SEX SPECIFIC GENE EXPRESSION IN *MUSCA DOMESTICA*

Neil M. White

THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

The University of Edinburgh

1996

ABSTRACT

The process of sex determination represents a paradigm of many developmental processes requiring a choice between several alternative differentiated fates. Somatic sex-determination in *Drosophila melanogaster* has been extensively characterised, and involves a hierarchical cascade of regulatory genes (namely *Sex-lethal*, *transformer*, and *doublesex*). It is not known, however, whether this process defines a developmental program conserved throughout Dipteran evolution. To address this issue, a molecular approach was employed to detect and clone putative sex-determining loci in Dipteran species distantly related to *Drosophila*.

Monoclonal antibodies were available for *Sex-lethal* (the primary sex-determining gene), as were *doublesex* (the terminal sex-determining gene) cDNA sequences. Western analysis revealed cross-reacting polypeptides (using anti-SXL) in *Calliphora erythrocephala* (the bluebottle) whole body extracts were not expressed in adult males or females, although expression was observed in unsexed pupae. Attention was therefore focused on isolating *doublesex (dsx)* homologues, since initial experiments suggested homologous sequences were present in both *Calliphora* and *Musca domestica* (the housefly) genomic DNA. However, despite using a variety of different approaches, we have been unable to isolate a *dsx* homologue from either of these species.

Since neither *Calliphora* nor *Musca* appeared to be amenable to cross-hybridisation analysis, a strategy was devised to determine if polypeptides functionally equivalent to DSX proteins were present in *Musca*. Transcription of the *Drosophila yolk protein (yp)* genes in the fat body is directly regulated by DSX proteins, such that transcription is activated in females and repressed in males. It has been shown in *Calliphora* that two *yp* genes in this species are expressed in an analogous manner, suggesting regulatory proteins (possibly DSX) are likely to be conserved. I report here the cloning of three independent *Musca domestica yolk protein* gene homologues, and their spatial and temporal expression profiles. Comparisons of Dipteran *yp* gene sequence conservation and the regulation of their expression are made.

These results, along with those from ongoing experiments directly related to the newly isolated *yp* genes described here, suggest the process of sex-determination in *Drosophila* may not represent a conserved developmental program in Dipteran evolution.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Mary Bownes, for her help and support during the course of my studies, and particularly during the more frustrating periods for her patience and understanding. Immense thanks of course to everyone in the Lab, both past and present (namely Angela, Bryce, Claudia, Colin, Craig, Debbie, Debiao, Diane, Dot, Elaine, Janis, Kathleen, Roger, Simone, and Wu-min). Rest assured I shall never forget the introduction of compulsory cake eating sessions on Friday afternoons - keep up the good work and intensity. I am also deeply indebted to everyone involved in washing up and media preparation, although too great in number to mention individually, I thank you all.

Finally, on a more personal note, I would like to thank Sharon for her continued support and understanding. Enormous gratitude is also extended to both my parents, and my brother, for simply being who they are. Special thanks, however, goes to my mother for several cash influxes without which I would be surely be poorer than I am now.

GENERAL ABBREVIATIONS

ABBREVIATION	DEFINITION
Amp	Ampicillin
ATP	Adenosine-5'-triphosphate
BCIP	5-Bromo-4-Chloro-Indolyl-Phosphate
bp	Base Pair(s)
°C	Degrees Centigrade
cDNA	Complementary Deoxyribonucleic Acid
cpm	Counts Per Minute
(d)dATP	2'(3'-di) Deoxyadenosine-5'-Triphosphate
(d)dGTP	2'(3'-di) Deoxyguanosine-5'-Triphospahate
(d)dCTP	2'(3'-di) Deoxycytidine-5'-Triphosphate
(d)dTTP	2'(3'-di) Deoxythymidine-5'-Triphosphate
dNTP(s)	Deoxynucleotide-5'-Triphosphate(s)
ddNTP(s)	Di deoxynucleotide-5'-Triphosphate(s)
ddH_2O	Double Distilled Water
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic Acid
DNAase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Diaminoethanetetra-Acetic Acid
FSB	Formaldehyde Sample Buffer
FW	Formula Weight
g	Gram(s)
HCl	Hydrochloric Acid
HEPES	N-2-Hydroxyethylpiperazine-N'-2-Ethanesulphc Acid
kb	Kilo-Base Pair(s)
kDa	Kilo-Dalton(s)
Klenow	Large Fragment Of DNA Polymerase I

rpm	Revolutions Per Minute
λ	*Lambda* Bacteriophage
L	Litre(s)
M	Molar
mA	Milliampere(s)
mCi	Millicurie(s)
mg	Milligram(s)
ml	Millilitre(s)
mM	Millimolar
mmol	Millimole(s)
min	Minute(s)
MOPS	Morpholinopropanesulphonic Acid
mRNA	Messenger Ribonucleic Acid
ml. wt.	Molecular Weight
ng	Nanogram(s)
nM	Nanomolar
nmol	Nanomole(s)
OLB	Oligo Labelling Buffer
OD	Optical Density
[^{32}P]	β-Emitting Isotope Of Phosphorus
pers. comm.	Personal Communication
PEG	Polyethylene Glycol
pg	Picogram(s)
pH	$-\log_{10}$ (Hydrogen Ion Concentration)
poly(a)$^+$ RNA	Polyadenylated Ribonucleic Acid
psi	Pounds Per Square Inch
RNA	Ribonucleic Acid
RNAase	Ribonuclease A
rRNA	Ribosomal Ribonucleic Acid
[^{35}S]	β-Emitting Isotope Of Sulphur

SDS	Sodium Dodecyl Sulphate
SDW	Sterile Double Distilled Water
Tris	Tris(Hydroxymethyl)-Amino-Methane
Triton-X100	Octylphenoxypolyethoxyethanol
Tween-20	Polyoxyethylene Sorbitan Monolaurate
U	Unit(s)
UTP	Uridine-5'-Triphosphate
UV	Ultraviolet Light
µg	Microgram(s)
µl	Microlitre(s)
µM	Micromolar
µmol	Micromole(s)
V	Volt(s)
v/v	Volume Per Volume
vol.	Volume(s)
w/v	Weight Per Volume
w.r.t.	With Respect To

AMINO ACID ABBREVIATIONS

AMINO ACID	SINGLE LETTER ABBREVIATION
Alanine (Ala)	A
Arginine (Arg)	R
Asparagine (Asn)	N
Aspartic Acid (Asp)	D
Cysteine (Cys)	C
Glutamine (Gln)	Q
Glutamic Acid (Glu)	E
Glycine (Gly)	G
Histidine (His)	H
Isoleucine (Ile)	I
Leucine (Leu)	L
Lysine (Lys)	K
Methionine (Met)	M
Phenylalanine (Phe)	F
Proline (Pro)	P
Serine (Ser)	S
Threonine (Thr)	T
Tryptophan (Trp)	W
Tyrosine (Tyr)	Y
Valine (Val)	V

CONTENTS

TABLE OF CONTENTS

1. INTRODUCTION

Over the course of the last 20 years, intensive research has revealed a complex yet beautifully orchestrated program governing the dimorphic sexual development of the fruit fly *Drosophila melanogaster*. A great deal of this information was obtained using the techniques of modern molecular biology, which allows investigation of interesting (and often lethal) mutants at both a molecular and morphological level. However, despite these advances, little is known of the determination of sex in other Dipteran species. Does the process of sex-determination in *Drosophila* define a process conserved throughout Dipteran evolution, or is it confined to closely related species? I report in this thesis attempts to answer these questions using molecular techniques in the bluebottle (*Calliphora erythrocephala*) and the common housefly (*Musca domestica*).

In this introduction I present a comprehensive review of both somatic and germ line sex determination in *Drosophila melanogaster*. Somatic sex-determination in this species involves a hierarchical cascade of zygotic regulatory genes (namely *Sex-lethal, transformer* and *doublesex*) which are activated differentially (i.e. in a sex-specific manner) in response to the ratio of X chromosome to autosomes. As we shall see in the subsequent text, the use of 'alternative splicing' to control gene expression (reviewed by Bingham *et al.*, 1988; Smith *et al.*, 1989) is a crucial aspect of somatic sex determination in *Drosophila*. The male and female regulatory cascades are illustrated in figures 1.2A and 1.2B respectively. Germline sex-determination in Drosophila however operates via an alternative pathway, and is depicted in Figure 1.8. A brief discussion of other Dipteran sex-determination mechanisms is included, although knowledge of these processes is rudimentary at this time. Particular emphasis is placed on *Musca domestica* sex determination, since this species is more characterised than most, and relates directly to the work reported here. Finally, the regulation of *yolk protein* gene expression is discussed as a model of downstream differentiation gene regulation, since these genes are regulated directly by proteins derived from the terminal sex determining gene, *doublesex*.

