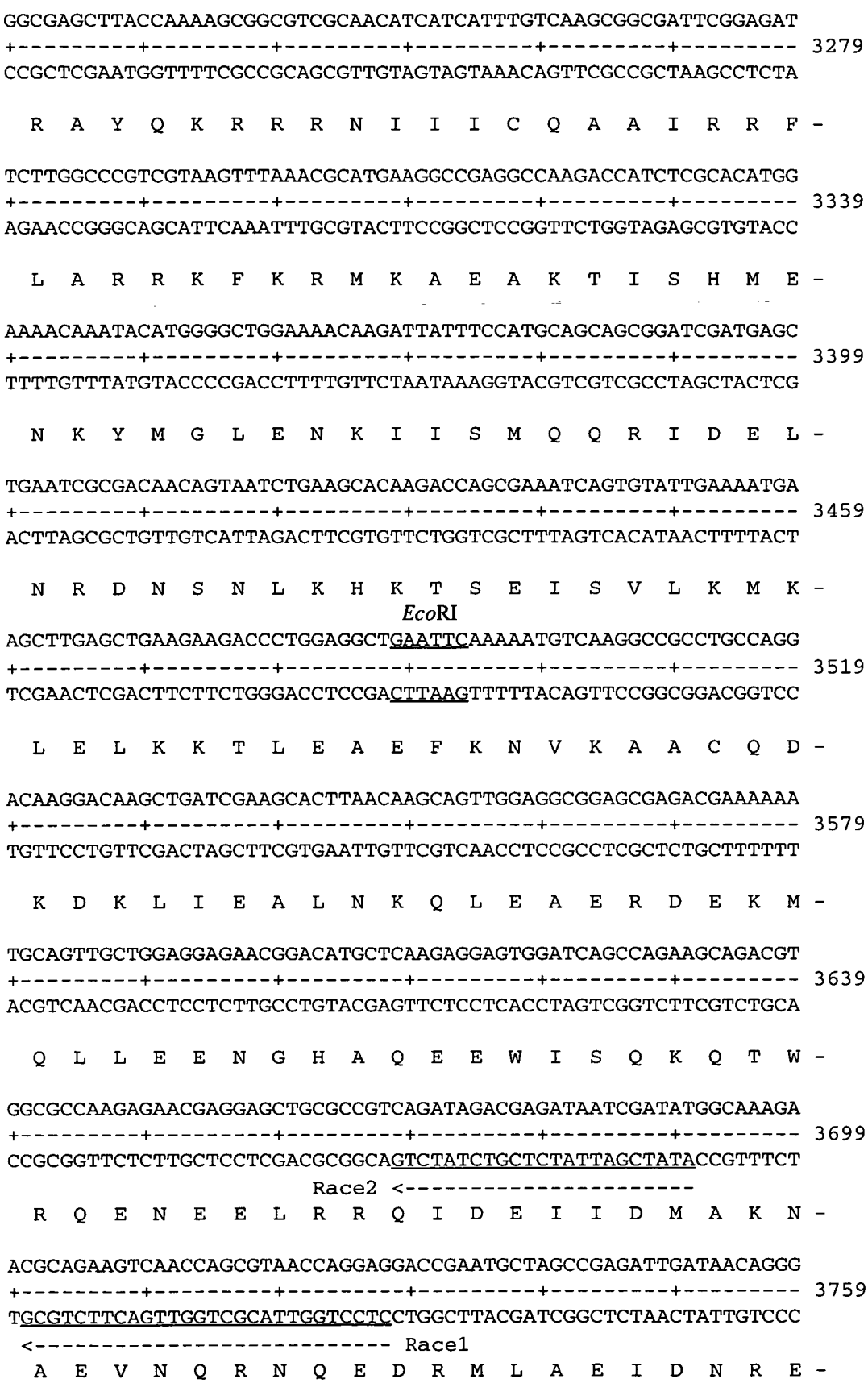
```
GgccCctTGGrATrAACCAAaCyTawTtTTmsggtsCmmytaayyAmCAmCyCCTGCmmA
---------+---------+---------+---------+---------+---------+ 660
CcggGgaACCyTAyTTGGTTtGrAtwAaAAksccasGkkrattrrTkGTkGrGGACGkkT


TnCCAT...
--------+
AnGGTA...
       <--------------------  Race4



700
-----------------> did2
 AAGTGAAAGAGGTCGAGCTAAAGGCCGATGGCAGCGATCTGCCTCCACTGCGCAATCCGG
 +---------+---------+---------+---------+---------+--------- 759
 TTCACTTTCTCCAGCTCGATTTCCGGCTACCGTCGCTAGACGGAGGTGACGCGTTAGGCC
     <--------------------  Race3
   V  K  E  V  E  L  K  A  D  G  S  D  L  P  P  L  R  N  P  A -

 CCATTCTGGTGGGACAGAACGACTTGACCACCCTGTCCTACCTGCATGAGCCGGGGGTGT
 +---------+---------+---------+---------+---------+--------- 819
 GGTAAGACCACCCTGTCTTGCTGAACTGGTGGGACAGGATGGACGTACTCGGCCCCCACA

   I  L  V  G  Q  N  D  L  T  T  L  S  Y  L  H  E  P  G  V  L -

 TGCACAATCTGCGTGTCCGCTTCTGCAAGCGCCAAATTATCTACACCTACTGCGGCATCA
 +---------+---------+---------+---------+---------+--------- 879
 ACGTGTTAGACGCACAGGCGAAGACGTTCGCGGTTTAATAGATGTGGATGACGCCGTAGT

   H  N  L  R  V  R  F  C  K  R  Q  I  I  Y  T  Y  C  G  I  I -

 TTCTGGTGGCCATCAACCCGTACGCGGAGATGCCTCTTTACGGGCCCAGCATAATCCGAG
 +---------+---------+---------+---------+---------+--------- 939
 AAGACCACCGGTAGTTGGGCATGCGCCTCTACGGAGAAATGCCCGGGTCGTATTAGGCTC
                                                  <---------
   L  V  A  I  N  P  Y  A  E  M  P  L  Y  G  P  S  I  I  R  A -

 CGTATCGGGGTCATGCTATGGGTGATCTGGAGCCGCACATCTTTGCCCTGGCGGAGGAGG
 +---------+---------+---------+---------+---------+--------- 999
 GCATAGCCCCAGTACGATACCCACTAGACCTCGGCGTGTAGAAACGGGACCGCCTCCTCC
 ------- did12
   Y  R  G  H  A  M  G  D  L  E  P  H  I  F  A  L  A  E  E  A -

 CGTACACAAAACTGGAGCGCGAGAACTGCAACCTGAGCATCATCGTCAGTGGGGAATCGG
 +---------+---------+---------+---------+---------+--------- 1059
 GCATGTGTTTTGACCTCGCGCTCTTGACGTTGGACTCGTAGTAGCAGTCACCCCTTAGCC

   Y  T  K  L  E  R  E  N  C  N  L  S  I  I  V  S  G  E  S  G -
```


GTGCGGGCAAAACGGTGTCCGCCAAATACGCCATGAGGTACTTTGCCGCTGTTGGAGGTT
CACGCCCGTTTTGCCACAGGCGGTTTATGCGGTACTCCATGAAACGGCGACAACCTCCAA
1119
A G K T V S A K Y A M R Y F A A V G G S -
CCGAGTCGGAAACCCAGGTCGAACGCAAGGTGCTGGCATCTTCGCCGATCATGGAAGCCT
GGCTCAGCCTTTGGGTCCAGCTTGCGTTCCACGACCGTAGAAGCGGCTAGTACCTTCGGA
1179
E S E T Q V E R K V L A S S P I M E A F -
did8 ----------->
TCGGAAATGCCAAGACAACCCGGAATGACAACAGTTCCCGCTTTGGGAAATTTACCAACC
AGCCTTTACGGTTCTGTTGGGCCTTACTGTTGTCAAGGGCGAAACCCTTTAAATGGTTGG
1239
G N A K T T R N D N S S R F G K F T N L -
TGCTGTTCCGGAACCAGATGGGTGTGATGTTCCTGCAGGGACCCACTATGCACACCTACC
ACGACAAGGCCTTGGTCTACCCACACTACAAGGACGTCCCTGGGTGATACGTGTGGATGG
1299
L F R N Q M G V M F L Q G P T M H T Y L -
TACTGGAGAAGTCACGTGTGGTGTACCAGGCCCAGGGAGACCGCAACTATCACATATTCT
ATGACCTCTTCAGTGCACACCACATGGTCCGGGTCCCTCTGGCGTTGATAGTGTATAAGA
1359
<----------
did6
L E K S R V V Y Q A Q G D R N Y H I F Y -
ATCAGCTGTGCGCGGCGCGATCGAAGTACCCTGAACTGGTGCTGGATCACCAGGACAAAT
TAGTCGACACGCGCCGCGCTAGCTTCATGGGACTTGACCACGACCTAGTGGTCCTGTTTA
1419
Q L C A A R S K Y P E L V L D H Q D K F -
TCCAGTTTCTGAACATGGGTGGCGCTCCTGAAATTGAACGAGTTTCGGATGCGGAGCAGT
AGGTCAAAGACTTGTACCCACCGCGAGGACTTTAACTTGCTCAAAGCCTACGCCTCGTCA
1479
Q F L N M G G A P E I E R V S D A E Q F -
did5 ----------->
TTAACGAAACCGTGCAGGCCATGACAGTTCTGGGCTTCTCCATTCAACAGATCGCTGATA
AATTGCTTTGGCACGTCCGGTACTGTCAAGACCCGAAGAGGTAAGTTGTCTAGCGACTAT
1539
N E T V Q A M T V L G F S I Q Q I A D I -
TCGTAAAGATCCTGGCAGGAATACTCCATTTAGGAAACATTCAGGTTTCCAAGAAGTTCA
AGCATTTCTAGGACCGTCCTTATGAGGTAAATCCTTTGTAAGTCCAAAGGTTCTTCAAGT
1599
V K I L A G I L H L G N I Q V S K K F N -

```
ACGAGGGCAGCGAAGAGGAGGACAGTGACTCTTGCGATATATTTCATAACGACATCCACC
+---------+---------+---------+---------+---------+--------- 1659
TGCTCCCGTCGCTTCTCCTCCTGTCACTGAGAACGCTATATAAAGTATTGCTGTAGGTGG

  E  G  S  E  E  E  D  S  D  S  C  D  I  F  H  N  D  I  H  L -

TGCAGATCACCGCCGATCTACTGCGGGTGAGCGCCGATGATCTGCGCCGGTGGCTTTTGA
+---------+---------+---------+---------+---------+--------- 1719
ACGTCTAGTGGCGGCTAGATGACGCCCACTCGCGGCTACTAGACGCGGCCACCGAAAACT

  Q  I  T  A  D  L  L  R  V  S  A  D  D  L  R  R  W  L  L  M -

TGCGTAAGATAGAGTCGGTCAATGAATATGTGCTGATACCGAATAGCATTGAGGCGGCTC
+---------+---------+---------+---------+---------+--------- 1779
ACGCATTCTATCTCAGCCAGTTACTTATACACGACTATGGCTTATCGTAACTCCGCCGAG

  R  K  I  E  S  V  N  E  Y  V  L  I  P  N  S  I  E  A  A  Q -

AGGCGGCTCGAGACGCTCTGGCCAAGCACATCTATGCGAAACTGTTTCAGTATATAGTCG
+---------+---------+---------+---------+---------+--------- 1839
TCCGCCGAGCTCTGCGAGACCGGTTCGTGTAGATACGCTTTGACAAAGTCATATATCAGC

  A  A  R  D  A  L  A  K  H  I  Y  A  K  L  F  Q  Y  I  V  G -

GTGTGCTGAACAAGAGCCTCAACAACGGTAGCAAGCAGTGCAGCTTCATTGGCGTCCTCG
+---------+---------+---------+---------+---------+--------- 1899
CACACGACTTGTTCTCGGAGTTGTTGCCATCGTTCGTCACGTCGAAGTAACCGCAGGAGC

  V  L  N  K  S  L  N  N  G  S  K  Q  C  S  F  I  G  V  L  D -

ATATCTACGGCTTCGAAACGTTCGAGGTGAACTCCTTTGAACAATTTTGCATAAACTATG
+---------+---------+---------+---------+---------+--------- 1959
TATAGATGCCGAAGCTTTGCAAGCTCCACTTGAGGAAACTTGTTAAAACGTATTTGATAC

  I  Y  G  F  E  T  F  E  V  N  S  F  E  Q  F  C  I  N  Y  A -
                      did9 ----------------->
CAAACGAAAAGCTTCAGCAGCAGTTCAACCAGCATGTCTTCAAGCTGGAGCAGGAGGAGT
+---------+---------+---------+---------+---------+--------- 2019
GTTTGCTTTTCGAAGTCGTCGTCAAGTTGGTCGTACAGAAGTTCGACCTCGTCCTCCTCA

  N  E  K  L  Q  Q  Q  F  N  Q  H  V  F  K  L  E  Q  E  E  Y -

ACCTTAAGGAAGGAATCACCTGGACGATGATTGACTTTTACGACAATCAACCGTGTATCG
+---------+---------+---------+---------+---------+--------- 2079
TGGAATTCCTTCCTTAGTGGACCTGCTACTAACTGAAAATGCTGTTAGTTGGCACATAGC

  L  K  E  G  I  T  W  T  M  I  D  F  Y  D  N  Q  P  C  I  D -
ATCTAATTGAATCTCGCTTGGGAGTGCTGGACCTGCTCGACGAGGAGTGTCGAATGCCAA
+---------+---------+---------+---------+---------+--------- 2139
TAGATTAACTTAGAGCGAACCCTCACGACCTGGACGAGCTGCTCCTCACAGCTTACGGTT

  L  I  E  S  R  L  G  V  L  D  L  L  D  E  E  C  R  M  P  K -
```

```
AGGGCTCGGACGAGAGCTGGGCTGGCAAGCTCATCGGAAAGTGCAATAAATTTCCGCATT
+---------+---------+---------+---------+---------+--------- 2199
TCCCGAGCCTGCTCTCGACCCGACCGTTCGAGTAGCCTTTCACGTTATTTAAAGGCGTAA

  G   S   D   E   S   W   A   G   K   L   I   G   K   C   N   K   F   P   H   F -

TCGAGAAGCCACGCTTTGGCACAACCAGCTTCTTTATCAAACATTTCTCGGACACGGTCG
+---------+---------+---------+---------+---------+--------- 2259
AGCTCTTCGGTGCGAAACCGTGTTGGTCGAAGAAATAGTTTGTAAAGAGCCTGTGCCAGC

  E   K   P   R   F   G   T   T   S   F   F   I   K   H   F   S   D   T   V   E -

AGTATGACGTGAACGGATTCTTGGAAAAGAATCGTGACACAGTCTCCAAGGAGTTGACGC
+---------+---------+---------+---------+---------+--------- 2319
TCATACTGCACTTGCCTAAGAACCTTTTCTTAGCACTGTGTCAGAGGTTCCTCAACTGCG
 did10 <-----------------
  Y   D   V   N   G   F   L   E   K   N   R   D   T   V   S   K   E   L   T   Q -

AAGTGCTAAGCGAGTCCAACATGTCTCTGGCCAAGCAGGTGATGACCCTGGAGGAAATAG
+---------+---------+---------+---------+---------+--------- 2379
TTCACGATTCGCTCAGGTTGTACAGAGACCGGTTCGTCCACTACTGGGACCTCCTTTATC

  V   L   S   E   S   N   M   S   L   A   K   Q   V   M   T   L   E   E   I   D -
          did4 ----------------->
ATACTCTGTGCGTGGATTCCGCTAAATCCTCCACCTTGGGCGGCCGCGTCGTGATCAGTG
+---------+---------+---------+---------+---------+--------- 2439
TATGAGACACGCACCTAAGGCGATTTAGGAGGTGGAACCCGCCGGCGCAGCACTAGTCAC

  T   L   C   V   D   S   A   K   S   S   T   L   G   G   R   V   V   I   S   A -

CTGGCCGCAAACAGGTGGTGCCATCCAAGCAGCATAGAAAAACGGTGGGATCGCAGTTCC
+---------+---------+---------+---------+---------+--------- 2499
GACCGGCGTTTGTCCACCACGGTAGGTTCGTCGTATCTTTTTGCCACCCTAGCGTCAAGG

  G   R   K   Q   V   V   P   S   K   Q   H   R   K   T   V   G   S   Q   F   Q -

AGGAGAGTCTGGCGTCGCTGATATCTACGTTACATGCCACAACTCCGCACTATGTGCGCT
+---------+---------+---------+---------+---------+--------- 2559
TCCTCTCAGACCGCAGCGACTATAGATGCAATGTACGGTGTTGAGGCGTGATACACGCGA

  E   S   L   A   S   L   I   S   T   L   H   A   T   T   P   H   Y   V   R   C -

GCATCAAGCCCAACGATGACAAAGTCGCCTTTAAGTGGGAGACGGCCAAGATCATACAGC
+---------+---------+---------+---------+---------+--------- 2619
CGTAGTTCGGGTTGCTACTGTTTCAGCGGAAATTCACCCTCTGCCGGTTCTAGTATGTCG

  I   K   P   N   D   D   K   V   A   F   K   W   E   T   A   K   I   I   Q   Q -

AGTTAAGGGCCTGTGGTGTGCTGGAAACGGTGCGCATCTCCGCAGCGGGATTCCCCTCGA
+---------+---------+---------+---------+---------+--------- 2679
TCAATTCCCGGACACCACACGACCTTTGCCACGCGTAGAGGCGTCGCCCTAAGGGGAGCT

  L   R   A   C   G   V   L   E   T   V   R   I   S   A   A   G   F   P   S   R -
```