During the following text, all occurrences of '*Drosophila*' , '*Musca*' and '*Calliphora*' refer to *Drosophila melanogaster*, *Musca domestica* and *Calliphora erythrocephala* respectively unless specific species are otherwise indicated.

1.1 TAXONOMY OF THE DIPTERA

The Diptera encompass the suborders *Nematocera* (frail flies with long segmented antennae and aquatic larvae [e.g. Mosquitoes]) and *Brachycera* (short antennae, a reduced larval stage and a free pupal stage [e.g. horseflies]). Within the *Brachycera* the infra-orders Tabanomorpha, Asilomorpha and Muscamorpha are found, and the Muscamorpha are further subdivided into the *Aschiza* and *Schizophora* divisions. The Muscamorpha, and particularly the *Schizophora*, are considered to be the most advanced Diptera. The *Schizophora* encompass the series Acalyptratae (including the Ephydroidae superfamily in which the Drosophilidae are found), the Calyptratae (including the Muscidae [e.g. the

housefly, *Musca domestica*]), the Sarcophigidae (e.g. the fleshfly) and the Calliphoridae (e.g. the bluebottle, *Calliphora erythrocephala*) among others. Figure 1.1 illustrates the phylogenetic relationships between the prominent Dipteran species discussed in this thesis.

1.1.1 Evolutionary time scales

Originally, the time of origin of a species was estimated using paleontological fossil evidence (sparse in the case of *Drosophila*), However, the onset of molecular techniques has allowed a comparative analysis of protein and/ or gene divergence between different taxa as an index of evolutionary time scale. Such molecular studies, which in themselves are speculative due to the influence of selective pressure, can provide interesting frameworks by which the reliability of paleontological studies can be judged (Beverley and Wilson, 1982). It is therefore of interest to analyse several gene and/ or protein sequences, such that erroneous results may be minimised (i.e. since individual genes will exhibit variable divergence rates, an average of several genes may give more reliable data). With this in mind, the analysis of Larval serum protein immuno-cross reactivity (Beverley and Wilson, 1982; 1984), α-glycerophosphate dehydrogenase (Beverley and Wilson, 1984; re-evaluated from Collier and MacIntyre, 1977), and Alcohol dehydrogenase (Villaroya and Juan, 1991) gene sequence divergence supports paleontological studies favourably and can be summarised as follows. During the Cretaceous period (68 - 130 Million Years [MYR] ago), the Tephritidae, Optidae, Sarcophigidae, Calliphoridae, Muscidae and Drosophilidae families arose, although continued divergence within the Drosophilidae family occurred in the Cainozoic era (i.e. the *Drosophila* radiation began about 62 MYR ago, and the *melanogaster* species group arose at least 26 MYR ago [Beverley and Wilson, 1984]). The Calliphoridae and Drosophilidae families are thought to have arisen from a common ancestor approximately 99 MYR ago (Beverley and Wilson, 1984).

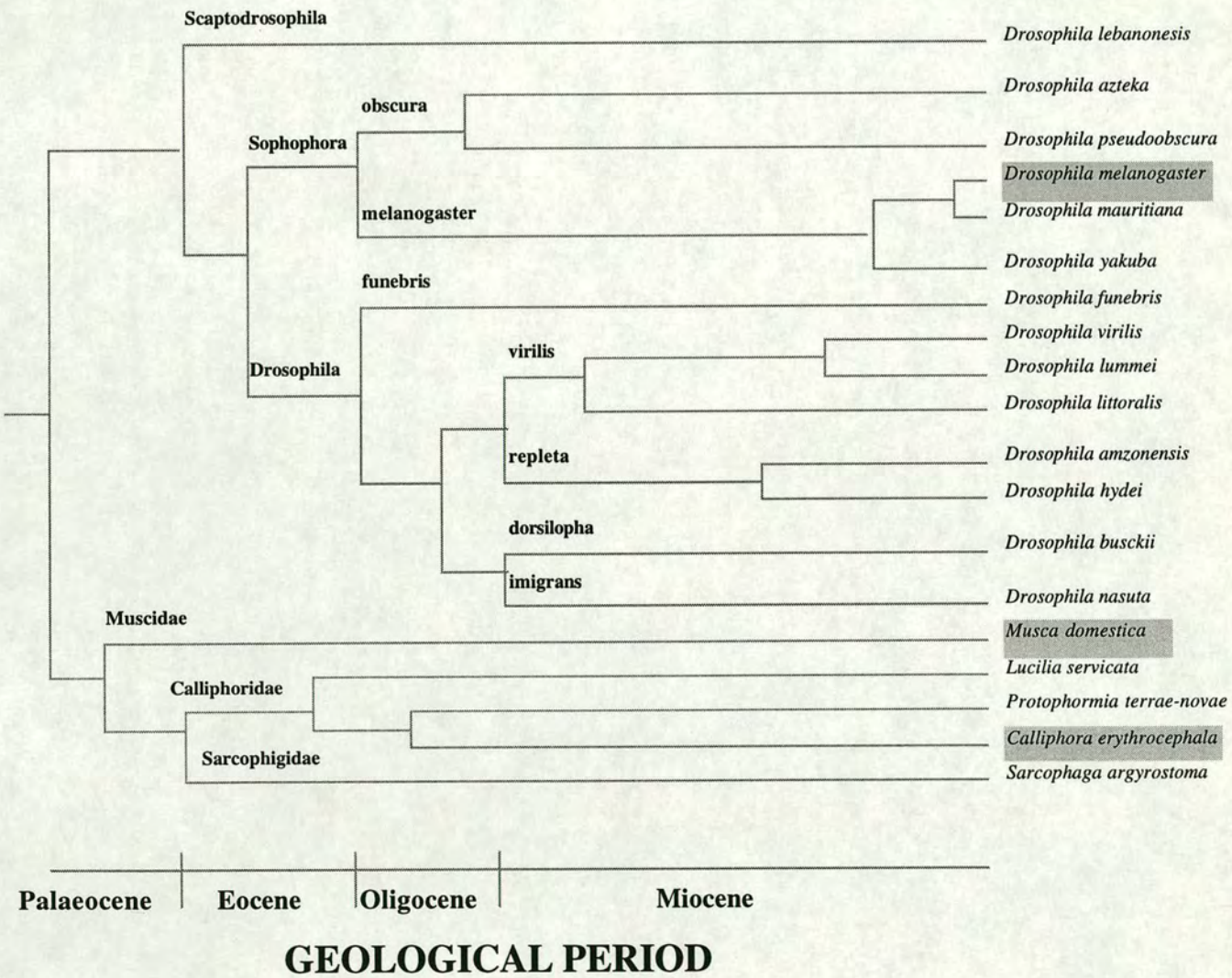


Figure 1.1. Phylogenetic tree of the prominent Dipteran species described in this thesis (shaded). For a more complete description please refer to the main text. Note that this figure is not to scale, and derives from Martinez (1991).

1.2 SOMATIC SEX DETERMINATION IN DROSOPHILA

The process of somatic sex determination in *Drosophila* involves a cascade of regulatory genes the state of activity of which ultimately governs all aspects of male or female development (for reviews see Baker and Belote, 1983; Slee and Bownes, 1990; Belote, 1992; Burtis and Wolfner, 1992; and Burtis, 1993). However, there must be some initial signal telling the developing embryo whether a male or female chromosomal constitution is present, such that this cascade is activated in an appropriate manner. In *Drosophila*, the primary determinant of sex is the ratio of X chromosomes to sets of autosomes (Bridges, 1921).

1.2.1 ASSESSMENT OF THE X/A RATIO

X chromosome dosage was first shown to be involved in the choice between male and female sexual fate by Bridges in 1916, where diplo-X individuals were shown to develop as females irrespective of the presence of a Y chromosome. These initial findings were resolved in 1921, when it was proposed that the X chromosome to autosome ratio determined sexual phenotype (Bridges, 1921). Individuals with an X/A ratio of ≥ 1.0 develop into females, X/A $\leq$ 0.5 develop as males, whereas flies whose chromosomal constitution generates intermediate X/A ratio's develop as intersexes.

In the most striking intersexes, those having a 2X/3A chromosomal constitution, individual cells are seen to follow either a male or female mode of development, with patches of mosaic tissue being characteristically large. This suggests sex is determined early in embryogenesis, since there is little cell mixing during *Drosophila* development. Interestingly, the penetrance of the mosaicism is tissue specific. For example, the forelegs only undergo female differentiation in extreme-female intersexes, while male differentiation is only observed in extreme-male intersexes (see review by Baker and Belote, 1983). This is suggestive of factors responding in a threshold specific manner.