GATGGCTCTATCCCGACTTCTATATGCGGTACCAGCTGCTGGTTTACCGCTCCAAACTCG
+---------+---------+---------+---------+---------+--------- 2739
CTACCGAGATAGGGCTGAAGATATACGCCATGGTCGACGACCAAATGGCGAGGTTTGAGC
did3 <----------
W L Y P D F Y M R Y Q L L V Y R S K L D -
ACAAAAACGACATGAGGCTGTCGTGCCGGAACATTGTGATGAAGTGGATCCAAGACGAAG
+---------+---------+---------+---------+---------+--------- 2799
TGTTTTTGCTGTACTCCGACAGCACGGCCTTGTAACACTACTTCACCTAGGTTCTGCTTC
K N D M R L S C R N I V M K W I Q D E D -
ATAAGTACCGATTTGGCAACACGCAGATTTTCTTCCGCGCCGGCCAAGTGGCCTTCCTTG
+---------+---------+---------+---------+---------+--------- 2859
TATTCATGGCTAAACCGTTGTGCGTCTAAAAGAAGGCGCGGCCGGTTCACCGGAAGGAAC
K Y R F G N T Q I F F R A G Q V A F L E -
AACAGGTTCGGGCTAATCTGCGCAAGAAGTATATCACCATTGTGCAGTCGGTTGTGCGGC
+---------+---------+---------+---------+---------+--------- 2919
TTGTCCAAGCCCGATTAGACGCGTTCTTCATATAGTGGTAACACGTCAGCCAACACGCCG
Q V R A N L R K K Y I T I V Q S V V R R -
did11 ----------
GATTCGTCTACCGGCGCCAGGTCCTGCGCATTCAGAAAGTAATTAATGGCATTCAGAAAC
+---------+---------+---------+---------+---------+--------- 2979
CTAAGCAGATGGCCGCGGTCCAGGACGCGTAAGTCTTTCATTAATTACCGTAAGTCTTTG
F V Y R R Q V L R I Q K V I N G I Q K H -
---->
ATGCGCGCGGATATCTTGCTCGCGAGCGTACTCAGAAAATGCGCGAAGCTCGTGCGGGAT
+---------+---------+---------+---------+---------+--------- 3039
TACGCGCGCCTATAGAACGAGCGCTCGCATGAGTCTTTTACGCGCTTCGAGCACGCCCTA
A R G Y L A R E R T Q K M R E A R A G L -
TAATCCTGTCGAAGTACGCCCGCGGTTGGTTATGCCGTCGTCGTTACTTGCGCCTACGCC
+---------+---------+---------+---------+---------+--------- 3099
ATTAGGACAGCTTCATGCGGGCGCCAACCAATACGGCAGCAGCAATGAACGCGGATGCGG
I L S K Y A R G W L C R R R Y L R L R H -
ACTCCATTTCCGGCATACAGACCTACGCCCGCGGCATGCTGGCGCGCAACAAGTTCCACG
+---------+---------+---------+---------+---------+--------- 3159
TGAGGTAAAGGCCGTATGTCTGGATGCGGGCGCCGTACGACCGCGCGTTGTTCAAGGTGC
S I S G I Q T Y A R G M L A R N K F H A -
CGATGCGGGATCACTACCGGGCAGTTCAGATCCAGCGTTTCGTGCGTGGTGCTTTGGCAC
+---------+---------+---------+---------+---------+--------- 3219
GCTACGCCCTAGTGATGGCCCGTCAAGTCTAGGTCGCAAAGCACGCACCACGAAACCGTG
did1 <----------
M R D H Y R A V Q I Q R F V R G A L A R -


GGCGAGCTTACCAAAAGCGGCGTCGCAACATCATCATTTGTCAAGCGGCGATTCGGAGAT
3279
CCGCTCGAATGGTTTTCGCCGCAGCGTTGTAGTAGTAAACAGTTCGCCGCTAAGCCTCTA
R A Y Q K R R R N I I I C Q A A I R R F -
TCTTGGCCCGTCGTAAGTTTAAACGCATGAAGGCCGAGGCCAAGACCATCTCGCACATGG
3339
AGAACCGGGCAGCATTCAAATTTGCGTACTTCCGGCTCCGGTTCTGGTAGAGCGTGTACC
L A R R K F K R M K A E A K T I S H M E -
AAAACAAATACATGGGGCTGGAAAACAAGATTATTTCCATGCAGCAGCGGATCGATGAGC
3399
TTTTGTTTATGTACCCCGACCTTTTGTTCTAATAAAGGTACGTCGTCGCCTAGCTACTCG
N K Y M G L E N K I I S M Q Q R I D E L -
TGAATCGCGACAACAGTAATCTGAAGCACAAGACCAGCGAAATCAGTGTATTGAAAATGA
3459
ACTTAGCGCTGTTGTCATTAGACTTCGTGTTCTGGTCGCTTTAGTCACATAACTTTTACT
N R D N S N L K H K T S E I S V L K M K -
EcoRI
AGCTTGAGCTGAAGAAGACCCTGGAGGCTGAATTCAAAAATGTCAAGGCCGCCTGCCAGG
3519
TCGAACTCGACTTCTTCTGGGACCTCCGACTTAAGTTTTTACAGTTCCGGCGGACGGTCC
L E L K K T L E A E F K N V K A A C Q D -
ACAAGGACAAGCTGATCGAAGCACTTAACAAGCAGTTGGAGGCGGAGCGAGACGAAAAAA
3579
TGTTCCTGTTCGACTAGCTTCGTGAATTGTTCGTCAACCTCCGCCTCGCTCTGCTTTTTT
K D K L I E A L N K Q L E A E R D E K M -
TGCAGTTGCTGGAGGAGAACGGACATGCTCAAGAGGAGTGGATCAGCCAGAAGCAGACGT
3639
ACGTCAACGACCTCCTCTTGCCTGTACGAGTTCTCCTCACCTAGTCGGTCTTCGTCTGCA
Q L L E E N G H A Q E E W I S Q K Q T W -
GGCGCCAAGAGAACGAGGAGCTGCGCCGTCAGATAGACGAGATAATCGATATGGCAAAGA
3699
CCGCGGTTCTCTTGCTCCTCGACGCGGCAGTCTATCTGCTCTATTAGCTATACCGTTTCT
Race2 <----------
R Q E N E E L R R Q I D E I I D M A K N -
ACGCAGAAGTCAACCAGCGTAACCAGGAGGACCGAATGCTAGCCGAGATTGATAACAGGG
3759
TGCGTCTTCAGTTGGTCGCATTGGTCCTCCTGGCTTACGATCGGCTCTAACTATTGTCCC
<---------- Race1
A E V N Q R N Q E D R M L A E I D N R E -

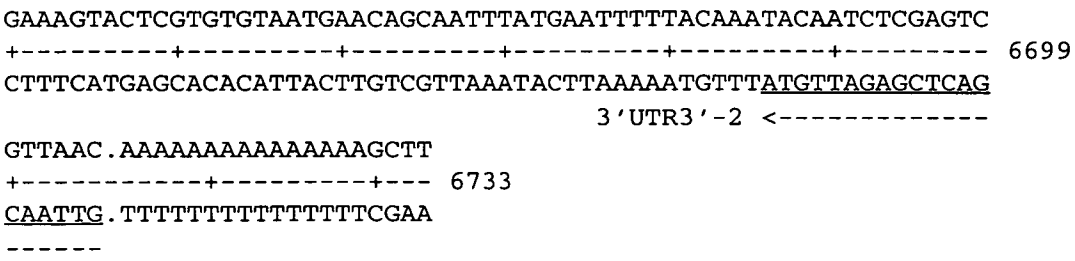
```
AGCTCAACGAGGCCTACCAACGAGCTATTAAGGACAAGGAGGTCATCGAGAACGAAAACT
+---------+---------+---------+---------+---------+--------- 3819
TCGAGTTGCTCCGGATGGTTGCTCGATAATTCCTGTTCCTCCAGTAGCTCTTGCTTTTGA

  L  N  E  A  Y  Q  R  A  I  K  D  K  E  V  I  E  N  E  N  F  -

TCATGCTGAAGGAAGAGCTCAGTCGATTAACGGCTGGCAGTTTCAGTTTGCACGGCCGCA
+---------+---------+---------+---------+---------+--------- 3879
AGTACGACTTCCTTCTCGAGTCAGCTAATTGCCGACCGTCAAAGTCAAACGTGCCGGCGT

  M  L  K  E  E  L  S  R  L  T  A  G  S  F  S  L  H  G  R  K  -

AGGCTAGCAACGCCTCCAGCCAAAACGAGGACGATGTGGGATACGCCTCCGCCAAGAACA
+---------+---------+---------+---------+---------+--------- 3939
TCCGATCGTTGCGGAGGTCGGTTTTGCTCCTGCTACACCCTATGCGGAGGCGGTTCTTGT

  A  S  N  A  S  S  Q  N  E  D  D  V  G  Y  A  S  A  K  N  T  -

CTCTGGATATCAATCGGCCCCCGGATTTGTTAAGCAAAAATTACTCGTACAATGACTCTA
+---------+---------+---------+---------+---------+--------- 3999
GAGACCTATAGTTAGCCGGGGGCCTAAACAATTCGTTTTTAATGAGCATGTTACTGAGAT

  L  D  I  N  R  P  P  D  L  L  S  K  N  Y  S  Y  N  D  S  T  -

CCAGTCTGGTGGTGAAGTTGAGATCCATTCTCGAGGAGGAGAAGCAAAAGCACAAGGTCT
+---------+---------+---------+---------+---------+--------- 4059
GGTCAGACCACCACTTCAACTCTAGGTAAGAGCTCCTCCTCTTCGTTTTCGTGTTCCAGA

  S  L  V  V  K  L  R  S  I  L  E  E  E  K  Q  K  H  K  V  L  -

TGCAGGAGCAGTACATTAAGTTGTCCAGTCGGCATAAGCCCACCGAGGATTCCTTCCGCG
+---------+---------+---------+---------+---------+--------- 4119
ACGTCCTCGTCATGTAATTCAACAGGTCAGCCGTATTCGGGTGGCTCCTAAGGAAGGCGC
                          601 <-----------------------
  Q  E  Q  Y  I  K  L  S  S  R  H  K  P  T  E  D  S  F  R  V  -

TCTCCGAGCTTGAGGTAGAGAATGAAAAGCTGCGCAGGAGGTACGATCAGCTGCGGACGA
+---------+---------+---------+---------+---------+--------- 4179
AGAGGCTCGAACTCCATCTCTTACTTTTCGACGCGTCCTCCATGCTAGTCGACGCCTGCT

  S  E  L  E  V  E  N  E  K  L  R  R  R  Y  D  Q  L  R  T  S  -

GCATTAAACACGGTGTTGAGATCAACGAGCTCAATGCACAGCATGCCGCCTTGCAGGAAG
+---------+---------+---------+---------+---------+--------- 4239
CGTAATTTGTGCCACAACTCTAGTTGCTCGAGTTACGTGTCGTACGGCGGAACGTCCTTC

  I  K  H  G  V  E  I  N  E  L  N  A  Q  H  A  A  L  Q  E  E  -

AGGTACGTAGGCGGCGCGAGGAGTGCATCCAATTAAAGGCAGTCCTTCTGCAGCAGAGCC
+---------+---------+---------+---------+---------+--------- 4299
TCCATGCATCCGCCGCGCTCCTCACGTAGGTTAATTTCCGTCAGGAAGACGTCGTCTCGG

  V  R  R  R  R  E  E  C  I  Q  L  K  A  V  L  L  Q  Q  S  Q  -
```

```
AGTCCATGAGATCGCTCGAGCCGGAAAGTCTACAGATGCGTGGCAACGATGTCAACGAAC
+---------+---------+---------+---------+---------+--------- 4359
TCAGGTACTCTAGCGAGCTCGGCCTTTCAGATGTCTACGCACCGTTGCTACAGTTGCTTG

  S  M  R  S  L  E  P  E  S  L  Q  M  R  G  N  D  V  N  E  L -

TGATGGAAGCCTTCCATTCCCAGAAGCTAATTAATCGTCAATTGGAGTCTGAGCTCAAGG
+---------+---------+---------+---------+---------+--------- 4419
ACTACCTTCGGAAGGTAAGGGTCTTCGATTAATTAGCAGTTAACCTCAGACTCGAGTTCC

  M  E  A  F  H  S  Q  K  L  I  N  R  Q  L  E  S  E  L  K  A -

CCATCACCGAGGAGCACAACAGTAAGCTCGTGGAGATGACACAGGAGATCGAGAGATTGA
+---------+---------+---------+---------+---------+--------- 4479
GGTAGTGGCTCCTCGTGTTGTCATTCGAGCACCTCTACTGTGTCCTCTAGCTCTCTAACT

  I  T  E  E  H  N  S  K  L  V  E  M  T  Q  E  I  E  R  L  N -

ACAATGAGAAGGATGAGCTGCAAAAAGTAATGTTCGAGAGCATCGACGAGTTCGAAGATT
+---------+---------+---------+---------+---------+--------- 4539
TGTTACTCTTCCTACTCGACGTTTTTCATTACAAGCTCTCGTAGCTGCTCAAGCTTCTAA

  N  E  K  D  E  L  Q  K  V  M  F  E  S  I  D  E  F  E  D  S -

CCAATGTGGATACGCTGAGACAGAACGATCGCTATCTGCGACGAGAACTGCAGAAGGCTG
+---------+---------+---------+---------+---------+--------- 4599
GGTTACACCTATGCGACTCTGTCTTGCTAGCGATAGACGCTGCTCTTGACGTCTTCCGAC

  N  V  D  T  L  R  Q  N  D  R  Y  L  R  R  E  L  Q  K  A  V -

TAGCCCAGTTCCTGCTCGTTCAGGAGGAGCTCAAACTGGCAAATGCCAAGCTTAAAGCTT
+---------+---------+---------+---------+---------+--------- 4659
ATCGGGTCAAGGACGAGCAAGTCCTCCTCGAGTTTGACCGTTTACGGTTCGAATTTCGAA

  A  Q  F  L  L  V  Q  E  E  L  K  L  A  N  A  K  L  K  A  Y -

ATCGGCAGGATGGAGGCCAGCTGGAGCACAAGATAGAGGAGGAGATGATTCGCAACAAGT
+---------+---------+---------+---------+---------+--------- 4719
TAGCCGTCCTACCTCCGGTCGACCTCGTGTTCTATCTCCTCCTCTACTAAGCGTTGTTCA

  R  Q  D  G  G  Q  L  E  H  K  I  E  E  E  M  I  R  N  K  S -

CCAACGGAACGTCCGCCGATGTAGGCGCGAATGTGACGAAGCAAAAGTCTCAGAATCCGC
+---------+---------+---------+---------+---------+--------- 4779
GGTTGCCTTGCAGGCGGCTACATCCGCGCTTACACTGCTTCGTTTTCAGAGTCTTAGGCG

  N  G  T  S  A  D  V  G  A  N  V  T  K  Q  K  S  Q  N  P  Q -

AAGGGCTGATGAAGTTCCACAGCAGCGATCTGGACAAGATCTTGCAACGCCTGCTTAGCG
+---------+---------+---------+---------+---------+--------- 4839
TTCCCGACTACTTCAAGGTGTCGTCGCTAGACCTGTTCTAGAACGTTGCGGACGAATCGC
                1311 <----------------------
  G  L  M  K  F  H  S  S  D  L  D  K  I  L  Q  R  L  L  S  A -
```

```
CCTTGACTCCACGCACAGTGGTCGGGCTCCTCCCTGGTTTTCCAGCATATCTCATCTTTA
+---------+---------+---------+---------+---------+--------- 4899
GGAACTGAGGTGCGTGTCACCAGCCCGAGGAGGGACCAAAAGGTCGTATAGAGTAGAAAT

  L  T  P  R  T  V  V  G  L  L  P  G  F  P  A  Y  L  I  F  M -

TGTGTATTCGATACACCGATCTGACAAATGCCGACGATGATGTGCGCGAGTTGCTAAGCA
+---------+---------+---------+---------+---------+--------- 4959
ACACATAAGCTATGTGGCTAGACTGTTTACGGCTGCTACTACACGCGCTCAACGATTCGT

  C  I  R  Y  T  D  L  T  N  A  D  D  D  V  R  E  L  L  S  K -

AGTTCGTTATTCAGATTAAGAAAATGCATCGTACGCCGCATCCGATCGAGAATCGTGTTA
+---------+---------+---------+---------+---------+--------- 5019
TCAAGCAATAAGTCTAATTCTTTTACGTAGCATGCGGCGTAGGCTAGCTCTTAGCACAAT

  F  V  I  Q  I  K  K  M  H  R  T  P  H  P  I  E  N  R  V  I -

TTTGGCTCGTCAATTCCATTACGCTGCTAAATCTTATGAAGCAATACGGCGACGTGGATG
+---------+---------+---------+---------+---------+--------- 5079
AAACCGAGCAGTTAAGGTAATGCGACGATTTAGAATACTTCGTTATGCCGCTGCACCTAC

  W  L  V  N  S  I  T  L  L  N  L  M  K  Q  Y  G  D  V  D  E -

AGTACGTCAAGTTCAATACTGAGAAGCAGAATCAGCAGCAGCTGAAGAACTTCAATCTCT
+---------+---------+---------+---------+---------+--------- 5139
TCATGCAGTTCAAGTTATGACTCTTCGTCTTAGTCGTCGTCGACTTCTTGAAGTTAGAGA
                    B100  <----------------
  Y  V  K  F  N  T  E  K  Q  N  Q  Q  Q  L  K  N  F  N  L  F -

TTGAATACCGTCGCGTAATTCTTGATTTATTTGTGAACCTGTACCAGGCGCTGATCATGC
+---------+---------+---------+---------+---------+--------- 5199
AACTTATGGCAGCGCATTAAGAACTAAATAAACACTTGGACATGGTCCGCGACTAGTACG

  E  Y  R  R  V  I  L  D  L  F  V  N  L  Y  Q  A  L  I  M  Q -

AGATCCAGGGTCTGTTGGACCCAAAAATAGTGCCAGCGATTCTCAACAATGATGAGATTC
+---------+---------+---------+---------+---------+--------- 5259
TCTAGGTCCCAGACAACCTGGGTTTTTATCACGGTCGCTAAGAGTTGTTACTACTCTAAG

  I  Q  G  L  L  D  P  K  I  V  P  A  I  L  N  N  D  E  I  Q -
                              1806 ---------------->
AGCGTGGGCGGCAGGCGCACGGAATGCGTAGTCGGGCCACGTCGATTGGAGCATCCTCGT
+---------+---------+---------+---------+---------+--------- 5319
TCGCACCCGCCGTCCGCGTGCCTTACGCATCAGCCCGGTGCAGCTAACCTCGTAGGAGCA

  R  G  R  Q  A  H  G  M  R  S  R  A  T  S  I  G  A  S  S  S -

CACCGGAGCACGGTGGCGGTCCGGCCTGGAAGCAACTGATCGGGCAGCTGGAGCATTTCT
+---------+---------+---------+---------+---------+--------- 5379
GTGGCCTCGTGCCACCGCCAGGCCGGACCTTCGTTGACTAGCCCGTCGACCTCGTAAAGA

  P  E  H  G  G  G  P  A  W  K  Q  L  I  G  Q  L  E  H  F  Y -
```

```
ACAAACAGTTTCAGCACTTCGGCTTGGACAACTGCTATGCGGAGCAGATATTCCATCAAC
+---------+---------+---------+---------+---------+--------- 5439
TGTTTGTCAAAGTCGTGAAGCCGAACCTGTTGACGATACGCCTCGTCTATAAGGTAGTTG

  K  Q  F  Q  H  F  G  L  D  N  C  Y  A  E  Q  I  F  H  Q  L -

TGCTTTACTTCATTTGCGCTGTGGCCCTTAATTGTCTGATGCTTAGGGGCGATATTTGCA
+---------+---------+---------+---------+---------+--------- 5499
ACGAAATGAAGTAAACGCGACACCGGGAATTAACAGACTACGAATCCCCGCTATAAACGT

  L  Y  F  I  C  A  V  A  L  N  C  L  M  L  R  G  D  I  C  I -

TATGGGAGACTGGCATGATAATCCGCTACAATATCGGCTGCATTGAGGATTGGGTGCGCA
+---------+---------+---------+---------+---------+--------- 5559
ATACCCTCTGACCGTACTATTAGGCGATGTTATAGCCGACGTAACTCCTAACCCACGCGT
      <------------------ 926Y
  W  E  T  G  M  I  I  R  Y  N  I  G  C  I  E  D  W  V  R  S -

GTAAAAAGATGTCTAACGATGTGCTGACAGCTTTGGCGCCTCTGAATCAGGTCTCCCAAT
+---------+---------+---------+---------+---------+--------- 5619
CATTTTTCTACAGATTGCTACACGACTGTCGAAACCGCGGAGACTTAGTCCAGAGGGTTA

  K  K  M  S  N  D  V  L  T  A  L  A  P  L  N  Q  V  S  Q  L -

TGCTGCAGTCTCGGAAGAGCGAGCAGGATGTTCAGACCATTTGTGATCTGTGTACTTCTC
+---------+---------+---------+---------+---------+--------- 5679
ACGACGTCAGAGCCTTCTCGCTCGTCCTACAAGTCTGGTAAACACTAGACACATGAAGAG

  L  Q  S  R  K  S  E  Q  D  V  Q  T  I  C  D  L  C  T  S  L -

TGAGCACGGCGCAGGTCCTCAAGGTGATGAAGTCCTACAAACTGGACGATTATGAGAGCG
+---------+---------+---------+---------+---------+--------- 5739
ACTCGTGCCGCGTCCAGGAGTTCCACTACTTCAGGATGTTTGACCTGCTAATACTCTCGC

  S  T  A  Q  V  L  K  V  M  K  S  Y  K  L  D  D  Y  E  S  E -

AAATAACGAACGTTTTTCTGGAGAAACTAACCGAGAAACTGAACGCCCGACAAATGCAAA
+---------+---------+---------+---------+---------+--------- 5799
TTTATTGCTTGCAAAAAGACCTCTTTGATTGGCTCTTTGACTTGCGGGCTGTTTACGTTT

  I  T  N  V  F  L  E  K  L  T  E  K  L  N  A  R  Q  M  Q  K -

AGAGCAATAGTGACGAATTCACCATAGACCAGAAGTTCATTCAGCCATTTAAGGTTGTCT
+---------+---------+---------+---------+---------+--------- 5859
TCTCGTTATCACTGCTTAAGTGGTATCTGGTCTTCAAGTAAGTCGGTAAATTCCAACAGA

  S  N  S  D  E  F  T  I  D  Q  K  F  I  Q  P  F  K  V  V  F -

TCAGGTATAGTGACATCAAGCTGGAGGATATTGAACTACCGTCGCATCTTAATCTGGACG
+---------+---------+---------+---------+---------+--------- 5919
AGTCCATATCACTGTAGTTCGACCTCCTATAACTTGATGGCAGCGTAGAATTAGACCTGC

  R  Y  S  D  I  K  L  E  D  I  E  L  P  S  H  L  N  L  D  E -
       3'UTR5' ----------------->
```

```
AGTTCCTTACAAAGATTTAAACGTCGCGTGCCTGCGTCTTGCCGGTCGAGAACTTTGCCA
+---------+---------+---------+---------+---------+--------- 5979
TCAAGGAATGTTTCTAAATTTGCAGCGCACGGACGCAGAACGGCCAGCTCTTGAAACGGT

  F   L   T   K   I   *

TTTTGGATGACATTACACTGCTTAGTTTTTGTGTATTATACGAGTATAACGATTCTAGAA
+---------+---------+---------+---------+---------+--------- 6039
AAAACCTACTGTAATGTGACGAATCAAAAACACATAATATGCTCATATTGCTAAGATCTT

                                        925Y ------------------>
ATTTAATTTGTTCTTATATTTAGAAAATCGATTAGATAATGAAACATTGAGTAGAGCCGT
+---------+---------+---------+---------+---------+--------- 6099
TAAATTAAACAAGAATATAAATCTTTTAGCTAATCTATTACTTTGTAACTCATCTCGGCA

TTTGCCGAAATGTTTGTGTAGTTCCTGTTTAGCTTAAATACCCGTTGTTTATCCACTCCA
+---------+---------+---------+---------+---------+--------- 6159
AAACGGCTTTACAAACACATCAAGGACAAATCGAATTTATGGGCAACAAATAGGTGAGGT

ATAGTATCCATTCCGAGTATCCAATGGCGTACCCGTCCCGAGATGCCAAGTGTTGTTTTT
+---------+---------+---------+---------+---------+--------- 6219
TATCATAGGTAAGGCTCATAGGTTACCGCATGGGCAGGGCTCTACGGTTCACAACAAAAA

TGTTGTACGAACTGTTTGTAGCAAAAGTAATGGAAGGTTAAGAACCAACCTTAATATGTG
+---------+---------+---------+---------+---------+--------- 6279
ACAACATGCTTGACAAACATCGTTTTCATTACCTTCCAATTCTTGGTTGGAATTATACAC

  2796 ------------------>
AAACGCAATAAATGCTTGGAGTTAGACTCTGACCATAAGAAATCATTCCTTGCCTAAACC
+---------+---------+---------+---------+---------+--------- 6339
TTTGCGTTATTTACGAACCTCAATCTGAGACTGGTATTCTTTAGTAAGGAACGGATTTGG
                                              2800 <------------

TTGTACCTTAAATCATAACTAAATATTTATATACAATGATTGTACTTAACCAGGCTGATT
+---------+---------+---------+---------+---------+--------- 6399
AACATGGAATTTAGTATTGATTTATAAATATATGTTACTAACATGAATTGGTCCGACTAA
-----
                                        2950 ------------------>
TGTTGCCTTAAAGACGTGTAAAGCCTACAGATGATAGAGTAAGACTCATAGAGCTCACCG
+---------+---------+---------+---------+---------+--------- 6459
ACAACGGAATTTCTGCACATTTCGGATGTCTACTATCTCATTCTGAGTATCTCGAGTGGC

ATGATGCATTGCTGTTGCATTCCATCTATTTATGTATTCTCATTTTCTGCGTAGCTTGTA
+---------+---------+---------+---------+---------+--------- 6519
TACTACGTAACGACAACGTAAGGTAGATAAATACATAAGAGTAAAAGACGCATCGAACAT

ACTCGTTAGTTGGGTTTCTTAGATATGTCGTTAGACTATGCAAGCCTAAACTACGTTGGC
+---------+---------+---------+---------+---------+--------- 6579
TGAGCAATCAACCCAAAGAATCTATACAGCAATCTGATACGTTCGGATTTGATGCAACCG

ATTTGTTTGAAACATTCTAAGCAGATTACCTAAACACATACAAATATTATATATTAGAAA
+---------+---------+---------+---------+---------+--------- 6639
TAAACAAACTTTGTAAGATTCGTCTAATGGATTTGTGTATGTTTATAATATATAATCTTT
                          3'UTR3'-1 <--------------------
```


GAAAGTACTCGTGTGTAATGAACAGCAATTTATGAATTTTTACAAATACAATCTCGAGTC
+---------+---------+---------+---------+---------+--------- 6699
CTTTCATGAGCACACATTACTTGTCGTTAAATACTTAAAAATGTTTATGTTAGAGCTCAG
3'UTR3'-2 <----------
GTTAAC.AAAAAAAAAAAAAAAGCTT
+-----------+---------+--- 6733
CAATTG.TTTTTTTTTTTTTTTCGAA

Appendix II Genomic Sequence of the *didum* Gene

Uncorrected genomic DNA sequence of the *didum* gene beginning at the *Xba*I site of pP1-X7. Introns have been identified in lower case lettering. A polymorphism identified in the 3'UTR has been noted as underlined lower case sequence (nt 7518-7537).

*Xba*I
TCTAGAGGGGATCGGGCATAACTCCCATTCATAAAACTTGTTAATTCGGCTCACTTCTTT
---------+---------+---------+---------+---------+---------+ 60
AGATCTCCCCTAGCCCGTATTGAGGGTAAGTATTTTGAACAATTAAGCCGAGTGAAGAAA

CGCAGGGCGCCAAGATCTGGGTGCCCCATGCGGATCTGGTGTGGGAGAGCGCCACCTTGG
---------+---------+---------+---------+---------+---------+ 120
GCGTCCCGCGGTTCTAGACCCACGGGGTACGCCTAGACCACACCCTCTCGCGGTGGAACC

AGGAGAGCTACCGCAAGGGCGCCGGCTTCTTGAAGATATGTACGGACTCCGGAAAACTGA
---------+---------+---------+---------+---------+---------+ 180
TCCTCTCGATGGCGTTCCCGCGGCCGAAGAACTTCTATACATGCCTGAGGCCTTTTGACT

AAGAGGTCAAGCTAAAGGCCGATGGCAGCGATCTGCCTCCACTGCGCAATCCGGCCATTC
---------+---------+---------+---------+---------+---------+ 240
TTCTCCAGTTCGATTTCCGGCTACCGTCGCTAGACGGAGGTGACGCGTTAGGCCGGTAAG

TGGTGGGACAGAACGACTTGACCACCCTGTCCTACCTGCATGAGCCGGGGGTGTTGCACA
---------+---------+---------+---------+---------+---------+ 300
ACCACCCTGTCTTGCTGAACTGGTGGGACAGGATGGACGTACTCGGCCCCCACAACGTGT

ATCTGCGTGTCCGCTTCTGCGAGCGCCAGATTATCTACACCTACTGCGGCATCATTCTGG
---------+---------+---------+---------+---------+---------+ 360
TAGACGCACAGGCGAAGACGCTCGCGGTCTAATAGATGTGGATGACGCCGTAGTAAGACC

TGGCCATCAACCCGTACGCGGAGATGCCTCTTTACGGGCCCAGCATAATCCGAGCGTATC
---------+---------+---------+---------+---------+---------+ 420
ACCGGTAGTTGGGCATGCGCCTCTACGGAGAAATGCCCGGGTCGTATTAGGCTCGCATAG

GGGGTCATGCTATGGGTGATCTGGAGCCGCACATCTTTGCCCTGGCGGAGGAGGCGTACA
---------+---------+---------+---------+---------+---------+ 480
CCCCAGTACGATACCCACTAGACCTCGGCGTGTAGAAACGGGACCGCCTCCTCCGCATGT

CGAAACTGGAGCGCGAGAACTGCAACCTGAGCATCATCGTCAGTGGGGAATCCGGTGCGG
---------+---------+---------+---------+---------+---------+ 540
GCTTTGACCTCGCGCTCTTGACGTTGGACTCGTAGTAGCAGTCACCCCTTAGGCCACGCC

GCAAAACGGTGTCCGCCAAATACGCCATGAGGTACTTTGCCGCTGTTGGAGGTTCCGAGT
---------+---------+---------+---------+---------+---------+ 600
CGTTTTGCCACAGGCGGTTTATGCGGTACTCCATGAAACGGCGACAACCTCCAAGGCTCA

```
CGGARACCCAGGTCAACGCAGGTGCTGGCATCTTCGCCRATCATGGAAAGCCTTCGGAAA
---------+---------+---------+---------+---------+---------+ 660
GCCTYTGGGTCCAGTTGCGTCCACGACCGTAGAAGCGGYTAGTACCTTTCGGAAGCCTTT

TGCCAAGACRACCCGGAATGACAACAGTTCCCGCTTTGGAAAGTTTACCAAGCTGCTGTT
---------+---------+---------+---------+---------+---------+ 720
ACGGTTCTGYTGGGCCTTACTGTTGTCAAGGGCGAAACCTTTCAAATGGTTCGACGACAA

CCGGAACCAAATGGGTGTGATGTTCCTGCAGGGAGCCACTATGCACACCTACCTACTGGA
---------+---------+---------+---------+---------+---------+ 780
GGCCTTGGTTTACCCACACTACAAGGACGTCCCTCGGTGATACGTGTGGATGGATGACCT

GAAGTCACGTGTGGTGTACCAGGCCCAGGGAGAGCGCAACTATCACATATTCTATCAGCT
---------+---------+---------+---------+---------+---------+ 840
CTTCAGTGCACACCACATGGTCCGGGTCCCTCTCGCGTTGATAGTGTATAAGATAGTCGA

GTGCGCGGCGCGATCGAAGTACCCTGAACTGGTGCTGggtatgtctggaagctccctcct
---------+---------+---------+---------+---------+---------+ 900
CACGCGCCGCGCTAGCTTCATGGGACTTGACCACGACccatacagaccttcgagggagga

atttcagcttagaaaatgaagtaatatttccattctctcaaaGATCACCAGGACAAATTC
---------+---------+---------+---------+---------+---------+ 960
taaagtcgaatcttttacttcattataaaggtaagagagtttCTAGTGGTCCTGTTTAAG

CAGTTTCTGAACATGGGTGGCGCTCCTGAAATTGAACGAGTTTCGGATGCGGAGCAGTTT
---------+---------+---------+---------+---------+---------+ 1020
GTCAAAGACTTGTACCCACCGCGAGGACTTTAACTTGCTCAAAGCCTACGCCTCGTCAAA

AACGAAACCGTGCAGGCCATGACAGTTCTGGGCTTCTCCATTCAACAGATCGCTGATATC
---------+---------+---------+---------+---------+---------+ 1080
TTGCTTTGGCACGTCCGGTACTGTCAAGACCCGAAGAGGTAAGTTGTCTAGCGACTATAG

GTAAAGATCCTGGCAGGGAATACTCCATTTAGGAAACATTCAGGTTTCCAAGAAGTTCAA
---------+---------+---------+---------+---------+---------+ 1140
CATTTCTAGGACCGTCCCTTATGAGGTAAATCCTTTGTAAGTCCAAAGGTTCTTCAAGTT

CGAGGGCAGCGAAGAGGAGGACAGTGACTCTTGCGATATATTTgtaagctcactttcgta
---------+---------+---------+---------+---------+---------+ 1200
GCTCCCGTCGCTTCTCCTCCTGTCACTGAGAACGCTATATAAAcattcgagtgaaagcat

aaacctgatgacaatttctgagttatatatgtactttaacttcatccatctcgtagCATA
---------+---------+---------+---------+---------+---------+ 1260
tttggactactgttaaagactcaatatatacatgaaattgaagtaggtagagcatcGTAT

ACGACATCCACCTGCAGATCACCGCCGATCTACTGCGGGTGAGCGCCGATGATCTGCGCC
---------+---------+---------+---------+---------+---------+ 1320
TGCTGTAGGTGGACGTCTAGTGGCGGCTAGATGACGCCCACTCGCGGCTACTAGACGCGG

GGTGGCTTTTGATGCGTAAGATAGAGTCGGTCAATGAATATGTGCTGATACCGAATAGCA
---------+---------+---------+---------+---------+---------+ 1380
CCACCGAAAACTACGCATTCTATCTCAGCCAGTTACTTATACACGACTATGGCTTATCGT

TTGAGGCGGCTCAGGCGGCTCGAGACGCTCTGGCCAAGCACATCTATGCGAAACTGTTTC
---------+---------+---------+---------+---------+---------+ 1440
AACTCCGCCGAGTCCGCCGAGCTCTGCGAGACCGGTTCGTGTAGATACGCTTTGACAAAG
```

AGTATATAGTCGGTGTGCTGAACAAGAGCCTCAACAACGGTAGCAAGCAGTGCAGCTTCA
---------+---------+---------+---------+---------+---------+ 1500
TCATATATCAGCCACACGACTTGTTCTCGGAGTTGTTGCCATCGTTCGTCACGTCGAAGT

TTGGCGTCCTCGATATCTACGGCTTCGAAACGTTCGAGGTGAACTCCTTTGAACAATTTT
---------+---------+---------+---------+---------+---------+ 1560
AACCGCAGGAGCTATAGATGCCGAAGCTTTGCAAGCTCCACTTGAGGAAACTTGTTAAAA

GCATAAACTATGCSAACGAAAAGCTTCAGCAGCAGTTCAACCAGCATGTCTTCAAGCTGG
---------+---------+---------+---------+---------+---------+ 1620
CGTATTTGATACGSTTGCTTTTCGAAGTCGTCGTCAAGTTGGTCGTACAGAAGTTCGACC

AGCAGGAGGAMTATCTTAAGGAAGGAATCACCTGGACGATGATTGACTTTTACGACAATC
---------+---------+---------+---------+---------+---------+ 1680
TCGTCCTCCTKATAGAATTCCTTCCTTAGTGGACCTGCTACTAACTGAAAATGCTGTTAG

AACCGTGTATCGATCTAATTGAATCTCGCTTGGGAGTGCTGGACCTGCTCGACGAGGAGT
---------+---------+---------+---------+---------+---------+ 1740
TTGGCACATAGCTAGATTAACTTAGAGCGAACCCTCACGACCTGGACGAGCTGCTCCTCA

GTCGAataagattcatttgacatagtacatatcattattaaaccgggtctatctgcagat
---------+---------+---------+---------+---------+---------+ 1800
CAGCTtattctaagtaaactgtatcatgtatagtaataatttggcccagatagacgtcta

gCCAAAGGGCTCGGACAACAGCTGGGCTGGCAAGCTCATCGGAAAGTGCAATAAATTTCC
---------+---------+---------+---------+---------+---------+ 1860
cGGTTTCCCGAGCCTGTTGTCGACCCGACCGTTCGAGTAGCCTTTCACGTTATTTAAAGG

SCATTTCGAGAAGCCACGCTTTGGCACAMCCAgtaagttmttttccgattctatgcacta
---------+---------+---------+---------+---------+---------+ 1920
SGTAAAGCTCTTCGGTGCGAAACCGTGTKGGTcattcaakaaaaggctaagatacgtgat

ttatttcgatgacaatctgttaattctatctttgagGCTTCTTTATCAAACATTTCTCGG
---------+---------+---------+---------+---------+---------+ 1980
aataaagctactgttagacaattaagatagaaactcCGAAGAAATAGTTTGTAAAGAGCC