But how exactly is the X/A ratio assessed? Since gynandromorphs (sexual mosaics) differentiate to produce both male and female cells, it can be inferred that most aspects of somatic sex are determined in a cell autonomous manner (i.e. diffusible molecules do not act to influence the sex of surrounding cells). Early experiments designed to identify elements acting to assess the chromosomal constitution relied upon the use of fly strains carrying duplications or deletions of various chromosomal regions (Dobzhansky and Schultz, 1931; Dobzhansky and Schultz, 1934). However, this is problematic since a gene need not have a direct role in assessing the X/A ratio to have an effect on the morphology or viability of the fly. Nevertheless, many feminising regions were identified on the X-chromosome, and prompted further investigation.

Many loci have now been identified which have profound effects on sexual phenotype. The majority of these loci however do not act to perturb sex determination in

a dose dependant manner, and as such cannot be involved in the assessment of the flies chromosomal balance. Indeed, these genes have now been characterised extensively, and are known to act in response to the X/A ratio signal, and will be discussed in later sections. Notably the gene *Sex-lethal* (*Sxl*), initially identified by Muller and Zimmering (1960), is now known to be the primary target for the X/A balance signal (see section 1.12.1). Loss of function *Sxl* mutations result in female-specific embryonic lethality, but have no discernible effect on male development (Cline, 1978; Marshall and Whittle, 1978).

Elements involved in assessing the zygotic chromosomal constitution are classified as being either 'numerator' (X-linked) or 'denominator' (autosomal) elements. Numerator elements, when raised in dose, will effectively increase the X/A ratio, whereas a decrease in dose results in lower X/A values. Denominator elements have the reciprocal effect. Elements identified to date acting as numerators are *sisterless-a* (Cline, 1986), *sisterless-b* (Cline, 1988), and *runt* (Duffy and Gergen, 1991). Only *deadpan* remains a confirmed denominator element (Younger-Shepherd *et al.*, 1992).

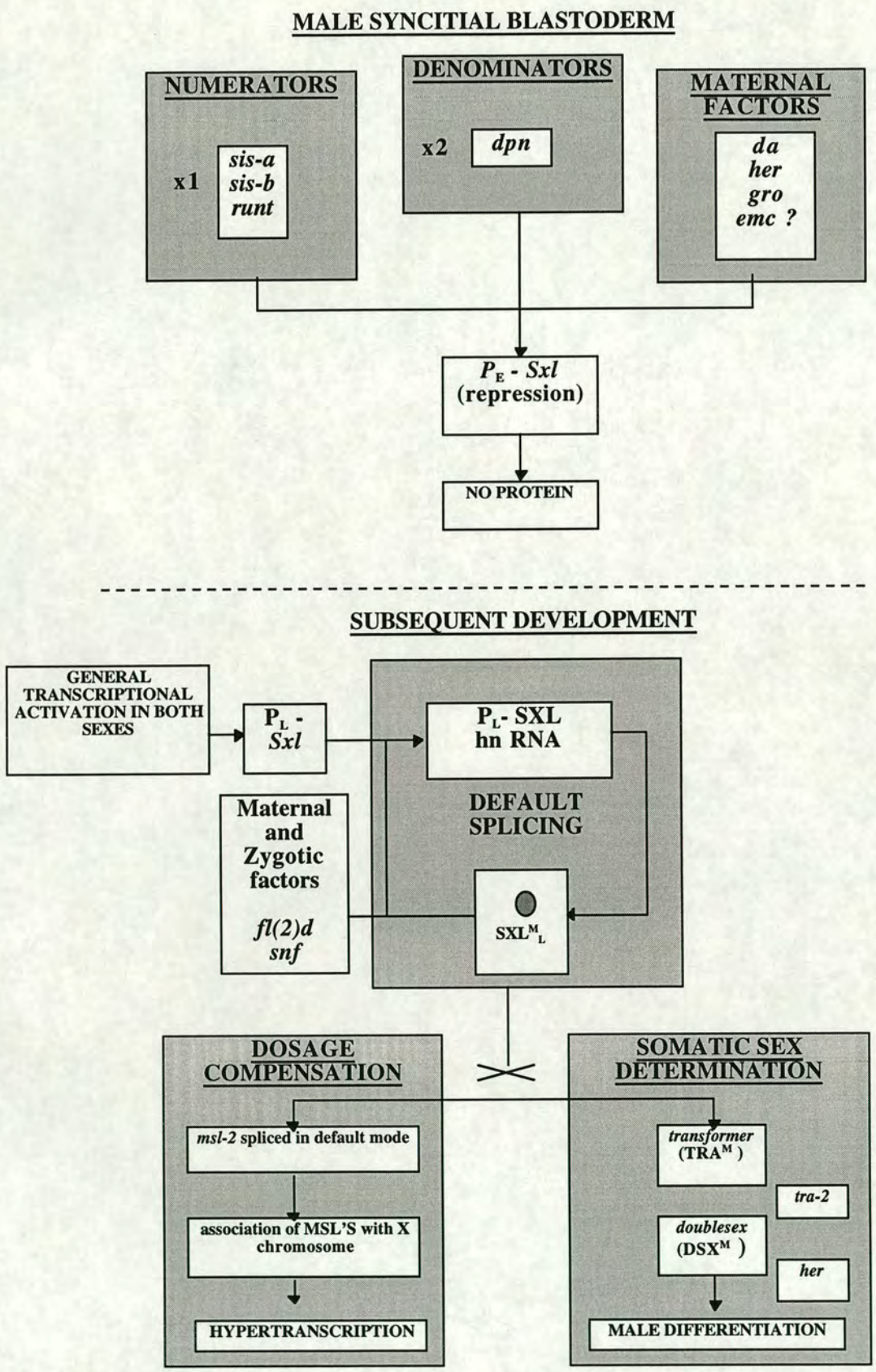


Figure 1.2A Effects of *Sex-lethal (Sxl)* activity on male differentiation and its involvement in somatic sex determination. Please refer to text for a description of the genes implicated. Protein products are marked as non-italic capitals. SXL^{M}_{L} derives from the 'late' promoter. TRA^{M} and DSX^{M} refer to the male-specific proteins derived from *transformer* and *doublesex*. P_E and P_L are the 'early' and 'late' promoters of *Sxl* respectively.