ACACGGTCGAKTATGACGTGAACGGATTCTTGGAAAAGAATCGTGACACAGTCTCCAAGG
---------+---------+---------+---------+---------+---------+ 2040
TGTGCCAGCTMATACTGCACTTGCCTAAGAACCTTTTCTTAGCACTGTGTCAGAGGTTCC

AATTGACGCAAGTGCTAAGCCGAGTCCAACCATGTCTCTGGCCCAATCMGGGTGGATGAC
---------+---------+---------+---------+---------+---------+ 2100
TTAACTGCGTTCACGATTCGGCTCAGGTTGGTACAGAGACCGGGTTAGKCCCACCTACTG

CCTGGGAAGGAAAATAAAATMCTCCTGTGCGTGGATYCCGCTAAATCCCCCCCCTTGGGC
---------+---------+---------+---------+---------+---------+ 2160
GGACCCTTCCTTTTATTTTAKGAGGACACGCACCTARGGCGATTTAGGGGGGGGAACCCG

GGCCCCTTCTTRATCABTGYTGGCCGCAAACaggtttgtgtttgaccarattcttcatgg
---------+---------+---------+---------+---------+---------+ 2220
CCGGGGAAGAAYTAGTVACRACCGGCGTTTGtccaaacacaaactggtytaagaagtacc

acccaacaatcgtattgctccaatttcgtccaattgtaatcgttgtcttatkgctatttg
---------+---------+---------+---------+---------+---------+ 2280
tgggttgttagcataacgaggttaaagcaggttaacattagcaacagaatamcgataaac

tattttccatacgcttgtcgttcgctcaccgccaactccctcggattgcccaacagcaag
---------+---------+---------+---------+---------+---------+ 2340
ataaaaggtatgcgaacagcaagcgagtggcggttgagggagcctaacgggttgtcgttc

ggaatgacacacgtcgaagagttagttaccaacttatttaaattaacagcaattgactaa
---------+---------+---------+---------+---------+---------+ 2400
ccttactgtgtgcagcttctcaatcaatggttgaataaatttaattgtcgttaactgatt

ccttaaacgaataaccacaaacattttaacaatttaattcgacttggaacatttgtttta
---------+---------+---------+---------+---------+---------+ 2460
ggaatttgcttattggtgtttgtaaaattgttaaattaagctgaaccttgtaaacaaaat

attaaagttactaatttagtccggattcaaatgacaaggctaaatggcccagtaatgctg
---------+---------+---------+---------+---------+---------+ 2520
taatttcaatgattaaatcaggcctaagtttactgttccgatttaccgggtcattacgac

attagaaagaatcaaagttgcttatagactaacaattcastacctccgcaaataaataat
---------+---------+---------+---------+---------+---------+ 2580
taatctttcttagtttcaacgaatatctgattgttaagtsatggaggcgtttatttatta

aataataaatgatgataggaatttgtcaaataggtgattcaatatattatcaataaaata
---------+---------+---------+---------+---------+---------+ 2640
ttattatttactactatccttaaacagtttatccactaagttatataatagttattttat

tgacacgcggacgaacgaacatattaaatgaattgggttaatgagaataagtttactcaa
---------+---------+---------+---------+---------+---------+ 2700
actgtgcgcctgcttgcttgtataatttacttaacccaattactcttattcaaatgagtt

aatgtaaaactagtttgaaactctaaatcgatcttatgtattcattttcattttgactgt
---------+---------+---------+---------+---------+---------+ 2760
ttacattttgatcaaactttgagatttagctagaatacataagtaaaagtaaaactgaca

gtttgcctatttccacacacgacctactaatttcgcctaatacttgttataAGGTGGTGC
---------+---------+---------+---------+---------+---------+ 2820
caaacggataaaggtgtgtgctggatgattaaagcggattatgaacaatatTCCACCACG

CATCCAAGCAGCATAGAAAAACGGTGGGATCGCAGTTCCAGGAGAGTCTGGCGTCGCTGA
---------+---------+---------+---------+---------+---------+ 2880
GTAGGTTCGTCGTATCTTTTTGCCACCCTAGCGTCAAGGTCCTCTCAGACCGCAGCGACT

TATCTACGTTACATGCCACAACTCCGCACTATGTGCGCTGCATCAAGgtaagatgtgctA
---------+---------+---------+---------+---------+---------+ 2940
ATAGATGCAATGTACGGTGTTGAGGCGTGATACACGCGACGTAGTTCcattctacacgaT

ttggatcttaagtacagataaaagctcattaaaatatgtttattttcgtgcagCCCAACG
---------+---------+---------+---------+---------+---------+ 3000
aacctagaattcatgtctattttcgagtaattttatacaaataaaagcacgtcGGGTTGC

ATGACAAAGTCGCCTTTAAGTGGGAGACGGCCAAGATCATACAGCAGTTAAGGGCCTGTG
---------+---------+---------+---------+---------+---------+ 3060
TACTGTTTCAGCGGAAATTCACCCTCTGCCGGTTCTAGTATGTCGTCAATTCCCGGACAC

GTGTGCTGGAAACGGTGCGCATCTTCCGCAGCGGGATTCCCCTCGAGATGGCTCTATCCC
---------+---------+---------+---------+---------+---------+ 3120
CACACGACCTTTGCCACGCGTAGAAGGCGTCGCCCTAAGGGGAGCTCTACCGAGATAGGG

GACTTCTATATGCGGTACCAGCTGCTGGTTTACCGCTCCAAACTCGACAAAAACGACATG
---------+---------+---------+---------+---------+---------+ 3180
CTGAAGATATACGCCATGGTCGACGACCAAATGGCGAGGTTTGAGCTGTTTTTGCTGTAC

AAGCTGTCGTGCCGGAACATTGTGATGAAGTGGATCCAAGACGAAGATAAGTACCGATTT
---------+---------+---------+---------+---------+---------+ 3240
TTCGACAGCACGGCCTTGTAACACTACTTCACCTAGGTTCTGCTTCTATTCATGGCTAAA

GGCAACACGCAGATTTTCTTCCGCGCCGGCCAAGTGGCCTTCCTTGAACAGGTTCGGGCT
---------+---------+---------+---------+---------+---------+ 3300
CCGTTGTGCGTCTAAAAGAAGGCGCGGCCGGTTCACCGGAAGGAACTTGTCCAAGCCCGA

AATCTGCGCAAGAAGTATATCACCATTGTGCAGTCGGTTGTGCGGCGATTCGTCTACCGG
---------+---------+---------+---------+---------+---------+ 3360
TTAGACGCGTTCTTCATATAGTGGTAACACGTCAGCCAACACGCCGCTAAGCAGATGGCC

CGCCAGTTCCTGCGCATTCAGAAAGTAATTAATGGCATTCAGAAACATGCGCGCGGATAT
---------+---------+---------+---------+---------+---------+ 3420
GCGGTCAAGGACGCGTAAGTCTTTCATTAATTACCGTAAGTCTTTGTACGCGCGCCTATA

CTTGCTCGCGAGCGTACTCAGAAAATGCGCGAAGCTCGTGCGGGATTAATCCTGTCGAAG
---------+---------+---------+---------+---------+---------+ 3480
GAACGAGCGCTCGCATGAGTCTTTTACGCGCTTCGAGCACGCCCTAATTAGGACAGCTTC

TACGCCCGCGGTTGGTTATGCCGTCGTCGTTACTTGCGCCTACGCCACTCCATTTCCGGC
---------+---------+---------+---------+---------+---------+ 3540
ATGCGGGCGCCAACCAATACGGCAGCAGCAATGAACGCGGATGCGGTGAGGTAAAGGCCG

ATACAGACCTACGCCCGCGGCATGCTGGCGCGCAACAAGTTCCACGCGATGCGGGATCAC
---------+---------+---------+---------+---------+---------+ 3600
TATGTCTGGATGCGGGCGCCGTACGACCGCGCGTTGTTCAAGGTGCGCTACGCCCTAGTG

TACCGGGCAGTTCAGATCCAGCGTTTCGTGCGTGGTGCTTTGGCACGGCGAGCTTACCAA
---------+---------+---------+---------+---------+---------+ 3660
ATGGCCCGTCAAGTCTAGGTCGCAAAGCACGCACCACGAAACCGTGCCGCTCGAATGGTT

AAGCGGSGTCGCAACATCATCATTTGTCAAGCGGCGATTCGGAGATTCTTGGCCCGTCGT
---------+---------+---------+---------+---------+---------+ 3720
TTCGCCSCAGCGTTGTAGTAGTAAACAGTTCGCCGCTAAGCCTCTAAGAACCGGGCAGCA

AAGTTTAAACGCATGAAGGCCGRGGCCARRACCATCTCGCACATGGAAAACAAAAACATG
---------+---------+---------+---------+---------+---------+ 3780
TTCAAATTTGCGTACTTCCGGCYCCGGTYYTGGTAGAGCGTGTACCTTTTGTTTTTGTAC

GGGCTGGAAAACAAGATTATTTCCATGCAGCAGCGGATCGATGAGCTGAATCGCGACAAC
---------+---------+---------+---------+---------+---------+ 3840
CCCGACCTTTTGTTCTAATAAAGGTACGTCGTCGCCTAGCTACTCGACTTAGCGCTGTTG

AGTAATYTGAAGCACAAGACCAGCGAAATCAGTgtattgaagtataaatttagcatctag
---------+---------+---------+---------+---------+---------+ 3900
TCATTARACTTCGTGTTCTGGTCGCTTTAGTCAcataacttcatatttaaatcgtagatc

tcgttcttgttcgcaggagttaatcatctttatcgtgcagAATGAAGCTTGAGCTGAAGA
---------+---------+---------+---------+---------+---------+ 3960
agcaagaacaagcgtcctcaattagtagaaatagcacgtcTTACTTCGAACTCGACTTCT

AGACCCTGGAGGCTGAATTCAAAAATGTCAAGGCCGCCTGCCAGGACAAGGACAAGCTGA
---------+---------+---------+---------+---------+---------+ 4020
TCTGGGACCTCCGACTTAAGTTTTTACAGTTCCGGCGGACGGTCCTGTTCCTGTTCGACT

TCGAAGCACTTAACAAGCAGTTGGAGGCGGAGCGAGACGAAAAAATGCAGTTGCTGGAGG
---------+---------+---------+---------+---------+---------+ 4080
AGCTTCGTGAATTGTTCGTCAACCTCCGCCTCGCTCTGCTTTTTTACGTCAACGACCTCC

AGAACGGACATGCTCAAGAGGAGTGGATCAGCCAGAAGCAGACTTGGCGCCAAGAGAACG
---------+---------+---------+---------+---------+---------+ 4140
TCTTGCCTGTACGAGTTCTCCTCACCTAGTCGGTCTTCGTCTGAACCGCGGTTCTCTTGC

AGGAGCTGCGCCGTCAGATAAACGAGATAATCGATATGGCAAAGAACGCAGAAGTCAACC
---------+---------+---------+---------+---------+---------+ 4200
TCCTCGACGCGGCAGTCTATTTGCTCTATTAGCTATACCGTTTCTTGCGTCTTCAGTTGG

AGSGTAACCRGGAGGACCGAATGCTAGCCGAGATTGATAACAGGGAGCTCAACGAGGCCT
---------+---------+---------+---------+---------+---------+ 4260
TCSCATTGGYCCTCCTGGCTTACGATCGGCTCTAACTATTGTCCCTCGAGTTGCTCCGGA

ACCAACGAGCTATTAAGGACAAGGAGGTCATCGAGAACGAAAACTTCATGCTGAAGGAAG
---------+---------+---------+---------+---------+---------+ 4320
TGGTTGCTCGATAATTCCTGTTCCTCCAGTAGCTCTTGCTTTTGAAGTACGACTTCCTTC

AGCTCAGTCGATTAACGGCTGGCAGTTTCAGTTTGCACGCCCGCAAGGCTAGCAACGCCT
---------+---------+---------+---------+---------+---------+ 4380
TCGAGTCAGCTAATTGCCGACCGTCAAAGTCAAACGTGCGGGCGTTCCGATCGTTGCGGA

CCAGCCAAAACGAGGACGATGTGGGATACGCCTCCGCCAAGAACACTCTGGATATCAATC
---------+---------+---------+---------+---------+---------+ 4440
GGTCGGTTTTGCTCCTGCTACACCCTATGCGGAGGCGGTTCTTGTGAGACCTATAGTTAG

GGCCCCCGGATTTGTTAAGCMAAAATTgtaagaacattacctgttgtytagtgtgtgtag
---------+---------+---------+---------+---------+---------+ 4500
CCGGGGGCCTAAACAATTCGKTTTTAAcattcttgtaatggacaacaratcacacacatc

agttgtaatccgtaaatttgtatgtttagACTCGTACAATGACTCTACCAGTCTGGTGGT
---------+---------+---------+---------+---------+---------+ 4560
tcaacattaggcatttaaacatacaaatcTGAGCATGTTACTGAGATGGTCAGACCACCA

GAAGTTGAGATCCATTCTCGAGGAGGAGAAGCAAAAGCACAAGGTCTTGCAGGAGCAGTA
---------+---------+---------+---------+---------+---------+ 4620
CTTCAACTCTAGGTAAGAGCTCCTCCTCTTCGTTTTCGTGTTCCAGAACGTCCTCGTCAT

CATTAAGTTGTCCAGTCGGCATAAGCCCACCGAGGGATTCCTTCCggtaagtatttaggg
---------+---------+---------+---------+---------+---------+ 4680
GTAATTCAACAGGTCAGCCGTATTCGGGTGGCTCCCTAAGGAAGGccattcataaatccc

aatccctcctaaatagacacaatatcgccgagctaaaatttcggtgttacttccccactt
---------+---------+---------+---------+---------+---------+ 4740
ttagggaggatttatctgtgttatagcggctcgattttaaagccacaatgaaggggtgaa

atcgctactaaatcttgaaaaaaactccattkgtctacatctatgtattattgccgcaca
---------+---------+---------+---------+---------+---------+ 4800
tagcgatgatttagaacttttttgaggtaamcagatgtagatacataataacggcgtgt

```
ctgctctccactcacataGCGTCTCCGAGCTTGAGGTAGAGAATGAAAAGCTGCGCAGCG
---------+---------+---------+---------+---------+---------+ 4860
gacgagaggtgagtgtatCGCAGAGGCTCGAACTCCATCTCTTACTTTTCGACGCGTCGC

AGTACGATCAGCTGCGAACGAGCATTAAACACGGTGTTGAGATCAACGAGCTCAATggta
---------+---------+---------+---------+---------+---------+ 4920
TCATGCTAGTCGACGCTTGCTCGTAATTTGTGCCACAACTCTAGTTGCTCGAGTTAccat

acctcccacgcacgccgtgcgatccacggagcggtcttactgacacacgtctttttccc
---------+---------+---------+---------+---------+---------+ 4980
tggagggtgcgtgcggcacgctaggtgcctcgccagaatgactgtgtgcagaaaaaaggg

CtcctccttccaatctttgtcattgcAGCACAGCATGCCGCCTTGCAGGAAGRAGTAMGT
---------+---------+---------+---------+---------+---------+ 5040
GaggaggaaggttagaaacagtaacgTCGTGTCGTACGGCGGAACGTCCTTCYTCATKCA

AGGCGGCGCGAGGAGTGCATCCAATTAAAGGCAGTCCTGCTGCAGCAGARCCAGTCCATG
---------+---------+---------+---------+---------+---------+ 5100
TCCGCCGCGCTCCTCACGTAGGTTAATTTCCGTCAGGACGACGTCGTCTYGGTCAGGTAC

AGATTGCTTGAGCCGGAAAGTTTACAGATGGGTGGCAACGATGTTAACGAAATGATGGAA
---------+---------+---------+---------+---------+---------+ 5160
TCTAACGAACTCGGCCTTTCAAATGTCTACCCACCGTTGCTACAATTGCTTTACTACCTT

GCCTTCCATTCCCAGRAGGTAATTAATyggtgagtgaaattggttgamtgtatgcccaaa
---------+---------+---------+---------+---------+---------+ 5220
CGGAAGGTAAGGGTCYTCCATTAATTArccactcactttaaccaactkacatacgggttt

gcatcatttttttttttaaattccMGTCAATTGGRGTTTGAGGTTAAGGCCATCACCGAG
---------+---------+---------+---------+---------+---------+ 5280
cgtagtaaaaaaaaaaatttaaggKCAGTTAACCYCAAACTCCAATTCCGGTAGTGGCTC

GGGCACAACAGTAAGGTTGTGGAGRTGACACAGGAGATTGAGAGATTGRACMATGAGAAG
---------+---------+---------+---------+---------+---------+ 5340
CCCGTGTTGTCATTCCAACACCTCYACTGTGTCCTCTAACTCTCTAACYTGKTACTCTTC

GATGAGCTGCCAAAGGTAATGTTTGAGAGCATTGACGAGTTTGAAGATTCCCAWGTGGAT
---------+---------+---------+---------+---------+---------+ 5400
CTACTCGACGGTTTCCATTACAAACTCTCGTAACTGCTCAAACTTCTAAGGGTWCACCTA

AAGCTGAGACAGAACGATTGSTATTTGCGACGAGAAATGCAGAAGGCTGTAGCCCAGTTC
---------+---------+---------+---------+---------+---------+ 5460
TTCGACTCTGTCTTGCTAACSATAAACGCTGCTCTTTACGTCTTCCGACATCGGGTCAAG

CTGCTTGTTCAGGAGGAGCTCAARCTGGCCAATGCCMAGCTTAAAGCTTATTGGCAGGAT
---------+---------+---------+---------+---------+---------+ 5520
GACGAACAAGTCCTCCTCGAGTTYGACCGGTTACGGKTCGAATTTCGAATAACCGTCCTA

GGAGGCCAGCTGGAGCACMAGATAGAGGAGGAGATGATTSGCAACAAGTCCAACGGAAAC
---------+---------+---------+---------+---------+---------+ 5580
CCTCCGGTCGACCTCGTGKTCTATCTCCTCCTCTACTAASCGTTGTTCAGGTTGCCTTTG

NTCCSCCGATGTAGGCGCGAATGTGASGAAGCGMAATTTYTGCADHHSCVAAGGGYBTAA
---------+---------+---------+---------+---------+---------+ 5640
NAGGSGGCTACATCCGCGCTTACACTSCTTCGCKTTAAARACGTHDDSGBTTCCCRVATT
```

```
AAAGTTCCACVGCAGCGATCTGGACAARATHTTGCAACGCCTGGTTAGCCCCTTGATTCC
---------+---------+---------+---------+---------+---------+ 5700
TTTCAAGGTGBCGTCGCTAGACCTGTTYTADAACGTTGCGGACCAATCGGGGAACTAAGG

ACGCACAGTGGTCGGGBTCHTTCCTGGTTTTCCAGCATATCTCATCTTTATGTGTATTCg
---------+---------+---------+---------+---------+---------+ 5760
TGCGTGTCACCAGCCCVAGDAAGGACCAAAAGGTCGTATAGAGTAGAAATACACATAAGc

gtgagttatggtcatggttgtggatgatggttaagttgattaattttatggctttttat
---------+---------+---------+---------+---------+---------+ 5820
cactcaataccagtaccaacacctactaccaattcaactaattaaaataccgaaaaaata

cacAGATACACCGATCTGACAAATGCCGACGATGATGTGCGCGAGTTGCTAAGCAAGTTC
---------+---------+---------+---------+---------+---------+ 5880
gtgTCTATGTGGCTAGACTGTTTACGGCTGCTACTACACGCGCTCAACGATTCGTTCAAG

GTTATTCAGATTAAGAAAATGCATCGTACGCCGCATCCGATCGAGAATCGTGTTATTTGG
---------+---------+---------+---------+---------+---------+ 5940
CAATAAGTCTAATTCTTTTACGTAGCATGCGGCGTAGGCTAGCTCTTAGCACAATAAACC

CTCGTCAATTCCAtaacgtgagttacaaaggtcataagctcttggaattcataaattaat
---------+---------+---------+---------+---------+---------+ 6000
GAGCAGTTAAGGTattgcactcaatgtttccagtattcgagaaccttaagtatttaatta

ccatttccgcgacctTTAGGCTGCTAAATCTTATGAAGCAATACGGCGACGTGGATGAGT
---------+---------+---------+---------+---------+---------+ 6060
ggtaaaggcgctggaAATCCGACGATTTAGAATACTTCGTTATGCCGCTGCACCTACTCA

ACGTCAAGTTCAATACTGAGAAGCAGAATCAGCAGCAGCTGAAGAACTTCAATCTCTTTG
---------+---------+---------+---------+---------+---------+ 6120
TGCAGTTCAAGTTATGACTCTTCGTCTTAGTCGTCGTCGACTTCTTGAAGTTAGAGAAAC

AATACCGTCGCGTAATTCTTGATTTAATTGTGAACCTGTACCAGGCGCTGATCATGCAGA
---------+---------+---------+---------+---------+---------+ 6180
TTATGGCAGCGCATTAAGAACTAAATTAACACTTGGACATGGTCCGCGACTAGTACGTCT

TCCAGGGTCTGTTGGACCCAAAAATAGTGCCAGCGATTCTCAACAATGATGAGATTCAGC
---------+---------+---------+---------+---------+---------+ 6240
AGGTCCCAGACAACCTGGGTTTTTATCACGGTCGCTAAGAGTTGTTACTACTCTAAGTCG

GTGGGCGGCAGGCGCACGGAATGCGTAGTCGGGCCACGTCGATTGGAGCATCCTCGTCAC
---------+---------+---------+---------+---------+---------+ 6300
CACCCGCCGTCCGCGTGCCTTACGCATCAGCCCGGTGCAGCTAACCTCGTAGGAGCAGTG

CGGAGCACGGTGGCGGTCCGGCCTGGAAGCAACTGATCGGGCAGCTGGAGCATTTCTACA
---------+---------+---------+---------+---------+---------+ 6360
GCCTCGTGCCACCGCCAGGCCGGACCTTCGTTGACTAGCCCGTCGACCTCGTAAAGATGT

AACAGTTCCAGCACTTCGGCTTGGACAACTGCTATGCGGAGCAGATATTCCATCAACTGC
---------+---------+---------+---------+---------+---------+ 6420
TTGTCAAGGTCGTGAAGCCGAACCTGTTGACGATACGCCTCGTCTATAAGGTAGTTGACG

TTTACTTCATTTGCGCTGTGGCCCTTAATTGTCTGATGCTTAGGGGCGATATTTGCATGT
---------+---------+---------+---------+---------+---------+ 6480
AAATGAAGTAAACGCGACACCGGGAATTAACAGACTACGAATCCCCGCTATAAACGTACA
```

```
GGGAGACTGGCATGATAATCCGCTACAATATCGGCTGCATTGAGGATTGGGTGCGCAGTA
---------+---------+---------+---------+---------+---------+ 6540
CCCTCTGACCGTACTATTAGGCGATGTTATAGCCGACGTAACTCCTAACCCACGCGTCAT

AAAAGATggtaggtcacgtttgcttattgccaagtttccaaactaactagcttttggttt
---------+---------+---------+---------+---------+---------+ 6600
TTTTCTAccatccagtgcaaacgaataacggttcaaaggtttgattgatcgaaaaccaaa

ttctaGTCTAACGATGTGCTGACAGCTTTGGCGCCTCTGAATCAGGTCTCCCAATTGCTG
---------+---------+---------+---------+---------+---------+ 6660
aagatCAGATTGCTACACGACTGTCGAAACCGCGGAGACTTAGTCCAGAGGGTTAACGAC

CAGTCTCGGAAGAGCGAGCAGGATGTTCAGACCATTTGTGATATGTGTACTTCTCTGAGC
---------+---------+---------+---------+---------+---------+ 6720
GTCAGAGCCTTCTCGCTCGTCCTACAAGTCTGGTAAACACTATACACATGAAGAGACTCG

ACGGCGCAGGTCCTCAAGGTGATGAAGTCCTACAAACTGGACGATTATGAGAGCGAAATA
---------+---------+---------+---------+---------+---------+ 6780
TGCCGCGTCCAGGAGTTCCACTACTTCAGGATGTTTGACCTGCTAATACTCTCGCTTTAT

ACGAACGTTTTTCTGGAGAAACTAACCGAGAAACTGAACGCCCGACAAATggtacgtgta
---------+---------+---------+---------+---------+---------+ 6840
TGCTTGCAAAAAGACCTCTTTGATTGGCTCTTTGACTTGCGGGCTGTTTAccatgcacat

aaatctcgaaatatacactttcgatttaccgattttgtcttacAGCAAAAGAGCAATAGT
---------+---------+---------+---------+---------+---------+ 6900
tttagagctttatatgtgaaagctaaatggctaaaacagaatgTCGTTTTCTCGTTATCA

GACGAATTCACCATAGACCAGAAGTTCATTCAGCCATTTAAGGTTGTCTTCAGGTATAGT
---------+---------+---------+---------+---------+---------+ 6960
CTGCTTAAGTGGTATCTGGTCTTCAAGTAAGTCGGTAAATTCCAACAGAAGTCCATATCA

GACATCAAGCTGGAGGATATTGAACTACCGTCGCATCTTAATCTGGACGAGTTCCTTACA
---------+---------+---------+---------+---------+---------+ 7020
CTGTAGTTCGACCTCCTATAACTTGATGGCAGCGTAGAATTAGACCTGCTCAAGGAATGT

AAGATTTAAACGTCGCGTGCCTGCGTCTTGCCGGTCGAGAACTTTGCCATTTGGGATGAC
---------+---------+---------+---------+---------+---------+ 7080
TTCTAAATTTGCAGCGCACGGACGCAGAACGGCCAGCTCTTGAAACGGTAAACCCTACTG

ATTACACTGCTTAGTTTTTGTGTATTTATACGAGTATAACGATTCTAGAAATTTAATTTG
---------+---------+---------+---------+---------+---------+ 7140
TAATGTGACGAATCAAAAACACATAAATATGCTCATATTGCTAAGATCTTTAAATTAAAC

TTCTTATATTTAGAAAATCGATTAGATAATGAAACATTGAGTAGAGCCGTTTTGCCGAAA
---------+---------+---------+---------+---------+---------+ 7200
AAGAATATAAATCTTTTAGCTAATCTATTACTTTGTAACTCATCTCGGCAAAACGGCTTT

TGTTTGTGTAGTTCCTGTTTAGCTTAAATACCCGTTGTTTATCCACTCCAATAGTATCCA
---------+---------+---------+---------+---------+---------+ 7260
ACAAACACATCAAGGACAAATCGAATTTATGGGCAACAAATAGGTGAGGTTATCATAGGT

TTCCGAGTATCCAATGGCGTACCCGTCCCGAGATGCCAAGTGTTGTTTTTTGTTGTACGA
---------+---------+---------+---------+---------+---------+ 7320
AAGGCTCATAGGTTACCGCATGGGCAGGGCTCTACGGTTCACAACAAAAAACAACATGCT
```

```
ACTATTTGTAGCAAAAGTAATGGAAGGTTAAGAACCAACCTTAATATGTGAAACGCAATA
---------+---------+---------+---------+---------+---------+ 7380
TGATAAACATCGTTTTCATTACCTTCCAATTCTTGGTTGGAATTATACACTTTGCGTTAT

AATGCTTGGAGTTAGACTCTGACCATAAGAAATCATTCCTTGCCTAAACCTTGTACCTTA
---------+---------+---------+---------+---------+---------+ 7440
TTACGAACCTCAATCTGAGACTGGTATTCTTTAGTAAGGAACGGATTTGGAACATGGAAT

AATCATAACTAAATATTTATATACAATGATTGTACTTAACCAGGCTGATTTGTTGCCTTA
---------+---------+---------+---------+---------+---------+ 7500
TTAGTATTGATTTATAAATATATGTTACTAACATGAATTGGTCCGACTAAACAACGGAAT

AAGAACGTGTAAAGACTacagatgatagagtaagactCATAGAGCTCACCGATGATGCAT
---------+---------+---------+---------+---------+---------+ 7560
TTCTTGCACATTTCTGAtgtctactatctcattctgaGTATCTCGAGTGGCTACTACGTA

TGCTGTTGCATTCCATCTATTTATGTATTCTCATTTTCTGCGTAGCTTGTAACTCGTTAG
---------+---------+---------+---------+---------+---------+ 7620
ACGACAACGTAAGGTAGATAAATACATAAGAGTAAAAGACGCATCGAACATTGAGCAATC

TTGGGTTTCTTAGATATGTCGTTAGACTATGCAAGCCTAAACTACGTTGGCATTTGTTTG
---------+---------+---------+---------+---------+---------+ 7680
AACCCAAAGAATCTATACAGCAATCTGATACGTTCGGATTTGATGCAACCGTAAACAAAC

AAACATTCTAAGCAGATTACCTAAACACATACAAATATTATATATTAGAAAGAAAGTACT
---------+---------+---------+---------+---------+---------+ 7740
TTTGTAAGATTCGTCTAATGGATTTGTGTATGTTTATAATATATAATCTTTCTTTCATGA

C.GTGTGTAATGAACAGCAATTTATGAATTTTTACAAATACAATCTCGAATCGTTAACaa
---------+---------+---------+---------+---------+---------+ 7800
G.CACACATTACTTGTCGTTAAATACTTAAAAATGTTTATGTTAGAGCTTAGCAATTGtt

tccacgaaracttttgataacttgttcttgggcaggtcaagggacggcaggcgctgcgtc
---------+---------+---------+---------+---------+---------+ 7860
aggtgcttytgaaaactattgaacaagaacccgtccagttccctgccgtccgcgacgcag

accataccggtagccaccgtcgctccgttctcgcsaattgtgaaggcttggcccggartc
---------+---------+---------+---------+---------+---------+ 7920
tggtatggccatcggtggcagcgaggcaagagcgsttaacacttccgaaccgggcctyag

atgaccattttccgcagcagcgtgacgcscacctggccatgctcacccggcataagcata
---------+---------+---------+---------+---------+---------+ 7980
tactggtaaaaggcgtcgtcgcactgcgsgtggaccggtacgagtgggccgtattcgtat

gcctcacttggaactgcaat
---------+---------+ 8000
cggagtgaaccttgacgtta
```