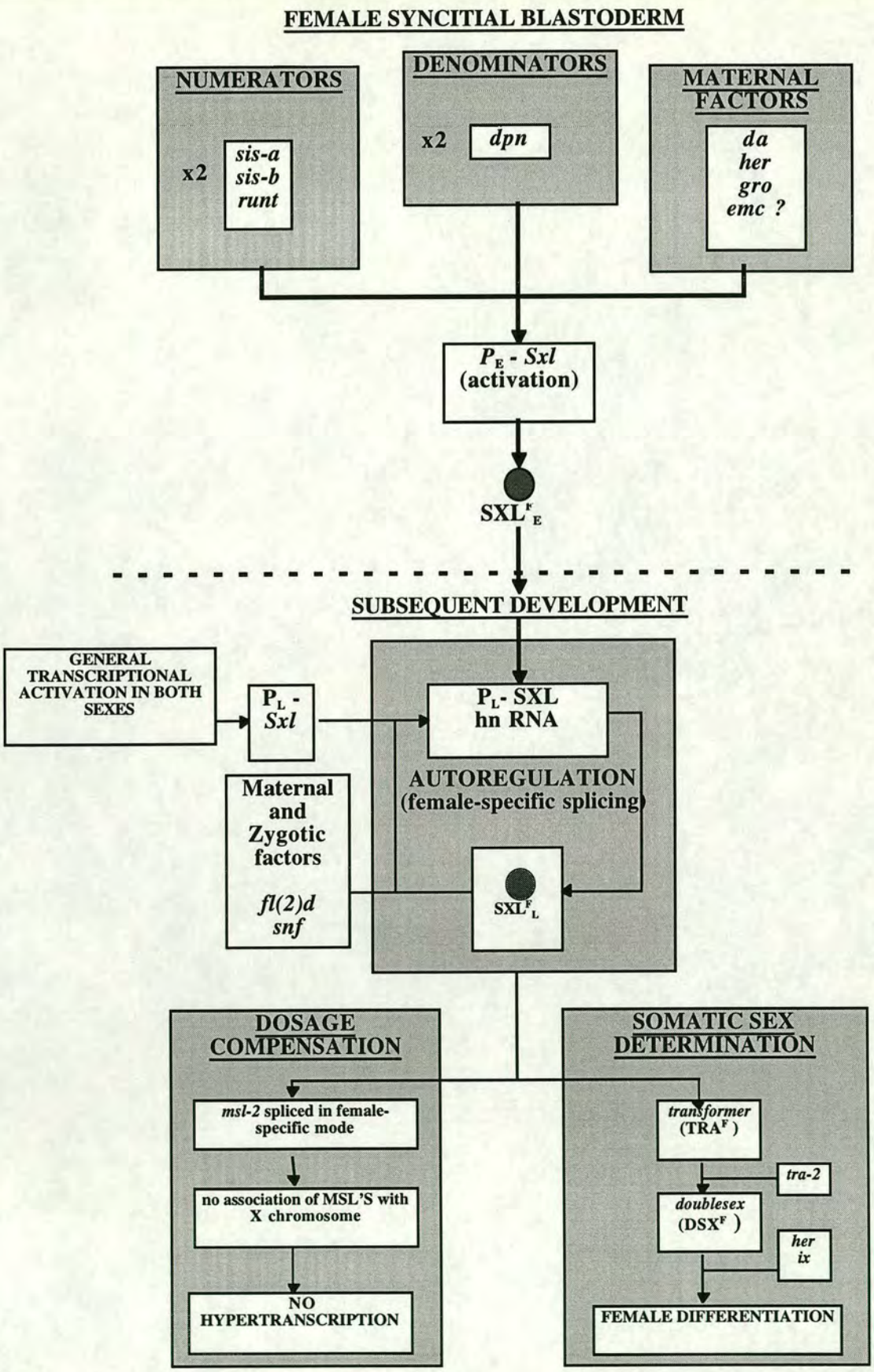


Figure 1.2B Effects of *Sex-lethal (Sxl)* activity on female differentiation and its involvement in somatic sex determination. Please refer to appendix text for a description of the genes implicated. Protein products are marked as non-italic capitals. SXL^F_E refers to male-specific protein derived from the 'early' promoter, whereas SXL^F_L derives from the 'late' promoter. TRA^F and DSX^F refer to the female-specific proteins derived from *transformer* and *doublesex*. P_E and P_L are the 'early' and 'late' promoters of *Sxl* respectively.

1.2.1.1 NUMERATOR ELEMENTS

1.2.1.1.1 sisterless-a

The *sisterless-a* (*sis-a*) gene was initially identified as a recessive female-specific embryonic lethal with a masculinising effect on triploid intersexes (Cline, 1986). Originally mapped to the distal breakpoint of *Df(1)N71*, the *sis-a* gene has now been cloned (Erickson and Cline, 1993), and encodes a predicted protein consisting of 189 amino acids (21 kDa) with extensive sequence similarity to the basic helix-loop-helix family of transcription factors (see Vinson *et al.*, 1989 for information regarding bZIP proteins).

Erickson and Cline (1993) analysed the temporal and spatial distribution of *sis-a* expression by *in-situ* hybridisation to whole mount embryos. Initially, transcripts first appear in all nuclei just prior to the migration of the pole cell nuclei to the periphery of the embryo (end of nuclear cycle 8), and are absent from pre-pole cell nuclei by stage 9. The signal remains closely associated with nuclei until the end of stage 10, but is distributed homogeneously throughout the embryo by stage 12. The most abundant levels of *sis-a* are detected during nuclear cycles 12 and 13, followed by a rapid decrease in intensity. Yolk nuclei, which accumulate high levels of *sis-a* (remaining tightly associated with the nuclei) show no apparent reduction in signal until 10 - 11 hours post fertilisation.

1.2.1.1.2 sisterless-b

During the subsequent analysis of *sis-a*, Cline (1988) demonstrated that duplications of an X-chromosomal region located in the *achaete-scute* gene complex were lethal to chromosomally male flies also carrying a duplication of *Sxl*.

Mutations within the *achaete-scute* gene complex (*AS-C*) fall into two main classes. Recessive loss of function *achaete* and *scute* alleles remove chaetae to varying degrees (Muller, 1955; Garcia-Bellido, 1979), while dominant gain of function alleles promote ectopic chaetae development (Note : chaetae are a set of chemoreceptors and mechanoreceptors distributed in positional [microchaetae] and density [macrochaetae] patterns in the adult cuticle [Sturtevant, 1921; Plunkett, 1926; Stern, 1954]). The complex is subdivided into the four principal regions of *achaete* (*ac*), *scute-α* (*sc-α*), *lethal of scute* (*lsc*) and *scute-β* (*sc-β*) (Garcia-Bellido, 1979). *AS-C* mutations are classified according to their effect on microchaetae (the *ac* alleles), macrochaetae (the *sc* alleles), or the embryonic central nervous system (*lethal of scute*), as reviewed by Garcia-Bellido (1981).

Cline (1988) initially localised *sis-b* to a 20 Kb interval at chromomere 1B3 near the *sc-α* functional unit. Cloned DNA encompassing the *AS-C* complex was shown to contain six independent transcriptional units (Campuzano *et al.*, 1985), and *sis-b* function has since been shown to be encoded by the T4 transcription unit of *sc-α*, although the T5

transcription unit can partially substitute for *sis-b* activity (Torres and Sanchéz, 1989, Parkhurst *et al.*, 1993). That *sis-b* activity resides in the *sc-α* gene has been confirmed using transgenes to complement *sis-b* mutants (Erickson and Cline, 1991; Parkhurst *et al.*, 1993). The nucleotide sequence of both the T4 and T5 transcription units shows homology to the *myc* family of proteins (Villares and Cabrera, 1987).

In-situ hybridisation (Erickson and Cline, 1993) shows that *sis-a* and *sis-b* have very similar expression patterns during early embryogenesis. Transcripts (*sis-b*) are first detected in nuclear cycle 9 (associated with the somatic nuclei through cycles 9 and 10), increasing in abundance until the end of cycle 12 (during cycle 11 the transcripts become distributed homogeneously throughout the embryo). Transcript abundance then decreases rapidly until by early nuclear cycle 14, little or no transcripts are apparent. This early expression pattern is consistent with a role in the determination of sex, since *sis-b* activity is required during the first 1 - 2.5 hours post fertilisation (cycles 9 - 14), the most sensitive period being between 1.5 -2 hours (cycles 12 -13).

Subsequent to the early expression pattern, a second round of expression is initiated during late nuclear cycle 14 consistent with the role of *sc-α* in neurogenesis.

1.2.1.1.3 runt

An interaction between the X-linked gene *runt* (19E2) and 1A1; 1B5-6 (i.e. the *AS-C* region) was observed during a screen for maternal dose dependant enhancers of a dominant *runt* segmentation phenotype (Duffy and Gergen, 1991). Females *trans*-heterozygous for a deficiency of the *sis-b* region (*Df[1]260-1*) and a null *runt* allele ($runt^{LB5}$) have severely reduced viability. This interaction was subsequently mapped to the *sc-α* transcriptional unit of the *AS-C* complex.

Transgenes carrying one of two insertions of *runt* (OP1 and OP2) were shown to suppress the *runt⁻ / sis-b⁻* deficiency interaction previously described (transposons are detailed in Gergen and Butler, 1988). This suppression is conferred by the coding region, since upstream sequence shows no suppression when transformed on its own. Also, Erickson and Cline (1991) have demonstrated that a transposon containing T4 sequence is also able to rescue the *runt⁻ / sis-b⁻* deficiency phenotype, further confirming that the vital interaction is with *sis-b*. This interaction is confined to *sis-b*, and not the *sc-α* function in neurogenesis, since surviving individuals from $sc^{10\text{-}1}$ / run^{LB5} x $Df(1)sc^{19}$ / run^{LB5} crosses show no apparent bristle defects (Duffy and Gergen, 1991).

Suppression of the *runt⁻/sis-b⁻* interaction is also observed with $Sxl^{M\#1}$ (a dominant male-lethal *Sxl* allele that constitutively expresses female *Sxl* function [Cline, 1979; Cline 1988]), suggesting *Sxl* is the target of the *runt / sis-b* interaction. Increasing the dose of *runt* (using the OP1 and OP2 transgenes) feminises triploid intersexes, whereas decreases in dose have the reciprocal effect. Therefore, all of the above data indicate that *runt* acts as a numerator element in the assessment of the X/A ratio.

In contrast to *sis-a* and *sis-b* however, which are expressed ubiquitously throughout the embryo, *runt* only appears to be required for *Sxl* activation in a spatially restricted domain. *In-situ* hybridisation shows a lack of *Sxl* expression in the central domain of *runt⁻/sis-b⁻* embryos, even though wild type expression levels are still observed at the anterior and posterior of the embryo (Duffy and Gergen, 1991). Consistent with this data, *runt* is initially expressed in a broad central domain, which becomes resolved into seven stripes of strong expression reflecting *runt*'s role in segmentation.