Appendix III cDNA Sequence of cDNA-34

DNA sequence of cDNA-34 encoding a putative sodium-dependent inorganic phosphate cotransporter. The translated predicted open reading frame is shown from nt 439 to nt 2013. Oligonucleotide primers have been marked at their respective positions.

```
AAGCTTTCTTCAGGAATATCTAGTCTTCATCTCTCGTAACGCGATAACGAAACCCACAGG
---------+---------+---------+---------+---------+---------+ 60
TTCGAAAGAAGTCCTTATAGATCAGAAGTAGAGAGCATTGCGCTATTGCTTTGGGTGTCC

GGAGCACGCCATATCATAAAGATAGTCAAACAGCTGACTGATCCGGTCTGCCTAAGTGCG
---------+---------+---------+---------+---------+---------+ 120
CCTCGTGCGGTATAGTATTTCTATCAGTTTGTCGACTGACTAGGCCAGACGGATTCACGC

AATTTCAACTGAGCGCTGCCGATCGCCGAACTCGTCGTTGTCTGTCGCATGTTGGGCAGG
---------+---------+---------+---------+---------+---------+ 180
TTAAAGTTGACTCGCGACGGCTAGCGGCTTGAGCAGCAACAGACAGCGTACAACCCGTCC

TTGATGAGCGATCACGAGCCGAGCTGCAGCATAGTCCGCCAGATCTGCGGCGCTTATCAG
---------+---------+---------+---------+---------+---------+ 240
AACTACTCGCTAGTGCTCGGCTCGACGTCGTATCAGGCGGTCTAGACGCCGCGAATAGTC

TGGCGTCCGGCTGGGATAAAGAGATACAGATACAGATACAGGCGGATCGACTGGAGGCAG
---------+---------+---------+---------+---------+---------+ 300
ACCGCAGGCCGACCCTATTTCTCTATGTCTATGTCTATGTCCGCCTAGCTGACCTCCGTC

ACTGGGTCTGGCATTCTTTTCCGGCAGCTGCCAGCAGCAGTTGATCGTCAGCGATCGAAT
---------+---------+---------+---------+---------+---------+ 360
TGACCCAGACCGTAAGAAAAGGCCGTCGACGGTCGTCGTCAACTAGCAGTCGCTAGCTTA

CACAGGGATTGGCGGCAGCTCCTGCAGATGCAGAGTTCAAATTTCCAGTGAAGTTCTCAA
---------+---------+---------+---------+---------+---------+ 420
GTGTCCCTAACCGCCGTCGAGGACGTCTACGTCTCAAGTTTAAAGGTCACTTCAAGAGTT

CCAGTCAATCCATTCACCATGGGCCAAGTCGAGGCGCGCACTGTGCTGTGGTACATGACA
---------+---------+---------+---------+---------+---------+ 480
GGTCAGTTAGGTAAGTGGTACCCGGTTCAGCTCCGCGCGTGACACGACACCATGTACTGT

                  M  G  Q  V  E  A  R  T  V  L  W  Y  M  T   -

TTCATCGGTTTCATCGTGAACTACATGATCCGGATCAATCTGAACATTACCATTGTGGAC
---------+---------+---------+---------+---------+---------+ 540
AAGTAGCCAAAGTAGCACTTGATGTACTAGGCCTAGTTAGACTTGTAATGGTAACACCTG

F  I  G  F  I  V  N  Y  M  I  R  I  N  L  N  I  T  I  V  D   -
```

```
ATGATTGCCGGGAAGGGTGCTATTACCTCAAATGAAACCCACGAGAACTCTACGGACTTG
---------+---------+---------+---------+---------+---------+ 600
TACTAACGGCCCTTCCCACGATAATGGAGTTTACTTTGGGTGCTCTTGAGATGCCTGAAC

M  I  A  G  K  G  A  I  T  S  N  E  T  H  E  N  S  T  D  L   -

GCTGCCCTTGCGGAAATGAATGAACGATTTTCGCTGGAGCGTTGGTTTTTGGACTGGGCG
---------+---------+---------+---------+---------+---------+ 660
CGACGGGAACGCCTTTACTTACTTGCTAAAAGCGACCTCGCAACCAAAAACCTGACCCGC

A  A  L  A  E  M  N  E  R  F  S  L  E  R  W  F  L  D  W  A   -

AATATTCCGTACGAAAAGAATGGATTTCACTGGAACGAGAAGCAGCAGGGCGCTCTGTTG
---------+---------+---------+---------+---------+---------+ 720
TTATAAGGCATGCTTTTCTTACCTAAAGTGACCTTGCTCTTCGTCGTCCCGCGAGACAAC

N  I  P  Y  E  K  N  G  F  H  W  N  E  K  Q  Q  G  A  L  L   -

GGATCATTTTTCTGGGCCCACTGGACACTGCAGATTCCCGGCGGCATCTTGGCCACCAAA
---------+---------+---------+---------+---------+---------+ 780
CCTAGTAAAAAGACCCGGGTGACCTGTGACGTCTAAGGGCCGCCGTAGAACCGGTGGTTT

G  S  F  F  W  A  H  W  T  L  Q  I  P  G  G  I  L  A  T  K   -

TATGGTACTAAACTGGTGTTTGGTTGGTCCAATGGCATCGGTGTGTTCTGCTGTTTCCTT
---------+---------+---------+---------+---------+---------+ 840
ATACCATGATTTGACCACAAACCAACCAGGTTACCGTAGCCACACAAGACGACAAAGGAA

Y  G  T  K  L  V  F  G  W  S  N  G  I  G  V  F  C  C  F  L   -

ATACCAATCGTATCGTACTGGAGCTACACGGGCTTGATTATCCTGCGAGTATTCCAGGGG
---------+---------+---------+---------+---------+---------+ 900
TATGGTTAGCATAGCATGACCTCGATGTGCCCGAACTAATAGGACGCTCATAAGGTCCCC

I  P  I  V  S  Y  W  S  Y  T  G  L  I  I  L  R  V  F  Q  G   -

TGGATAACGGGCTTGGCCTGGCCATCGATGCACGTGCTCACTGCCAAATGGATTCCGCCC
---------+---------+---------+---------+---------+---------+ 960
ACCTATTGCCCGAACCGGACCGGTAGCTACGTGCACGAGTGACGGTTTACCTAAGGCGGG

W  I  T  G  L  A  W  P  S  M  H  V  L  T  A  K  W  I  P  P   -

AACGAGCGCAGCAAGTTTGTCAGTGCTTACTTGGGCAGTTCTGTGGGCGTGGCCCTGTTC
---------+---------+---------+---------+---------+---------+ 1020
TTGCTCGCGTCGTTCAAACAGTCACGAATGAACCCGTCAAGACACCCGCACCGGGACAAG

N  E  R  S  K  F  V  S  A  Y  L  G  S  S  V  G  V  A  L  F   -

TATCCGATCTTCGGCTACATCATCGATTGGACACGGTGGGAGTGGGTCTACTATATCTGC
---------+---------+---------+---------+---------+---------+ 1080
ATAGGCTAGAAGCCGATGTAGTAGCTAACCTGTGCCACCCTCACCCAGATGATATAGACG

Y  P  I  F  G  Y  I  I  D  W  T  R  W  E  W  V  Y  Y  I  C   -
```

```
GGAATCGTGGGCACTCTCTGGTTCATCGCCTGGCAGTTCCTTGTGTTCGATAGTCCCGCT
---------+---------+---------+---------+---------+---------+ 1140
CCTTAGCACCCGTGAGAGACCAAGTAGCGGACCGTCAAGGAACACAAGCTATCAGGGCGA

G  I  V  G  T  L  W  F  I  A  W  Q  F  L  V  F  D  S  P  A  -

GAGCATCCCCGCATCGCTGACTCGGAGAGAAAGTTCATCGAGAAGTCCTTGGGTGCCTCC
---------+---------+---------+---------+---------+---------+ 1200
CTCGTAGGGGCGTAGCGACTGAGCCTCTCTTTCAAGTAGCTCTTCAGGAACCCACGGAGG

E  H  P  R  I  A  D  S  E  R  K  F  I  E  K  S  L  G  A  S  -

ATTCAGGGCAGCAAGGGACCCACACCCTGGAAGGCGATTGCCACCTCTCGTCCCGTTTGG
---------+---------+---------+---------+---------+---------+ 1260
TAAGTCCCGTCGTTCCCTGGGTGTGGGACCTTCCGCTAACGGTGGAGAGCAGGGCAAACC
                                                <-------------
I  Q  G  S  K  G  P  T  P  W  K  A  I  A  T  S  R  P  V  W  -

CTGAATGTGGTCGCCCAATGGGGTGGCATCTGGGGTCTCTTCACTTTGATGACCCATGCG
---------+---------+---------+---------+---------+---------+ 1320
GACTTACACCAGCGGGTTACCCCACCGTAGACCCCAGAGAAGTGAAACTACTGGGTACGC
---- Dpo1
L  N  V  V  A  Q  W  G  G  I  W  G  L  F  T  L  M  T  H  A  -

           Dpo5 ----------------->
CCGACTTACTTCAGACTGATTCACCACTGGAACATTCGAGCGACTGGCTTCCTGTCCGGA
---------+---------+---------+---------+---------+---------+ 1380
GGCTGAATGAAGTCTGACTAAGTGGTGACCTTGTAAGCTCGCTGACCGAAGGACAGGCCT

P  T  Y  F  R  L  I  H  H  W  N  I  R  A  T  G  F  L  S  G  -

CTGCCGCATCTTATGAGAATGCTCTTCGCCTACGTCTTCTCCATATTCGCCGACTATCTT
---------+---------+---------+---------+---------+---------+ 1440
GACGGCGTAGAATACTCTTACGAGAAGCGGATGCAGAAGAGGTATAAGCGGCTGATAGAA

L  P  H  L  M  R  M  L  F  A  Y  V  F  S  I  F  A  D  Y  L  -

CTGCGCACGGATAAGATGAGTCGCACAAATGTCCGCAAGTTGGCCACATTCATTTGCTGT
---------+---------+---------+---------+---------+---------+ 1500
GACGCGTGCCTATTCTACTCAGCGTGTTTACAGGCGTTCAACCGGTGTAAGTAAACGACA

L  R  T  D  K  M  S  R  T  N  V  R  K  L  A  T  F  I  C  C  -

GGCACTAAAGGCTTAATCGTATTAGCACTAGCTTACTTTGGCTACAACGCGACGGCTGCC
---------+---------+---------+---------+---------+---------+ 1560
CCGTGATTTCCGAATTAGCATAATCGTGATCGAATGAAACCGATGTTGCGCTGCCGACGG

G  T  K  G  L  I  V  L  A  L  A  Y  F  G  Y  N  A  T  A  A  -

ATCGTGTTGGTCACAGTGGCCACCATGCTTCACGGCGCCGTGTCCTCGGGTCCCTTGGCC
---------+---------+---------+---------+---------+---------+ 1620
TAGCACAACCAGTGTCACCGGTGGTACGAAGTGCCGCGGCACAGGAGCCCAGGGAACCGG

I  V  L  V  T  V  A  T  M  L  H  G  A  V  S  S  G  P  L  A  -
```

```
TCCATGGTCGATCTGTCGCCAAACTATGCGGGCATTGTCCTGGGCGTGAGTGGAATGATT
---------+---------+---------+---------+---------+---------+ 1680
AGGTACCAGCTAGACAGCGGTTTGATACGCCCGTAACAGGACCCGCACTCACCTTACTAA

S  M  V  D  L  S  P  N  Y  A  G  I  V  L  G  V  S  G  M  I  -

GGTGGAATGCCGGGCTTCATATCGCCCTTCATCGTGGGCCAACTTACTCACAATAATCAA
---------+---------+---------+---------+---------+---------+ 1740
CCACCTTACGGCCCGAAGTATAGCGGGAAGTAGCACCCGGTTGAATGAGTGTTATTAGTT

G  G  M  P  G  F  I  S  P  F  I  V  G  Q  L  T  H  N  N  Q  -

ACCATTGATGCCTGGAAGAATGTGTTCCTGCTCACCTCGTTGATGTTAACAGGCAGTGGC
---------+---------+---------+---------+---------+---------+ 1800
TGGTAACTACGGACCTTCTTACACAAGGACGAGTGGAGCAACTACAATTGTCCGTCACCG

T  I  D  A  W  K  N  V  F  L  L  T  S  L  M  L  T  G  S  G  -

ATCTTGTATGTGCTTTTCTCGGAATCCAAATTGCAGCCATGGAACAGTGGCTGTCACCAG
---------+---------+---------+---------+---------+---------+ 1860
TAGAACATACACGAAAAGAGCCTTAGGTTTAACGTCGGTACCTTGTCACCGACAGTGGTC

I  L  Y  V  L  F  S  E  S  K  L  Q  P  W  N  S  G  C  H  Q  -

TTGCCTGATTCCGGGCTAAAGGAACTTCAGAATCTGGGACGCGATCAGGACGATGAGGAG
---------+---------+---------+---------+---------+---------+ 1920
AACGGACTAAGGCCCGATTTCCTTGAAGTCTTAGACCCTGCGCTAGTCCTGCTACTCCTC

L  P  D  S  G  L  K  E  L  Q  N  L  G  R  D  Q  D  D  E  E  -


GAGAAGAAGCCTCTTAAATCTGACCATGATAAGGAAACCCCTATAGTCGCTGAACAGGAG
---------+---------+---------+---------+---------+---------+ 1980
CTCTTCTTCGGAGAATTTAGACTGGTACTATTCCTTTGGGGATATCAGCGACTTGTCCTC

E  K  K  P  L  K  S  D  H  D  K  E  T  P  I  V  A  E  Q  E  -

ACCAAAACAAAATCTGATTGCGATGGAAAGTGACATGTGTGGTCGCACTCTGTTTAATTG
---------+---------+---------+---------+---------+---------+ 2040
TGGTTTTGTTTTAGACTAACGCTACCTTTCACTGTACACACCAGCGTGAGACAAATTAAC

T  K  T  K  S  D  C  D  G  K  *                              -

TCATAATATCGTGTAGATTATGGCAACGAACAACGTCTGATAAGGCTTATCGTCATTTTG
---------+---------+---------+---------+---------+---------+ 2100
AGTATTATAGCACATCTAATACCGTTGCTTGTTGCAGACTATTCCGAATAGCAGTAAAAC

TGTAAATAACTACCCGTGAATCCACTGAGTTTAGCATCTTTAACCTCTAGATTTAAGTGA
---------+---------+---------+---------+---------+---------+ 2160
ACATTTATTGATGGGCACTTAGGTGACTCAAATCGTAGAAATTGGAGATCTAAATTCACT

GTTGCTAACTATACCATAGGATTAGTTCCCTGTTCTCCCCCAAGATTACCTTCTTATACA
---------+---------+---------+---------+---------+---------+ 2220
CAACGATTGATATGGTATCCTAATCAAGGGACAAGAGGGGGTTCTAATGGAAGAATATGT

TATAAATACAAATGTATATGATTTATTACCAGCTTTTGTTTATTATTTAAACAACACTTT
```

```
---------+---------+---------+---------+---------+---------+ 2280
ATATTTATGTTTACATATACTAAATAATGGTCGAAAACAAATAATAAATTTGTTGTGAAA

CGGCGTAAAAGTCACTTCATGTCGGCATGGCGTTAATCAAATAAAATGTTTTATGTTTAC
---------+---------+---------+---------+---------+---------+ 2340
GCCGCATTTTCAGTGAAGTACAGCCGTACCGCAATTAGTTTATTTTACAAAATACAAATG

GGAAGCTAAAAATGCTTTGCTGCTGTCTGGGAAAGTTTTTGCTTAAGCCAGGTGATTCAA
---------+---------+---------+---------+---------+---------+ 2400
CCTTCGATTTTTACGAAACGACGACAGACCCTTTCAAAAACGAATTCGGTCCACTAAGTT

CTAATCCCGACTAAGATAATTTTTTTCAGAAGGTCAAAAGGGATTTTTTAATATCTCAGC
---------+---------+---------+---------+---------+---------+ 2460
GATTAGGGCTGATTCTATTAAAAAAAGTCTTCCAGTTTTCCCTAAAAAATTATAGAGTCG

ATATTTATTTGCAGTTATTAAAAATAATGTTAATTAAATATTTGAGTGTGACGCCACGCG
---------+---------+---------+---------+---------+---------+ 2520
TATAAATAAACGTCAATAATTTTTATTACAATTAATTTATAAACTCACACTGCGGTGCGC

AGTACTTTTCTTTAACGTAAAACATTATAACTCGCTTTATTTTTAGACATTGCTATGTAT
---------+---------+---------+---------+---------+---------+ 2580
TCATGAAAAGAAATTGCATTTTGTAATATTGAGCGAAATAAAAATCTGTAACGATACATA

TCTTTAATTTTCGTATTAATACTACTCCCCTAAGAAAATATTAACATCCAAAATTAAGCT
---------+---------+---------+---------+---------+---------+ 2640
AGAAATTAAAAGCATAATTATGATGAGGGGATTCTTTTATAATTGTAGGTTTTAATTCGA

AATTAAAGAATTTAAAATTTTGAATTAGGAATTC
---------+---------+---------+---- 2674
TTAATTTCTTAAATTTTAAAACTTAATCCTTAAG
```

Bibliography

Ambrosio L, Mahowald A P, Perrimon N. (1989). Requirement of the *Drosophila raf* homolog for *torso* function. Nature. 342: 288-291

Ausubel F M, Brent R, Kingston R E, Moore D D, Seidman J G, Smith J A, Struhl K. (1989). Current Protocols in Molecular Biology.
John Wiley and Sons, New York

Bähler M, Kroschewski R, Stoffler H-E, Behrmann T. (1994). Rat myr4 defines a novel subclass of myosin I: identification, distribution, localization and mapping of calmodulin binding sites with differential calcium sensitivity. Journal of Cell Biology 126: 375-389

Bakken A H. (1973). A cytological and genetic study of oogenesis in *Drosophila melanogaster*. Developmental Biology 33: 100-122

Bardsley A, McDonald K, Boswell R E. (1993). Distribution of *tudor* protein in the *Drosophila* embryo suggests separation of functions based on site of localization. Development 119: 207-219

Barnett T, Pachl C, Gergen J P, Wensink P C. (1980). The isolation and characterization of *Drosophila* yolk protein genes. Cell 21: 729-738

Belote J M, Baker B S. (1983). The dual functions of a sex determination gene in *Drosophila melanogaster*. Developmental Biology 95: 512-517

Belvin M, Jin Y, Anderson K V. (1995). *Cactus* protein degradation mediates *Drosophila* dorsal-ventral signaling. Genes and Development 9: 783-793

Bement W M, Hasson T, Wirth J A, Cheney R E, Mooseker M S. (1994). Identification and overlapping expression of multiple unconventional myosin genes in vertebrate cell-types. Proceedings of the National Academy of Sciences of the USA. 91: 6549-6553.

Bement W M, Mooseker M S. (1995) TEDS Rule: a molecular rationale for differential regulation of myosins by phosphorylation of the heavy chain head. Cell Motility and the Cytoskeleton. 31: 87-92

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

Bier E, Vaessin H, Shephard S, Lee K, McCall K, Barbel S, Ackerman L, Carretto R, Uemura T, Grell E, Jan L Y, Jan Y N. (1989). Searching for pattern and mutation in the *Drosophila* genome with a P-*lacZ* vector. Genes and Development 3: 1223-1287

Birnboim H C, Doly J. (1979). A rapid alkaline lysis extraction procedure for screening recombinant DNA. Nucleic Acids Research 7: 1513-1523

Bopp D, Horabin J I, Lersch R A, Cline, T W, Schedl P. (1993). Expression of the *Sex-lethal* gene is controlled at multiple levels during *Drosophila* oogenesis. Development 118: 797-812.

Bownes M, Dale L. (1982). in *A Handbook of* Drosophila *Development.* ed R Ransom Elsevier Biomedical Press, Amsterdam pp31-68

Bownes M. (1994). The regulation of the yolk protein genes, a family of sex differentiation genes in *Drosophila melanogaster.* BioEssays 16: 745752

Bownes M, Ronaldson E, Mauchline D. (1996). 20-Hydroxyecdysone, but not Juvenile Hormone, regulation of *yolk protein* gene expression can be mapped to *cis*-acting DNA sequences. Developmental Biology. 173: 475-489

Brand A H, Perrimon N. (1994). Raf acts downstream of the EGF receptor to determine dorsoventral polarity during *Drosophila* oogenesis. Genes and Development 8: 629-639

Bridges C B. (1916) Non-disjunction as proof of the chromosome theory of heredity. Genetics 1: 1-52

Brockerhoff S E, Stevens R C, Davis T N. (1994). The unconventional myosin, myo2p, is a calmodulin target at sites of cell-growth in *Saccharomyces cerevisiae.* Journal of Cell Biology 124: 315-323.

Brown N H, Kafatos F C. (1988). Functional cDNA libraries from *Drosophila melanogaster.* Journal of Molecular Biology 203: 425-437

Bu D-F, Erlander M G, Hitz B C, Tillakaratne N J K, Kaufman D L, Wagner-McPherson C B, Evans G A, Tobin A J. (1992). Two human glutamate decarboxylases, 65-kDa GAD and 67-kDa GAD, are each encoded by a single gene. Proceedings of the National Academy of Sciences of the USA 89: 2115-2119

Cant K, Knowles B A, Mooseker M S, Cooley L. (1994). *Drosophila singed*, a Fascin homolog, is required for actin bundle formation during oogenesis and bristle extension. Journal of Cell Biology 125: 369-380.

Cassanova J, Furriols M, McCormack C A, Struhl G. (1995). Similarities between *trunk* and *spätzle*, putative extracellular ligands specifying body pattern in *Drosophila.* Genes and Develoment 9: 2539-2544

Cavener D R. (1987). Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucleic Acids Research 15: 1353-1361

Chasan R, Jin Y, Anderson K V. (1992). Activation of the *easter* zymogen is regulated by five other genes to define dorsal-ventral polarity in the *Drosophila* embryo. Development 115: 607-616

Chen I L, Edwards K, Lin R C, Coats L W, Kiehart D P. (1991). *Drosophila* myosin heavy chain at 35B,C. Journal of Cell Biology 115: 330a (Abstract)

Cheney R E, Mooseker M S. (1992). Unconventional Myosins. Current Opinion in Cell Biology 4: 27-35

Cheney R E, Riley MA, Mooseker M S. (1993a). Phylogenetic analysis of the myosin superfamily. Cell Motility and the Cytoskeleton 24: 215-223

Cheney R E, Oshea M K, Heuser J E, Coelho M V, Wolenski J S, Espreafico E M, Forscher P, Larson R E, Mooseker M S. (1993b). Brain myosin-V is a 2-headed unconventional myosin with motor- activity. Cell 75: 13-23.

Cheung H-K, Serano T L, Cohen R S. (1992). Evidence for a highly selective RNA transport system and its role in establishing the dorsal-ventral axis of the *Drosophila* egg. Development. 114: 653-661

Cheyette B N R, Green P J, Martin K, Garren H, Hartenstein V, Zipursky S L. 1994). The *Drosophila sine oculis* locus encodes a homeodomain-containing protein required for the development of the entire visual system. Neuron 12: 977--996

Chong S S, Kozak C A, Liu L, Bordeau J E, Hughes M R, Kristjansson K, Dunn S.T. (1995). Cloning, genetic mapping and expression analysis of a mouse renal sodium-dependent phosphate cotransporter. American Journal of Physiology 268: 1038-1045

Chou T-B, Perrimon N. (1992). Use of a yeast site-specific recombinase to produce female germline chimeras in *Drosophila*. Genetics 131:643-653

Chou T-B, Noll E, Perrimon N. (1993). Autosomal P[ovo^{D1}] dominant female-sterile insertions in *Drosophila* and their use in generating germ-line chimeras. Development 119:1359-1369

Chomczynski P, Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. Analytical Biochemistry 162: 156-159

Christerson L B, McKearin D M. (1994). *orb* is required for anteroposterior and dorsoventral patterning during *Drosophila* oogenesis. Genes and Development 8: 614-628.

Clark K A, McKearin D M. (1996). The *Drosophila stonewall* gene encodes a putative transcription factor essential for germ-cell development. Development 122: 937-950.

Clark L, Giniger E, Ruohola-Baker H, Jan L Y, Jan Y N. (1994). Transient posterior localisation of a kinesin fusion protein reflects anterioposterior polarity of the *Drosophila* oocyte. Current Biology 4: 289-300

Clark-Maguire S, Mains P E. (1994) *mei-1*, a gene required for meiotic spindle formation in *Caenorhabditis elegans*, is a member of a family of ATPases. Genetics 136: 533-546

Cline T W. (1976). A sex-specific, temperature-sensitive maternal effect of the *daughterless* mutation in *Drosophila melanogaster*. Genetics 84: 723-742.

Cline T W. (1983). The interaction between *daughterless* and *Sex-lethal* in triploids: A lethal sex-transforming maternal effect linking sex determination and dosage compensation in *Drosophila melanogaster*. Developmental Biology 72: 266-275.

Cline T W. (1988). Evidence that *Sisterless-a* and *Sisterless-B* are 2 of several discrete numerator elements of the X/A sex determination signal in *Drosophila* that switch *Sxl* between 2 alternative stable expression states. Genetics 119: 829-862.

Cline T W. (1993). The *Drosophila* sex determination signal - How do flies count to 2. Trends in Genetics 9: 385-390.

Cohen B, McGuffin M E, Pfeifle C, Segal D, Cohen S M. (1992). *apterous*, a gene required for imaginal disc development in *Drosophila* encodes a member of the LIM family of developmental regulatory proteins. Genes and Development 6: 715-729

Cooley L, Verheyen E, Ayers K. (1992). *chickadee* encodes a Profilin required for intracellular cytoplasm transport during *Drosophila* oogenesis. Cell 69: 173-184.

Cooley L, Theurkauf W E. (1994). Cytoskeletal functions during *Drosophila* oogenesis. Science 266: 590-596.

Cronmiller C, Schedl P, Cline T W. (1988). Molecular characterization of *Daughterless*, a *Drosophila* sex determination gene with multiple roles in development. Genes and Development 2: 1666-1676.

Cummings C A, Cronmiller C. (1994). The *Daughterless* gene functions together with *Notch* and *Delta* in the control of ovarian follicle development in *Drosophila*. Development 120: 381-394.

Delidakis C, Kafatos F C. (1987). Amplification of a chorion gene cluster in *Drosophila* is subject to multiple *cis*-regulatory elements and to long range position effects. Journal of Molecular Biology 197: 11-26

Devereux J, Haeberli P, Smithies O. (1984). A comprehensive set of sequence analysis programs for VAX. Nucleic Acids Research 12: 387-395

Ding D, Parkhurst S M, Lipshitz H D. (1993a). Different genetic requirements for anterior RNA localization revealed by the distribution of *adducin-like* transcripts during *Drosophila* oogenesis. Proceedings of the National Academy of Sciences of the USA 90: 2512-2516

Ding D, Parkhurst S M, Halsell S R, Lipshitz H D. (1993b). Dynamic *Hsp83* RNA localization during *Drosophila* oogenesis and embryogenesis. Molecular and Cellular Biology 13: 3773-3781

Doyle H J, Bishop J M. (1993) Torso, a receptor tyrosine kinase required for embryonic pattern formation, shares substrates with sevenless and EGF-R pathways in *Drosophila*. Genes and Development 7: 633-646

Driever W, Nüsslein-Volhard C. (1989). Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the *bicoid* morphogen. Nature 340: 363-367

Dubnau J, Struhl G. (1996). RNA recogntion and translational regulation by a homeodomain protein. Nature 379: 694-699

Duffy J B, Gergen J P. (1991). The *Drosophila* segmentation gene *runt* acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene *Sex-lethal*. Genes and Development. 5: 2176-2187

Edwards K A, Kiehart D P. (1996). *Drosophila* nonmuscle myosin II has multiple essential roles in imaginal disc and egg chamber morphogenesis. Development 122: 1499-1511

Ephrussi A, Dickinson L K, Lehmann R. (1991). *oskar* organises the germ plasm and directs localization of the posterior determinant *nanos*. Cell 66: 37-60

Ephrussi A, Lehmann R. (1992). Induction of germ cell formation by *oskar*. Nature 358: 387-392

Erdélyi M, Michon A-M, Guichet A, Bogucka Glotzer J, Ephrussi A. (1995). Requirement for *Drosophila* cytoplasmic tropomyosin in *oskar* mRNA localization. Nature 377: 524-527

Erickson J W, and Cline T W. (1993). A BZIP protein, Sisterless-a, collaborates with BHLH transcription factors early in *Drosophila* development to determine sex. Genes and Development 7: 1688-1702.

Erlander M G, Tobin A J. (1991). The structural and functional heterogeneity of glutamic acid decarboxylase: A review. Neurochemical Research 16: 215-226

Espreafico E M, Cheney R E, Matteali M, Nascimento M C, De Camilli P V, Larsen R E, Mooseker M S. (1992). Primary structure and cellular localization of chicken brain myosin-V, an unconvential myosin with calmodulin light chains. Journal of Cell Biology 119: 1541-1557

Espindola F S, Espreafico E M, Coelho M V, Martins A R, Costa F R C, Mooseker M S, Larson R E. (1992). Biochemical and immunological characterization of p190-calmodulin complex from vertebrate brain: a novel calmodulin binding protein. Journal of Cell Biology 118: 359-368

Fargnoli J, Waring G L. (1982). Identification of vitelline membrane proteins in *Drosophila melanogaster*. Developmental Biology 92: 306-314

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

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

Flickinger T W, Salz H K. (1994). The *Drosophila* sex determination gene *snf* encodes a nuclear protein with sequence and functional similarity to the mammalian U1A snRNP protein. Genes and Development. 8: 914-925

Flybase (1993). A *Drosophila* genetic database. http://flybase.bio.indiana.edu:82/ or FTP.BIO.INDIANA.EDU

Freeman M. (1991). First, trap your enhancer. Current Biology 1: 378-381

Frohnhöfer H G, Lehmann R, Nüsslein-Volhard C. (1986). Manipulating the anterioposterior pattern of the *Drosophila* embryo. Journal of Embryology and Experimental Morphology. 97: 169-179

Gait M J. (1984). An Introduction to Modern Methods of DNA Synthesis, in: *Oligonucleotide Synthesis: A Practical Approach*, ed Gait M J. IRL Press, Oxford pp 1-22

Gavis E R, Lehmann R. (1992). Localization of *nanos* RNA controls embryonic polarity. Cell 71:301-313

Gavis E R, Lehmann R. (1994). Translational regulation of *nanos* by RNA localization. Nature 369: 315-318

Ghosh S, Clifford A M, Riviera L R, Tempst P, Nolan G P, Baltimore D. (1990). Cloning of the p50 DNA binding subunit of NF-kB: homology to *rel* and *dorsal*. Cell 62: 1019-1029

Gibson F, Walsh J, Mburu P, Varela A, Brown K A, Antonio M, Beisel, K W, Steel K P, Brown S D M. (1995). A Type-VII myosin encoded by the mouse deafness gene *shaker-1*. Nature 374: 62-64.

Golumbeski G S, Bardsley A, Tax F, Boswell R E. (1991). *Tudor*, a posterior group gene of *Drosophila melanogaster* encodes a novel protein and a messenger RNA localized during mid-oogenesis. Genes and Development 5: 2060-2070

González-Reyes A, St Johnston D. (1994). Role of oocyte position in establishment of anterior-posterior polarity in *Drosophila*. Science 266: 639-642.

González-Reyes A, Elliott H, St Johnston D. (1995). Polarization of both major body axes in *Drosophila* by *Gurken-Torpedo* signalling. Nature 375: 654-658.

Goode S, Wright D, Mahowald A P. (1992). The neurogenic locus *brainiac* cooperates with the *Drosophila* EGF receptor to establish the ovarian follicle and to determine its dorsal-ventral polarity. Development 116: 177-192

Govind S, Brennan L, Steward R. (1993). Homeostatic balance between Dorsal and Cactus proteins in the *Drosophila* embryo. Development 117: 135-148

Granadino B,.Juán A S, Sanchez L. (1992)..Evidence of a dual function in *fl(2)d*, a gene needed for *Sex-lethal* expression in *Drosophila melanogaster*. Genetics 130: 597-612

Grossniklaus U, Bellen H J, Wilson C, Gehring W J. (1989). P-Element-mediated enhancer detection applied to the study of oogenesis in *Drosophila*. Development 107: 189-200

Haarer B K, Petzold A, Lillie S H, Brown S S. (1994). Identification of *MYO4*, a second class V myosin gene in yeast. Journal of Cell Science 107: 1055-1064

Harlow E, Lane D. (1988). Antibodies: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Hartl D L, Nurminsky D I, jones R W, Lozovskaya E R. (1994). Genome structure and evolution in *Drosophila*: applications of the framework P1 map. Proceedings of the National Academy of Sciences of the USA 91: 6824-6829

Hashimoto C, Hudson K L, Anderson K V. (1988). The *Toll* gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell 52:269-279

Hashimoto C, Gerttula S, Anderson K V. (1991). Plasma membrane localization of the *Toll* protein in the syncytial *Drosophila* embryo: importance of transmembrane signaling for dorsal-ventral pattern formation. Development 111: 1021-1028

Hasson T, and Mooseker M S. (1994). Porcine myosin-VI - Characterization of a new mammalian unconventional myosin. Journal of Cell Biology. 127: 425-440.

Hawkins N C, Thorpe J, Schüpbach T. (1996). *Encore*, a gene required for the regulation of germ-line mitosis and oocyte differentiation during *Drosophila* oogenesis. Development 122: 281-290.

Hilfiker A, Amrein H, Dubendorfer A, Schneiter R, Nothiger R. (1995). The gene *Virilizer* is required for female-specific splicing controlled by *Sxl*, the master gene for sexual development in *Drosophila*. Development. 121: 4017-4026.

Hedgé J, Stephenson E C. (1993). Distribution of *swallow* protein in egg chambers and embryos of *Drosophila melanogaster*. Development 119: 457-470

Heitzler P, Coulson D, Saenz-Robles M-T, Ashburner M, Roote J, Simpson P, Gubb D. (1993). Genetic and cytogenetic analysis of the 43A-E region containing the segment polarity gene *costa* and the cellular polarity genes *prickle* and *spiny-legs* in *Drosophila melanogaster*. Genetics 135:105-115

Hong C C, Hashimoto C.(1995). An unusual mosaic protein with a protease domain, encoded by the *nudel* gene, is involved in defining embryonic dorsoventral polarity in *Drosophila*. Cell 82: 785-794

Horowitz J A, Hammer J A. (1990). A new *Acanthamoeba* myosin heavy chain. Journal of Biological Chemistry. 265: 20646-20652

Horowitz H, Berg C A. (1996). The *Drosophila pipsqueak* gene encodes a nuclear BTB-domain-containing protein required early in oogenesis. Development 122: 1859-1871

Hsu T, Bagni C, Sutherland J D, Kafatos F C. (1996). The transcription factor CF2 is a mediator of EGF-R-activated dorsoventral patterning in *Drosophila* oogenesis. Genes and Development 10: 1411-1421

Huang W-M, Fourquet L R, Wu E, Wu J-Y. (1990). Molecular cloning and amino acid sequence of brain L-glutamate decarboxylase. Proceedings of the National Academy of Sciences of the USA 87: 8491-8495

Hülskamp M, Schroeder C, Pfeifle V, Jäckle H, Tautz D. (1989). Posterior segmentation of the *Drosophila* embryo in the absence of a maternal posterior organizer gene. Nature 338: 629-632

Irish V, Lehmann R, Akam M. (1989). The *Drosophila* posterior group gene *nanos* functions by repressing *hunchback* activity. Nature 338: 646-648

Jackson F R, Newby L M, Kulkarni S J. (1990). *Drosophila* GABAergic systems: Sequence and expression of glutamic acid decarboxylase. Journal of Neurochemistry 54: 1068-1078

Jansen R-P, Dowzer C, Michaelis C, Galova M, Nasmyth K. (1995). Mother cell-specific *HO* expression in budding yeast depends on the unconventional myosin myo4p and other cytoplasmic proteins. Cell 84: 687-697.

Jingman J, Petri W H. (1993). Developmental control elements in the promoter of a *Drosophila* vitelline membrane gene. Developmental Biology 156: 557-565

Johnston G C, Pendergast J A, Singer R A. (1991). The *Saccharomyces cerevisiae MYO2* gene encodes an essential myosin for vectorial transport of vesicles. Journal of Cell Biology 113: 539-551

Jongens T A, Hay B, Jan L Y, Jan Y N. (1992). The *germ cell-less* gene-product - a posteriorly localized component necessary for germ-cell development in *Drosophila.* Cell 70: 569-584

Jongens T A, Ackerman L D, Swedlow J R, Jan L Y, Jan Y N. (1994). *germ cell-less* encodes a cell-type-specific nuclear pore-associated protein and functions early in the germ-cell specification pathway of *Drosophila.* Genes and Development 8:.2123-2136

Jung G, Hammer J A. (1994) The actin binding site in the tail domain of *Dictyostelium* myosin IC (myoC) resides within the glycine- and proline-rich sequence (tail homology region 2). FEBS Letters 342:197-202

Kania M A, Bonner A S, Duffy J B, Gergen J P. (1990) The *Drosophila* segementation gene runt encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system. Genes and Development 4: 1701-1703

Kellerman K A, Miller K G. (1992). An unconventional myosin heavy-chain gene from *Drosophila melanogaster.* Journal of Cell Biology 119: 823-834.

Kelley R C. (1993). Initial organization of the *Drosophila* dorsoventral axis depends on an RNA-binding protein encoded by the *squid* gene. Genes and Development 7: 948-960

Kim-Ha J, Smith J L, Macdonald P M. (1991). *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. Cell 66: 23-35

Kim-Ha J, Webster P J, Smith J L, Macdonald P M. (1993). Multiple RNA regulatory elements mediate distinct steps in localization of *oskar* mRNA. Development 119: 169-178

Kim-Ha J, Kerr K, Macdonald P M. (1995) Translational regulation of *oskar* messenger-RNA by Bruno, an ovarian RNA-binding protein, is essential. Cell 81: 403-412

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

Kinkema M, Wang, H Y, Schiefelbein J. (1994). Molecular analysis of the myosin gene family in *Arabidopsis thaliana.* Plant Molecular Biology 26: 1139-1153.

Kinkema M, and Schiefelbein J. (1994). A myosin from a higher-plant has structural similarities to class-V myosins. Journal of Molecular Biology 239: 591-597.

Knight A E, Kendrick-Jones J. (1993) A myosin-like protein from a higher plant. Journal of Molecular Biology 231: 148-154

Knight A E. (1994). The diversity of myosin-like proteins. PhD Thesis. University of Cambridge

Kobayashi S, Yamada M, Asoaka M, Kitamura T. (1996). Essential role of the posterior morphogen *nanos* for germline development in *Drosophila*. Nature 380: 708-711

Kuznetsov S A, Langford G M, Weiss D G. (1992). Actin-dependent organelle movement in squid axoplasm. Nature 356: 722-725

Lamarche N, Hall A. (1994). GAPs for rho-related GTPases. Trends in Genetics 10: 436-440

Langford G M. (1995). Actin- and microtubule-dependent organelle motors: interrelationships between the two motility systems. Current Opinion in Cell Biology 7: 82-88

Lantz V, Ambrosio L, Schedl P. (1992). The *Drosophila orb* gene is predicted to encode sex-specific germline RNA-binding proteins and has localized transcripts in ovaries and early embryos. Development 115: 75-88

Lantz V, Chang J S, Horabin J I, Bopp D, Schedl P. (1994). The *Drosophila orb* RNA-binding protein is required for the formation of the egg chamber and establishment of polarity. Genes and Development 8: 598-613

Lasko P. (1994). Molecular genetics of oogenesis. R G Landes and Co, Austin

Lathe R. (1985). Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. Journal of Molecular Biology 183: 1-12

Lederberg E M, Cohen S N. (1974). Transformation of *Salmonella typhimurium* by plasmid deoxyribonucleic acid. Journal of Bacteriology 119: 1072-1074

Letsou A, Alexander S, Orth K, Wasserman S A. (1991) Genetic and molecular characterization of *tube*, a *Drosophila* gene maternally required for embryonic dorsoventral polarity. Proceedings of the National Academy of Sciences of the USA 88: 810-814

Liang L, Diehl-Jones W, Lasko P. (1994). Localization of *vasa* protein in the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. Development 120: 1201-1211

Lin H, Yue L, Spradling A. (1994). The *Drosophila* fusome, a germ-line specific organelle contains membrane skeletal proteins and functions in cyst formation. Development 120: 947-956

Lindsley D L, Zimm G G. (1992). The Genome of *Drosophila melanogaster*. Academic Press Inc, San Diego

Lupas A, van Dyke M, Stock J. (1991). Predicting coiled coils from protein sequences. Science 252: 1162-1164

Luria S E, Burrous J N. (1957). Hybridization between *Escherichia coli* and *Shigella.* Journal of Bacteriology 74: 461-476

Macdonald P M, Kerr K, Smith J L, Leask A. (1993). RNA regulatory element BLE1 directs the early steps of *bicoid* mRNA localization. Development 118: 1233-1243

MacDougall C, Harbison D, Bownes M. (1995). The developmental consequences of alternate splicing in sex determination and differentiation in *Drosophila.* Developmental Biology 172: 353-376.

Mahajan-Miklos S, Cooley L. (1994). The Villin-like protein encoded by the *Drosophila quail* gene is required for actin bundle assembly during oogenesis. Cell 78: 291-301

Mahowald A P. (1968). Polar granules of *Drosophila* II. Ultrstructural changes during early embryogenesis. Journal of Experimental Zoology 167: 237-262

Mahowald A P, Kambysellis M P. (1980). Oogenesis in *The Genetics and Biology of* Drosophila. ed Asburner M, Wright T R F. Vol 2 Academic Press, New York. pp 141-224

Markussen F H, Michon A M, Breitwieser W, Ephrussi A. (1995). Translational control of *oskar* generates short Osk, the isoform that induces pole plasm assembly. Development 121: 3723-3732

Marsh J L, Wieschaus E. (1978). Is sex determination in germ line and soma controlled by separate genetic mechanisms? Nature. 272. 249-251

McKearin D M, Spradling A C. (1990). *bag-of-marbles* - a *Drosophila* gene required to initiate both male and female gametogenesis. Genes and Development 4: 2242-2251.

Ménel-Ninio M, Terracol R, Kafatos F.C. (1991). The *ovo* gene of *Drosophila* encodes a zinc finger protein required for female germ line development. EMBO Journal 10: 2259-2266.

Mercer J A, Seperack P K, Strobel M C, Copeland N G, Jenkins N A. (1991). Novel myosin heavy shains encoded by murine *dilute* coat colour locus. Nature 349: 709-713

Mermall V, McNally J G, Miller K G. (1994). Transport of cytoplasmic particles catalyzed by an unconventional myosin in living *Drosophila* embryos. Nature 369: 560-562.

Mermall V, Miller K G. (1995). The 95F unconventional myosin Is required for proper organization of the *Drosophila* syncytial blastoderm. Journal of Cell Biology 129: 1575-1588.

Miyamoto K I, Tatsumi S, Sonoda T, Yamamoto H, Minami H, Taketani Y, Takeda E. (1995). Cloning and functional expression of a Na^+-dependent phosphate co-transporter from human kidney: cDNA cloning and functional expression. Biochemistry Journal 305: 81-85

Montell C, Rubin G M. (1988). The *Drosophila ninaC* locus encodes two photoreceptor cell specific proteins with domains homologous to protein kinases and the myosin heavy chain head. Cell 52: 757-772

Mooseker M S, Cheney R E. (1995). Unconventional myosins. Annual Review of Cell and Developmental Biology 11: 633-675

Morgan, N S, Skovronsky D M, Artavanistsakonas S, Mooseker M S. (1994). The molecular-cloning and characterization of *Drosophila-melanogaster* Myosin-Ia and Myosin-Ib. Journal of Molecular Biology 239: 347-356.