It is intriguing to note that *runt* encodes a protein lacking any evident DNA binding domain (Kania *et al.*, 1990), since both *sis-a* and *sis-b* proteins contain such domains. It is evident, however, that a low level of *Sxl* expression is observed in individuals completely lacking *runt* activity, suggesting *runt* may be involved in amplifying the effects of other numerator elements, as opposed to directly assessing the X/A ratio itself.

Since *runt* acts only in the central domain of the embryo, Duffy and Gergen (1991) have proposed that other factors must be involved in *Sxl* regulation at the anterior and posterior. Interestingly, embryos lacking *runt* activity but also carrying four copies of *bicoid* show a broader domain of *Sxl* expression than *runt⁻* embryos with wild type *bicoid* dose (i.e. the central domain lacking *Sxl* expression is less extensive). Also, *torso* gain-of-function mutants (which result in expansion of the embryonic termini at the expense of the segmental body regions) show uniform expression of *Sxl* irrespective of *runt* activity. However, neither *bicoid* nor *torso* need have a direct effect on *Sxl* expression. Therefore, no real conclusions can be drawn from this data except that *Sxl* expression is regulated by different factors in different domains of the embryo.

1.2.1.2 DENOMINATOR ELEMENTS

1.2.1.2.1 deadpan

Bier *et al.* (1989) first identified *deadpan* in an enhancer trap screen, and subsequently localised the gene to chromosome 2. It is expressed in embryonic and adult neuroblasts, and flies homozygous for a null P-*lacW* insertion mutation (*dpn^{1}*) or hemizygous (*dpn^{1} / Df[2R]193A*) die at various stages throughout development (the relative viability of *dpn^{1}* adults ranged from 1.5% in males to 8-10% in females). This male lethality is even more pronounced in flies carrying two or three copies of *scute^{+}* (no male flies were ever recovered), yet female viability is higher than in their homozygous *dpn^{1}* counterparts (Younger-Shepherd *et al.*, 1992). Males with two copies of *sc^{+}* and one copy of *dpn^{+}* are less than half as viable as siblings carrying two copies of both genes. In general, the viability of males is seen to decrease in relation to an increase in the ratio of *sc^{+}* to *dpn^{+}*, suggesting *dpn* acts as a denominator element.

Younger-Shepherd *et al.* (1992) hypothesised that decreasing *sc^{+}* dose relative to *dpn^{+}* in females would promote a more male-like level of dosage compensation (since *Sxl* represses hypertranscription of X-chromosomes in females [Cline, 1978; Cline, 1983; Lucchesi and Skripsky, 1981; Bernstein *et al.*, 1994]). As the lethal effects of *Sxl* are a

consequence of improper dosage compensation, and not the determination of sexual morphology, dosage compensation levels can be correlated to viability (Cline, 1979a; Lucchesi and Skripsky, 1981). Females carrying two doses of sc^+ relative to one dose of dpn^+, or one copy of sc^+ relative to one or two copies of dpn^+, appear to have wild type morphology and viability. However, females carrying one dose of sc^+ relative to three doses of dpn^+ have poor viability (only 10% survived), but no effect is observed in males (Younger-Shepherd et al., 1992). This effect is purely zygotic, as no difference in these results is observed if the *dpn* duplications are derived maternally or paternally.

Since *dpn* is proposed to act as a repressor of *Sxl* activation in males, reduced male viability as a consequence of *dpn* mutations should be rescued by the female-lethal $Sxl^{F\#1}$ allele. This is indeed found to be the case, with progeny showing restored viability. Further confirmation that *dpn* regulates *Sxl* is seen in male dpn^- embryos stained with anti-SXL antibody (Younger-Shepherd *et al.*, 1992), where patchy expression of SXL protein is observed (no protein is normally detected in males).

Consistent with a role in *Sxl* regulation, *dpn* transcripts (2.3 Kb) are first detected homogeneously distributed throughout the embryo in early stage 12, with transcript abundance increasing rapidly until late stage 12 (*sis-a*, *sis-b* and *runt* transcripts are also most abundant during this stage). This broad expression pattern is resolved in stage 13 to give a transient gap gene-like expression, ultimately resulting in a pattern of eight striped segments approximately three hours into embryogenesis. Subsequent to the commencement of gastrulation, the pair-rule *dpn* expression pattern is lost, and expression is only seen thereafter in defined cells of the developing nervous system (Bier *et al.*, 1992).

1.2.1.3 THE INFLUENCE OF MATERNALLY ACTING FACTORS

In addition to the embryonic requirement for zygotically derived proteins, several maternally derived factors, which when mutated, perturb sex determination have been identified. These loci include *daughterless* (Bell, 1954), *fl(2)*d (Granadino *et al.*, 1990, 1991), fs*(1)1621* (Cline, 1988; also known as *liz* [Steinmann-Zwicky, 1988] and *sans-fille* [Oliver *et al.*, 1988]), *hermaphrodite* (Pultz *et al.*, 1994), *groucho* (Paroush *et al.*, 1994), and possibly *extramachrochaetae* (Younger-Shepherd *et al.*, 1992; Cline, 1993).

1.2.1.3.1 daughterless

Loss of function mutations at the *daughterless* (*da*; 2-41.5) locus are intriguing in that not only do they result in a lethal maternal effect, but the lethality is restricted to female progeny. At 25°C, homozygous *da* mothers produce only male (1X2A) progeny, while daughters (2X2A) die as embryos regardless of their genotype with respect to *da* (Bell, 1954; Sandler, 1972; Mason, 1973; Cline, 1976; Cline, 1980). The survival of gynandromorphs derived from homozygous *da* mothers reared at 25°C is related primarily to the amount of female tissue they possess, with extensive variation in sensitivity to the *da* maternal effect seen in different tissues. For example, genitalia and

the anterior region are particularly susceptible (Cline, 1976). A zygotic role for *da* (the recessive lethal effect) unrelated to the determination of sex is also inferred since *da/da* progeny from *da/+* mothers are less viable than their *da/+* siblings.

Cline (1976) demonstrated that the temperature sensitive period (TSP) of the *daughterless* maternal sex ratio effect commences in oogenesis several days prior to, and extends to within 9 hours of, oviposition. However, consistent with a role in sex determination, the critical TSP appeared to be in the first three hours post-fertilisation, as embryos were less sensitive to increases in temperature subsequent to this. The recessive lethal effect in contrast has a TSP commencing shortly after, and ending approximately 14 hours subsequent to, fertilisation. Thus the maternal lethal and zygotic recessive lethal effects of *da* are clearly distinguishable (the TSP of the maternal effect occurs in the 3rd hour of embryogenesis, whereas the TSP of the recessive lethal effect is just beginning at this stage).

The *daughterless* gene has now been cloned, and encodes a 74kDa protein with significant similarity to both the BICOID and PAIRED proteins (Cronmiller *et al.*, 1988). This is intriguing since variations in *bicoid* dose affect the spatial expression of *Sxl* in *runt*$^{-}$ embryos (Duffy and Gergen, 1991). However, one would assume that *bicoid* would antagonise the activity of *daughterless*, but in such embryos an increase in *bicoid* dose leads to an expanded domain of *Sxl* activation, not a reduction. Further investigation is required to determine if this is a natural role of *bicoid*, or if the elevated dose (and therefore elevated protein levels) are merely acting in a manner analogous to *da* as a consequence of the observed protein similarities. Independent characterisation of the *da* locus also revealed a critical role in the development of the peripheral nervous system and parts of the central nervous system (Caudy *et al.*, 1988a). In a discovery with important implications in the regulation of sex determination, *da* was shown to have homology to, and interact with, genes of the *AS-C* complex (Caudy *et al.*, 1988b; Dambly-Chaudière *et al.*, 1988; Murre *et al.*, 1989a). Indeed Murre *et al.* (1989b) demonstrated that DNA binding activity is absent from proteins encoded by the *AS-C* (T3) and *da* when acting independently, but is present when these proteins are mixed. Thus a possible model accommodating the involvement of both *da* and *AS-C* genes in both sex determination and neurogenesis is conceivable. Essentially, DA would act as a ubiquitous transcription factor which becomes activated by dimerisation with specific DNA binding proteins. Such DNA binding proteins would confer different specificitys and be expressed in a temporal and spatial manner to allow regulation of different target genes in a tissue-specific manner. Repression of such activators could be achieved by sequestration of partner DNA binding proteins with those having similar dimerisation motifs to DA. The possible involvement of DA/SIS-B heterodimers in the regulation of *Sxl* expression will be discussed in subsequent sections.

A relationship between *da* and *Sxl* is inferred because the male-lethal $Sxl^{M\#1}$ allele rescues diplo-X flies from the lethal maternal effect of *da* (Cline, 1978; Cline, 1980; Cline

1983). Also, triploid intersex progeny from *da* mothers are masculinised relative to progeny derived from wild type mothers. Phenotypically, this masculinisation is analogous to that observed in triploid intersexes with reduced zygotic Sxl^+ activity. Cline (1983) further demonstrated that both *Sxl* and *da* affect the rate of transcription from X-linked loci such as *Bar (B)* and *Hairy wing (Hw)*. For example, transcription of these two loci in triploid intersexes also carrying $Sxl^{M\#1}$ was at a rate equivalent to that observed in 2X2A flies. The same conclusions were drawn by Gergen (1987) in a separate analysis of the segmental expression of *runt* in mutant *da* and *Sxl* embryos.

Characterisation of null alleles has shown that *daughterless* is required zygotically in all somatic cells for the development of both sexes, but is not required for the development of the germ line (Cronmiller and Cline, 1987). *da* is active in the germ line to provide developing oocytes with the maternal factor required for the survival of female progeny. However, expression associated with the somatically derived follicle cells surrounding the developing oocyte is correlated with a role in egg chamber morphogenesis (Cline, 1976; Cummings and Cronmiller, 1994).

1.2.1.3.2 female-lethal-2-d [fl(2)d]

The *fl(2)d* region is involved in the maintenance pathway of *Sxl* activity in female flies, whereby of two EMS induced mutations, $fl(2)d^1$ homozygosity (temperature sensitive allele) results in female lethality and male semi-lethality, whereas $fl(2)d^2$ homozygosity is lethal in both sexes (Granadino *et al.*, 1990; Granadino *et al.*, 1991). Deficiency mapping has localised the *fl(2)d* region to between 50A2-5 and 50F-1A1 on the second chromosome.

An interaction with *Sxl* was initially noted since in homozygous X/X; $fl(2)d^1/+$ flies, transcripts derived from the *Sxl* locus are spliced in a male default mode. Also, the effects of $fl(2)d^1$ are suppressed by $Sxl^{M\#1}$ in females ($Sxl^{M\#1}$ constitutively expresses female specific *Sxl* function), but rescue males from the feminising effects of this *Sxl* allele. In homozygous $fl(2)d^1$ flies raised at 18°C (sterile; oogenesis arrests at stage 8-9), $Sxl^{M\#1}$ restores oogenesis although the flies remain sterile due to a failure in oviposition (Granadino *et al.*, 1991). Thus *fl(2)d* appears to be involved in the regulation of female-specific splicing of *Sxl* transcripts.

1.2.1.3.3 sans-fille (4F - 4F11)

Originally identified as a female-sterile mutation [*fs(1)1621*] by Gans *et al.* (1975), *sans-fille* (*snf*, Oliver *et al.*, 1988; also known as *liz* [Steinmann-Zwicky, 1988]) is involved in *Sxl* expression in the germ line and soma. This is evident since the *snf* phenotype is suppressed by constitutive *Sxl* alleles (Steinmann-Zwicky, 1988; Salz, 1992). Also, the two germ line specific transcripts of *Sxl* are absent from *snf* mutant females (Salz, 1992). The sterility associated with the *snf* mutation is due to uncontrolled proliferation of germarial cells, ultimately giving rise to ovarian tumours (a phenotype also observed in *Sxl* mutants).

A somatic interaction between *snf* and *Sxl* is only uncovered when the probability of *Sxl* activation is reduced by mutations in numerator elements (Cline, 1988). Using flies heterozygous for a variety of *Sxl* alleles, Oliver *et al.* (1988) were able to determine that *snf* is involved in the maintenance, rather than initiation, pathway of *Sxl* expression. This somatic interaction derives from maternally contributed *snf* protein since homozygous snf^{8H} females crossed to males carrying a null *Sxl* allele yields no female progeny (snf^{8H} homozygous females do not show the female sterility associated with the snf^{1621} allele, but retain the female-lethal synergistic interaction with *Sxl*). This maternally derived *snf* protein is presumably also involved in the activation of zygotic *snf* expression, since snf^{+} $Sxl^{M\#1}$/Y males can survive in the absence of maternal *snf* product (Steinmann-Zwicky, 1988). Consistent with a role in the maintenance of *Sxl* expression, *snf* mutant embryos show aberrant *Sxl* transcript splicing in post 5hr old embryos. This is co-ordinate with expression, where in such embryos SXL^{F} protein is only detected in 0-5hr embryos (Albrecht and Salz, 1993).

More recently, Flickinger and Salz (1994) have cloned the *snf* gene and shown that it encodes a nuclear protein with functional similarity to the mammalian U1A snRNP protein. Thus, *snf* is likely to be involved in general splicing reactions, and indeed a *snf* null phenotype is a non sex-specific lethal. Characterisation of the original mutation, snf^{1621}, reveals a missense mutation in the RNA recognition motif suggesting this protein would have either reduced affinity for, or an alteration in the recognition of, its target sequences. The second option seems most likely since this allele appears to specifically affect the splicing of *Sxl* transcripts, and does not display the null phenotype.

1.2.1.3.4 virilizer

The chromosome containing the temperature-sensitive mutation *virilizer* (vir^{ts}) was initially isolated in an EMS induced mutagenesis screen for female-sterile mutations on the second chromosome (Schüpbach and Wieschaus, 1989). When raised at 29°C, diplo-X flies homozygous for this mutant chromosome are transformed into sterile intersexes, whereas haplo-X flies are unaffected. Hilfiker and Nöthiger (1991) have recently localised the gene more precisely to map position 2-103·9, and the mutation is uncovered by *Df(2R)bw-S46*.

Since flies raised at 29°C of the genotype X/X; vir^{ts}/vir^{ts} also transformed with a hs-*tra* construct (encoding TRA^{F}) develop as females, *vir* must act upstream of *tra* (Hilfiker and Nöthiger, 1991). However, such flies remain sterile even at 25°C. Also, X/X; dsx^{D}/+; vir^{ts}/vir^{ts} flies raised at 25°C show the *dsx* phenotype characteristic of the simultaneous presence of both male- and female-specific DSX proteins (as described by Nagoshi and Baker, 1990). These flies when raised at 29°C are strongly masculinised, suggesting vir^{ts} affects the production of DSX^{F} protein. Although the ovaries of X/X; vir^{ts}/vir^{ts} flies raised at 29°C develop poorly and never complete oogenesis, no sexual transformation of the gonads is observed. This suggest *vir* is not involved in germ line sex determination.

Hilfiker and Nöthiger (1991) determined that the TSP of *virilizer* is in the 3rd larval instar, consistent with a role in the maintenance of female differentiation.

1.2.1.3.5 hermaphrodite

The *hermaphrodite* (*her*) locus (36A3 - 36A11) uncovers a complex pleiotropic phenotype dependant on both maternal and zygotic activity (Pultz *et al.*, 1994). Homozygous her^{1} females are transformed into true intersexes (i.e. individual cells express both male and female characteristics) at semi-restrictive (25°C) and restrictive (29°C) temperatures. Weaker zygotic effects of *her* on male development are also apparent, with viability in both sexes being dependant on the relative severity of the *her* allele.

The maternal effects of *her* are similar to those observed with *da*, such that the viability of XX embryos derived from mutant mothers is severely reduced at elevated temperatures (Redfield, 1926; Pultz *et al.*, 1994). In all *her* alleles analysed, the phenotypic effects of the mutation are enhanced over a deficiency of the region, indicating partial loss-of-function (hypomorphic) as opposed to hypermorphic activity. Also, the maternal and zygotic effects of *her* are clearly separable since one allele, $her^{1(2)mat}$, is only deficient for the maternal function (Pultz *et al.*, 1994). An interaction between maternal *her* and zygotic numerator elements has also been demonstrated, since the viability of *sis-a* embryos derived from *her*⁻ mothers is severely reduced relative to those derived from wild type mothers. Reciprocally, *her*⁻ mothers can rescue the lethality of males associated with a duplication of zygotic $sisterless^{+}$ genes (Pultz and Baker, 1995).

These data all suggest that *her* is involved in the regulation of *Sxl* expression in the developing embryo. Indeed, constitutive *Sxl* alleles can rescue the female-lethality of defective maternal *her*. Further analysis by Pultz and Baker (1995) has shown that *her* is involved in the activation of *Sxl* in female embryos, as opposed to maintaining *Sxl* expression throughout development. In those female embryos analysed, only patchy expression of SXL protein is detected, although variability is observed in penetrance (i.e. the extreme posterior of the embryo maintains *Sxl* expression).