Morgan N S, Heintzelman M B, Mooseker M S. (1995). Characterization of Myosin-1a and Myosin-1b, two unconventional myosins associated with the *Drosophila* brush border cytoskeleton. Developmental Biology 172: 51-71

Morisato D, Anderson K. (1993). A proteolytically processed form of protein is required for the activation of *Toll*. 34th Annual *Drosophila* Research Conference, San Diego pp249

Morisato D, Anderson K V. (1994). The *spätzle* gene encodes a component of the extracellular signaling pathway establishing the dorsal-ventral pattern of the *Drosophila* embryo. Cell 76: 677-688

Morisato D, Anderson K V. (1995) Signaling pathways that establish the dorsal-ventral pattern of the *Drosophila* embryo. Annual Review of Genetics 29: 371-399

Murata Y, Wharton R P. (1995). Binding of Pumilio to maternal *hunchback* mRNA is required for posterior patterning in *Drosophila* embryos. Cell 80: 747-756

Murre C, McCaw P S, Vaessin H, Caudy M, Jan L Y, Jan Y N, Carrera C V, Buskin J N, Hauschka S D, Lassar A B, Weintraub H, Baltimore D. (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell. 58: 537-544

Newmark P A, Boswell R E. (1994). The *mago nashi* locus encodes an essential product required for germ plasm assembly in *Drosophila* . Development 120: 1301-1313

Ni B, Rosteck P R, Nadi S, Paul S M. (1994). Cloning and expression of a cDNA encoding a brain-specific Na^+-dependent inorganic phosphate cotransporter. Proceedings of the National Academy of Sciences of the USA 91: 5607-5611

Nolan G P, Ghosh S, Liou H C, Tempst P, Baltimore D. (1991). DNA binding and 1kB inhibition of the cloned p66 subunit of NF-kB, a *rel*-related polypeptide. Cell 64: 961-969

Nöthiger R, Jonglez M, Leuthold M, Meier-Gerschwiler P, Weber T. (1989). Sex determination in the germ line of *Drosophila* depends on genetic signals and inductive somatic factors. Development 107: 505-518

O'Kane C J, Gehring W J. (1987). Dectection *in situ* of genomic regulatory elements in *Drosophila*. Proceedings of the National Academy of Sciences of the USA 84: 9123-9127

Oliver B, Perrimon N, Mahowald A P. (1987). The *ovo* locus is required for sex-specific germ line maintenance in *Drosophila*. Genes and Development. 1: 913-923.

Oliver B, Pauli D, Mahowald A P. (1990). Genetic evidence that the *ovo* locus is involved in *Drosophila* germ line sex determination. Genetics 125: 535-550

Oliver B, Kim Y J, Baker B S. (1993). *Sex-Lethal*, master and slave - a hierarchy of germ-line sex determination in *Drosophila*. Development 119: 897-908.

Oliver B, Singer J, Laget V, Pennetta G, Pauli D. (1994). Function of *Drosophila* ovo^{+} in germ-line sex determination depends on X-chromosome number. Development 120: 3185-3195.

Oliver T N, Corey D P, Derfler B H, Pennisi C M, Cheney R E. (1996). Myosin-X: An unconventional myosin with PH domains. Molecular Biology of the Cell 7: 227-227.

Parks S, Spradling A. (1987). Spatially regulated expression of chorion genes during *Drosophila* oogenesis. Genes and Development 1: 497-509

Paroush Z, Finley R L, Kidd T, Wainwright S M, Ingham P W, Brent R, and Ish-Horowicz D. (1994). Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with Hairy-Related Bhlh proteins. Cell 79: 805-815.

Perkins L A, Larsen I, Perrimon N. (1992). *corkscrew* encodes a putative protein tyrosine phosphatase that functions to transduce the terminal signal from the receptor tyrosine kinase Torso. Cell 70: 225-236

Perrimon N, Mohler D, Engstrom L, Mahowald A P. (1986). X-linked female-sterile loci in *Drosophila melanogaster*. Genetics 113: 695-712

Perrimon N. (1993). The Torso receptor protein-tyrosine kinase signalling pathway: An endless story. Cell 74: 219-222

Peterson M D, Novak K D, Reedy M C, Ruman J L, Titus M A. (1995). Molecular genetic analysis of myoC, a *Dictyostelium* myosin I. Journal of Cell Science 108: 1093-1103

Pollard T D, Korn E D. (1973) *Acanthamoeba* Myosin I. Isolation from *Acanthamoeba castellanii* of an enzyme similar to muscle myosin. Journal of Biological Chemistry 248: 4682-4690

Pollard T D, Doberstein S K, Zot H G. (1991). Myosin-I. Annual Review of Physiology 53: 653-681

Porter J A, Hicks J L, Williams D S, Montell C. (1992). Differential localizations and requirements for the two *Drosophila ninaC* kinase/myosins in photoreceptor cells. Journal of Cell Biology 116: 683-693

Proudfoot N. (1991). Poly(A) signals. Cell 64: 671-674

Pultz M A, Carson G S, Baker B S. (1994). A genetic-analysis of *hermaphrodite*, a pleiotropic sex determination gene in *Drosophila melanogaster*. Genetics 136: 195-207.

Ran B, Bopp R, and Suter B. (1994). Null alleles reveal novel requirements for *Bic-D* during *Drosophila* oogenesis and zygotic development. Development 120: 1233-1242.

Reinhard J, Scheel A A, Dickman D, Hall A, Ruppert C, Bähler M. (1995). A novel type of myosin implicated in signalling by rho family GTPases. EMBO Journal 14: 697-704

Rivera-Pomar R, Lu X, Perrimon N, Taubert H, Jäckle H. (1995). Activation of posterior gap gene expression in the *Drosophila* blastoderm. Nature 376: 253-256

Rivera-Pomar R, Niessing D, Schmidt-Ott U, Gehring W J, Jäckle H. (1996). RNA binding and translational suppression by *bicoid*. Nature 379: 746-749

Robinson D N, Cant K, Cooley L. (1994). Morphogenesis of *Drosophila* ovarian ring canals. Development 120: 2015-2025

Rodesch C, Geyer P K, Patton J S, Bae E, Nagoshi R N. (1995). Developmental analysis of the *ovarian tumor* gene during *Drosophila* oogenesis. Genetics 141: 191-202

Rogers S, Wells R, Rechsteiner M. (1986). Amino-acid sequences common to rapidly degraded proteins - the PEST hypothesis. Science 234: 364-368

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

Rongo C, Gavis E R, Lehmann R. (1995). Localization of *oskar* RNA regulates *oskar* translation end requires *oskar* protein. Development 121: 2737-2746

Roth S, Hiromi Y, Godt D, Nüsslein-Volhard C. (1991). *cactus*, a maternal gene required for proper formation of the dorsoventral morphogen gradient in *Drosophila* embryos. Development 112: 371-388

Roth S. (1994). Proteolytic generation of a morphogen. Current Biology 4: 755-757

Roth S, Nueman-Silberberg F S, Barcelo G, Schüpbach T. (1995). *cornichon* and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in *Drosophila*. Cell 81: 967-978

Ruohola -Baker H, Grell E, Chou T-B, Baker D, Jan L Y, Jan Y N. (1993). Spatially localised Rhomboid is required for the establishment of the dorsal-ventral axis in *Drosophila* oogenesis. Cell 73: 953-965

Ruppert C, Kroschewski R, Bähler M. (1993). Identification, characterization and cloning of myr 1, a mammalian myosin-I. Journal of Cell Biology 120: 1393-1403

Ryner L C, Swain A. (1995). Sex in the 90's. Cell 81: 483-493

Salz H K, Cline T W, Schedl P. (1987). Functional changes associated with structural alterations induced by mobilization of a P element inserted in the *Sex-lethal* gene of *Drosophila*. Genetics 117: 221-231

Sambrook J, Frisch E F, Maniatis T. (1989). Molecular Cloning. A Laboratory Manual, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Sander K, Lehmann R. (1988). *Drosophila* nurse cells produce a posterior signal required for embryonic segmentation and polarity. Nature 335: 68-70

Sanders G, Lichte B, Meyer H E, Kilimann M W. (1992) cDNA encoding the chicken ortholog of the mouse *dilute* gene product. sequence comparison reveals a myosin I subfamily with conserved C-terminal domains. FEBS Letters 311: 295-298

Sanger F, Nicklen S, Coulson A R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the USA 74: 5463-5467

Sass G L, Comer A R, Searles L L. (1995). The *ovarian tumor* protein isoforms of *Drosophila-melanogaster* exhibit differences in function, expression, and localization. Developmental Biology 167: 201-212.

Savant-Bhonsale S, Montell D J. (1993). *torso-like* encodes the localized determinant of *Drosophila* terminal pattern formation. Genes and Development 7: 2548-2555

Schneider D S, Hudson K L, Lin T Y, Anderson K V. (1991). Dominant and recessive mutations define functional domains of *Toll*, a transmembrane protein required for dorsal-ventral polarity in the *Drosophila* embryo. Development 120: 1243-1250

Schneider D S, Jin Y, Morisato D, Anderson K V. (1994). A processed form of the Spätzle protein defines dorsal-ventral polarity in the *Drosophila* embryo. Development 120: 1243-1250

Schüpbach T. (1982). Autosomal mutations that interfere with sex determination in somatic cells of *Drosophila* have no direct effect on the germline. Developmental Biology 89: 117-127

Schüpbach T, Wieschaus E. (1989). Female sterile mutations on the 2nd chromosome of *Drosophila melanogaster* I. Maternal effect mutations. Genetics 121: 101-117

Schüpbach T, Wieschaus E. (1991). Female sterile mutations on the 2nd chromosome of *Drosophila melanogaster* .II. Mutations blocking oogenesis or altering egg morphology. Genetics 129: 1119-1136.

Serano T L, Cohen R S. (1995). Gratuitous mRNA localization in the *Drosophila* oocyte. Development 121: 3013-3021

Shelton C A, Wasserman S A. (1993). *pelle* encodes a protein kinase required to establish dorsoventral polarity in the *Drosophila* embryo. Cell 72: 515-525

Shirras A D, Bownes M. (1989). *cricklet*, a locus regulating a number of adult functions of *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the USA 86: 4559-4563

Siegel V, Jongens T A, Jan L Y, Jan Y N. (1993). *pipsqueak*, an early acting member of the posterior group of genes, affects *vasa* level and germ cell-somatic cell interaction in the developing egg chamber. Development 119: 1187-1202

Sisson J C, Scott M P. (1996). Characterization of the *costal2* segment polarity gene product. American Conference of *Drosophila* Research 37: 3

Slee R, Bownes M. (1990). Sex Determination in *Drosophila melanogaster*. Quarterly Review of Biology 65: 175-204

Smith J L, Wilson J E, Macdonald P M. (1992). Overexpression of *oskar* mRNA directs ectopic activation of *nanos* and presumptive pole cell formation in *Drosophila* embryos. Cell 70: 849-859

Sondergaard L, Mauchline D, Egetoft P, white N, Wulff P, Bownes M. (1995). Nutritional response in a *Drosophila* yolk protein gene promoter. Molecular and General Genetics 248: 25-32

Spradling A C, Digan M E, Mahowald A P, Scott M, Craig E A. (1980). Two clusters of genes for major chorion proteins of *Drosophila melanogaster*. Cell. 19: 905-914

Spradling A C. (1993a). Developmental genetics of oogenesis. In *The Development of* Drosophila melanogaster. ed M Bate and A Martinez-Arias. Cold Spring Harbor Laboratory Press, Cold Spring Harbor pp1-70

Spradling A C. (1993b). Germline Cysts: Communes That Work. Cell 72: 549-551

Sprenger F, Stevens L M, Nüsslein-Volhard C. (1989). The *Drosophila* gene *torso* encodes a putative receptor tyrosine kinase. Nature 338: 478-483

St Johnston D, Driever W, Berleth T, Richstein S, Nüsslein-Volhard C. (1989). Multiple steps in the localization od *bicoid* RNA to the anterior pole of the *Drosophila* oocyte. Development Supplement 13-19

St Johnston D, Beuchl D, Nüsslein-Volhard C. (1991). *staufen*, a gene required to localise maternal RNAs in the *Drosophila* egg. Cell 66: 51-63

St Johnston D, Nüsslein-Volhard C. (1992). The orgin of pattern and polarity in the *Drosophila* embryo. Cell 68: 201-219

St Johnston D. (1995). The intracellular localization of messenger RNAs. Cell 81: 161-170

Steel K P, Brown S D M. (1994). Genes and deafness. Trends in Genetics 10: 428-435

Stein D, Roth S, Vogelsang E, Nüsslein-Volhard C. (1991). The polarity of the dorsoventral axis in the *Drosophila* embryo is defined by an extracellular signal. Cell 65: 725-735

Steinhauer W R, Kalfayan L J. (1992). A specific *ovarian tumor* protein isoform is required for efficient differentiation of germ-cells in *Drosophila* oogenesis. Genes and Development 6: 233-243.

Steinmann-Zwicky M, Schmid H, Nöthiger R. (1989). Cell-autonomous and inductive signals can determine the sex of the germ line in *Drosophila* by regulating the gene *Sxl*. Development 120: 707-716

Steinmann-Zwicky M. (1992). How do germ-cells choose their sex - *Drosophila* as a paradigm. Bioessays 14: 513-518.

Stevens L M, Frohnhöfer H G, Klingler M, Nüsslein-Volhard C. (1990). Localized requirement for *torso-like* expression in follicle cells for development of terminal anlagen of the *Drosophila* embryo. Nature 346: 660-663

Steward R. (1989). Relocalization of the *dorsal* protein from the cytoplasm to the nucleus correlates with its function. Cell 59: 1179-1188

Stroumbakis N D, Li Z, Tolias P P. (1994). RNA- and single stranded DNA-binding (SSB) proteins expressed during *Drosophila melanogaster* oogenesis: a homolog of bacterial and eukaryotic mitochondrial SSBs. Gene 143: 171-177

Sugiyama S, Okada M. (1990). Cytoplasmic factors determining anterioposterior polarity in *Drosophila* embryos. Roux's Archives of Developmental Biology 198: 402-410

Suter B, Romberg L M, Steward R. (1989). *Bicaudal-D*, a *Drosophila* gene involved in developmental asymmetry: localized transcript accumulation in ovaries and sequence similarity to myosin heavy chain tail domains. Genes and Development 3: 1957-1968

Suter B, Steward R. (1991). Requirement for phosphorylation and localization of the *Bicaudal-D* protein in *Drosophila* oocyte differentiation. Cell 67: 917-926.

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

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

Theurkauf W E, Alberts B M, Jan Y N, Jangens T A. (1993). A central role for microtubules in the differentiation of *Drosophila* Oocytes. Development 118: 1169-1180

Theurkauf W E. (1994). Microtubules and cytoplasm organization during *Drosophila* oogenesis. Developmental Biology 165: 352-360.

Tolias P P, Kafatos F C. (1990) Functional dissection of an early *Drosophila* chorion gene promoter: Expression throughout the follicular epithelium is under spatially composite regulation. EMBO Journal 9: 1457-1464

Tolias P P, Konsolaki M, Halfon M S, Stroumbakis N D, Kafatos F C. (1993). Elements controlling follicular expression of the c36 chorion gene during *Drosophila* oogenesis. Molecular and Cellular Biology 13: 5898-5906

Van Deusen E B. (1976). Sex determination in germ line chimeras of *Drosophila melanogaster*. Journal of Embryology and Experimental Morphology 37: 173-185

Verheyen E M, Cooley L. (1994). Profilin mutations disrupt multiple actin-dependent processes during *Drosophila* development. Development 120: 717-728

Wang C, Lehmann R. (1991). *Nanos* is the localized posterior determinant in *Drosophila*. Cell 66: 637-647

Wang S, Hazelrigg T. (1994). Implications for *bcd* localization from spatial distribution of *exu* protein in *Drosophila* oogenesis. Nature 369: 400-403

Webster P J, Suen J, Macdonald P M. (1994). *Drosophila virilis oskar* transgenes direct body patterning but not pole cell-formation or maintenance of messenger-RNA localization in *Drosophila melanogaster*. Development 120: 2027-2037

Weigel D, Jürgens G, Klingler M, Jäckle H. (1990). Two gap genes mediate maternal terminal pattern information in *Drosophila*. Science 248: 495-498

Weil D, Blanchard S, Kaplan J, Guilford P,Gibson F, Walsh J, Mburu P, Varela A, Levilliers J, Weston M D, Kelley P M, Kimberling W J, Wagenaar M, Levi-Acobas F, Larget-Piet D, Munnich A, Steel K P, Brown S D M, Petit C. (1995). Defective myosin VIIA gene responsible for Usher syndrome type IB. Nature 374: 80-81

Werner A, Moore M L, Mantei N, Biber J, Semenza G, Murer H. (1991). Cloning and expression of cDNA for a Na/P_i cotransporter system of kidney cortex. Proceedings of the National Academy of Sciences of the USA 88: 9608-9612

Wilson C, Kurth Pearson R, Bellen H J, O'Kane C J, Grossniklaus U, Gehring W. (1989). P-element mediated enhancer detection: An efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. Genes and Development 3: 1301-1313

Wharton R P, Struhl G. (1989). Structure of the *Drosophila Bicaudal-D* protein and its role in localizing the posterior determinant *nanos*. Cell 59: 881-892

Wharton R P, Struhl G. (1991). RNA regulatory elements mediate control of *Drosophila* body pattern by the posterior morphogen *nanos*. Cell 67: 955-967

Xue F, Cooley L. (1993). *kelch* encodes a component of intercellular bridges in *Drosophila* egg chambers.. Cell 72: 681-692

Younger-Shepherd S, Vaessin H, Bier E, Jan L Y, Jan Y N. (1992). *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in *Drosophila* sex determination. Cell 70: 911-922.

Yue L, Spradling A C. (1992). *hu-li tao shao*, a gene required for ring canal formation during *Drosophila* oogenesis, encodes a homolog of Adducin. Genes and Development 6: 2443-2454

Zhao L P, Koslovsky J S, Reinhard J, Bähler M, Witt A E, Provance D W, Mercer J A. (1996). Cloning and characterization of myr-6, an unconventional myosin dilute/myosin-V family. Proceedings of the National Academy of Sciences of the USA 93: 10826-10831.