Consistent with the observation that different *her* alleles appear defective in either or both maternal and zygotic function, neither $Sxl^{M\#1}$ nor hs-tra^{F} is able to rescue the intersexuality of XX/ *her*⁻ individuals. Also, no apparent defects in *dsx* transcription are observed in XX/*her*⁻ intersexes, as judged by Northern analysis (Pulz and Baker, 1995). Therefore, the sexual transformation observed in XX/*her*⁻ flies appears to be independent of the sex determination genes isolated to date, and suggests *her* acts downstream of or in parallel to *doublesex*. Pultz and Baker (1995) propose that *her* may interact with *dsx* to regulate both male and female differentiation, since weak feminisation of males is observed in *her*⁻ individuals.

1.2.1.3.6 groucho

groucho, initially reported as being a maternal effect gene with a role in neurogenesis (Hartley *et al.*, 1988; Delidakis *et al.*, 1991; Schrons *et al.*, 1992) has recently been implicated in the regulation of *Sex-lethal* activity (Paroush *et al.*, 1994). Using a yeast two-hybrid system, Paroush *et al.* have shown that GROUCHO protein complexes with HAIRY protein, as well as the bHLH proteins DEADPAN and those derived from the *Enhancer of split-C* locus. Additionally, in embryos derived from mothers with *gro*$^{-}$ germ cells, *Sxl* is seen to be mis-expressed in haplo-X (i.e. male) embryos. Female-specific SXL protein is detected in all embryos analysed (whereas it is normally restricted to diplo-X individuals), and is apparent by blastoderm cycle 14. This suggests *gro* functions prior to this stage, and is likely to be involved in the X/A ratio dependent regulation of *Sxl* expression. How *gro* may be involved in this regulation is discussed in section 1.2.2.1.

1.2.2 ZYGOTIC SEX REALISER LOCI

1.2.2.1 Sex-lethal

The X-chromosomal gene *Sex-lethal* (*Sxl*; 1-19.2) was first shown to have dramatic effects on sexual differentiation by Muller and Zimmering in 1960. Strong loss of function mutations (e.g. $Sxl^{F\#1}$) are lethal to diplo-X embryos and kill during embryogenesis. Reciprocally, strong gain of function mutations (e.g. $Sxl^{M\#1}$) induce male lethality during early larval development (Cline, 1978; Marshal and Whittle, 1978).

Extensive characterisation has revealed that *Sxl* plays a critical and divergent role in *Drosophila* development. Not only does it regulate sexual differentiation in response to the X/A balance signal, but it is also crucial for dosage compensation (Cline, 1978, 1979a, 1979b, 1984, 1993; Lucchesi and Skripsky, 1981; Sánchez and Nöthiger, 1982; Gergen, 1987; Bopp *et al.*, 1991). Indeed, the lethality associated with *Sxl* mutations is a consequence of improper dosage compensation and not a result of gross morphological defects.

As discussed in section 1.2.1, both maternal and zygotic factors assess the chromosomal constitution of the developing embryo. How exactly the X/A balance ratio is represented molecularly is unknown, although it seems likely that the presence of numerator/DA heterodimers will constitute an activation signal. This hypothesis is supported by the observation that *daughterless* and *sis-b* proteins interact *in-vitro* to produce heterodimers exhibiting DNA-binding activity (Murre *et al.*, 1989b). An *in vivo* interaction between SIS-B and DA proteins has also recently been reported in a yeast two-hybrid system (Liu and Belote, 1995; Deshpadne *et al.*, 1995).

Genetic and molecular studies suggest that numerator element proteins (particularly SIS-B) interact with maternally derived *daughterless* protein to form heterodimers capable of activating transcription at the *Sxl* 'early' promoter (and other

loci during neurogenesis etc.). This interaction, initially, was thought to be hindered in male (1X:2A) embryos by association of *sis-b* and *deadpan* (denominator) derived proteins, such that SIS-B/DA heterodimers were unable to form. This hypothesis however now appears to be unlikely, since recent reports demonstrate that SIS-B and DPN do not interact with one another either *in-vitro* or *in-vivo* (Liu and Belote, 1995; Deshpadne *et al.*, 1995). Indeed, these authors show that SIS-A/DPN complexes form readily *in-vivo*. Additionally, Paroush *et al.* (1994) report that protein derived from another maternal effect gene, *groucho*, complexes with DPN proteins *in vitro*, and is critical for repression of *Sxl* activation in haplo-X individuals.

These data appear to correlate well with a model whereby GRO/DPN heterodimers repress *Sxl* transcription at the 'early' promoter by steric hindrance of SIS-B/DA heterodimers. In diplo-X individuals, increased SIS-A concentration would result in the preferential formation of SIS-A/DPN heterodimers as opposed to GRO/DPN complexes. This, in conjunction with elevated SIS-B/DA concentrations, would result in a vastly increased probability of *Sxl* activation. This model predicts that GRO/DPN complexes would have a higher affinity for a shared promoter binding site than SIS-B/DA heterodimers. It is not known at present, however, whether GRO/DPN heterodimers have DNA-binding activity. Gel retardation assays could provide some interesting preliminary answers to this question, and to those regarding the relative affinities of the complexes discussed here to *Sxl* promoter elements.

The *Sxl* region has now been cloned (Maine *et al.*, 1985a), and spans approximately 23Kb. Northern analysis reveals a complex array of overlapping transcripts (Salz *et al.*, 1989), which are expressed in a stage-, tissue- and sex-specific manner. A complete description of the major transcripts derived from *Sxl* is shown in Figure 1.3, along with schematic diagrams illustrating the complex splicing associated with this locus (derived from Bell *et al.*, 1988; Salz *et al.*, 1989; and Keyes *et al.*, 1992). One intriguing aspect of *Sxl* activity is that once initiated, expression becomes independent of the X/A balance signal (Cline, 1984; Maine *et al.*, 1985). Indeed, ectopic expression of SXL^F (using a heat shock transgene) can *trans*-activate an endogenous wild type *Sxl* gene (Bell *et al.*, 1991). These results are consistent with the observation that in somatic tissue, sex is determined in a cell autonomous manner and is established irreversibly early in development.

Analysis of cDNAs derived from adult males and females reveals the sex-specific nature of *Sxl* activity is attributable to the inclusion of a male-specific exon in male transcripts. This male-specific exon introduces several in-frame stop codons in an otherwise long open reading frame, and translation therefore yields a truncated and presumably non-functional protein of only 48 amino acids (aa). Female transcripts, in contrast, encode a predicted protein of some 354aa with extensive homology (in two principle domains) to a conserved RNA binding motif found in many ribonucleoproteins (Bell *et al.*, 1988), many of which act as developmental regulators (Bandziulis *et al.*,

1989). Therefore, in a simplistic sense we can view the activity of *Sxl* as being 'ON' in females and 'OFF' in males.

The male-specific exon, accounting for some 190bp, is consistent with the relative difference in all adult transcript sizes derived from males and females. That the inclusion of this exon is necessary for male development is supported by the observation that $Sxl^{M\#1}$ - $Sxl^{M\#5}$ (constitutive, hypermorphic male-lethal alleles) all contain transposon insertions within 1kb of this exon (Maine *et al.*, 1985b). Indeed, these alleles have since been shown to represent three independent mutations, all of which affect the sex-specific splicing of transcripts derived from *Sxl* (Bernstein *et al.*, 1995).

The presence of a putative RNA binding domain led Bell *et al.* (1988) to propose that SXL^F would interact with transcripts derived from both downstream genes (notably *transformer*), and the *Sxl* locus itself. In both cases, this prediction has been confirmed. In the case of *transformer* (*tra*), SXL^F blocks the utilisation of a default 3' splice acceptor site such that a weaker secondary 3' splice acceptor is employed (refer to section 1.2.2.2). Initially, a similar model was proposed for the regulation of *Sxl* transcript splicing, but this process is now known to be more complex. What function the 'early' transcripts may play in the complex activity of *Sxl*, and how expression is maintained in each sex is discussed in section 1.2.2.1.1.

Consistent with the proposed role of *Sxl* in RNA processing, SXL^F protein is preferentially localised to the nucleus, and is perturbed in $sis\text{-}b^-$ embryos and those derived from da^- mothers (Bopp *et al.*, 1991).

1.2.2.1.1 Activation and maintenance of *Sxl* expression

The presence of both 'early' and 'late' transcripts derived from *Sxl* was initially intriguing since it suggests an intrinsic difference in the relative activities of the encoded proteins. Particularly, the transient expression of the 'early' transcripts coincides with the stage at which assessors of the X/A balance signal are known to influence the determination of sexual identity. Keyes *et al.* (1992) have since shown that the appearance of these transcripts is in fact dependant on these factors. In both embryos derived from da^- mothers, and those lacking zygotic $sisterless^+$ activity, these transcripts are present at greatly reduced levels. cDNAs representative of the three 'early' transcripts were isolated, and revealed that they contained novel 5' sequences relative to the adult transcripts. Essentially an early exon (E1) is present instead of the late exons (L1, L2, L3 [male-specific]) found in adult transcripts (see Figure 1.3). This exon is located between L1 and L2 in genomic DNA, and is included in these transcripts due to initiation derived from a promoter located 5kb downstream of that used in later development. Presumably, this early promoter (P_E) is only utilised in response to the X/A balance signal, whereas the late promoter (P_L) is activated in both sexes by general transcriptional activators throughout the remainder of development.

In these 'early' transcripts, splicing from E1 occurs in frame to late exon 4, bypassing the male-specific exon (L3). Thus all 'early' transcripts encode female-specific *Sxl* protein (SXL^F), and are only present in diplo-X embryos. The significance of these transcripts only becomes apparent when we consider the autoregulatory activity of *Sxl*, as revealed by the analysis of proteins encoded by the adult 'late' transcripts. Sakamoto *et al.* (1992) demonstrated SXL^F regulates splicing of *Sxl* hnRNA to maintain female-specific expression (using co-transfection experiments in Kc cultured cells). Further analysis suggests that similar to the regulation of *tra* splicing, a blockage mechanism is operating in females to prevent the inclusion of the male-specific exon (L3) from *Sxl* transcripts in females. However, in contrast to *tra* (where SXL^F blocks the use of a default 3' splice acceptor), both introns either side of L3 are required for efficient excision (Horabin and Schedl, 1993a, 1993b; Wang and Bell, 1994). Also, the intron between exons L3 and 4 appears to be more crucial than the intron between exons L2 and L3, implicating a 5' splice donor site mediated regulation.

Consistent with the results obtained with *tra*, uridine repeat elements (typically poly-U_8 sequences) are required for SXL^F regulation. Gel shift experiments with such repeat elements (both singular and as doublets separated by a short spacer) suggest SXL^F may form heterodimers, and that the amino-acyl end of the protein is important for this interaction (Wang and Bell, 1994). Also, Samuels *et al.* (1994) have shown that SXL^F associates with large ribonucleoprotein complexes, suggesting complex formation may be intrinsic to some aspects of SXL^F activity.

Once SXL^F is produced, it actively promotes the exclusion of the male-specific exon (L3) from *Sxl* transcripts and maintains *Sxl* expression in a female-specific mode. Thus the 'early' transcripts (which do not require the removal of L3) encoding constitutive SXL^F lock *Sxl* expression in a female mode, and this state is maintained throughout the remainder of development by autoregulation. How other factors such as those encoded by *snf* and *fl(2)d* interact with SXL^F to regulate this splicing is unknown, although they do appear to be critical (see sections 1.2.1.3.2 and 1.2.1.3.3).

A schematic diagram summarising the activation and maintenance of *Sxl* activity is shown in Figure 1.4.

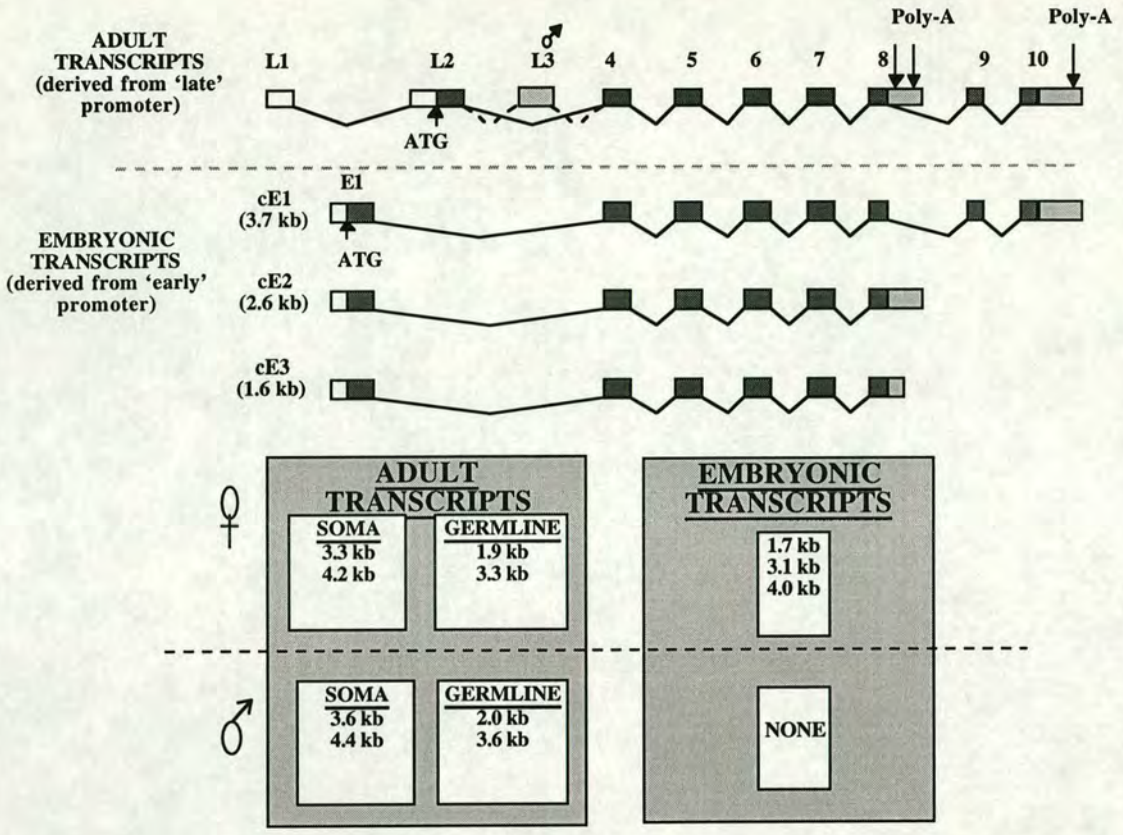


Figure 1.3. Alternative splicing of *Sxl* transcripts and relevant transcript sizes. Exons L1, L2 and L3 are included in transcripts derived from the 'late' promoter. Exon L3 is male-specific and introduces in-frame stop codons. Exon E1 is only included in transcripts derived from the 'early' promoter. cE1, cE2 and cE3 refer to isolated cDNAs.

1.2.2.1.2 THE ROLE OF *SEX-LETHAL* IN DEVELOPMENT

1.2.2.1.2.1 Somatic sex determination

We have seen that *Sxl* is activated divergently in diplo-X and haplo-X individuals, but what is the consequence of this differential expression on development? Classical genetic experiments have revealed interactions with several other genes known to induce sexual transformation when mutated. In particular, *transformer (tra)* was shown to be epistatic to *Sxl*, whereas *doublesex (dsx)* is epistatic to both of these loci (Baker and Ridge, 1980; Nöthiger *et al.*, 1987; McKeown *et al.*, 1988). Figures 1.2A and 1.2B show schematic representations of the somatic sex determination hierarchy in *Drosophila*, and illustrate the epistatic relationships deduced for relevant loci.

Principally, in females, SXL^F actively regulates the splicing of *tra* hnRNA such that a female-specific transcript is produced (encoding TRA^F). This protein, in conjunction with *transformer*-2 protein, regulate splicing of transcripts derived from the *dsx* locus in a sex-specific manner (section 1.2.2.4.1.; for reviews see Baker and Belote, 1983; Slee and Bownes, 1990, Belote, 1992; Burtis and Wolfner, 1992; and Burtis, 1993). Therefore, in females, this cascade ultimately results in *doublesex* female-specific protein (DSX^F). In males, where all transcripts are spliced in a default manner (and thus SXL^F and

TRA^F are never produced), male-specific *doublesex* protein (DSX^M) is generated. The proteins derived from *dsx* are proposed to directly regulate the majority of genes involved in sex-specific differentiation. Please refer to sections 1.2.2.2, 1.2.2.3, and 1.2.2.4 for more detailed discussion of *tra*, *tra-2* and *dsx*. The involvement of *Sxl* in germ line sex determination, as opposed to somatic sex determination, is discussed in section 1.3.1 since much less is known about this aspect of *Sxl* function.

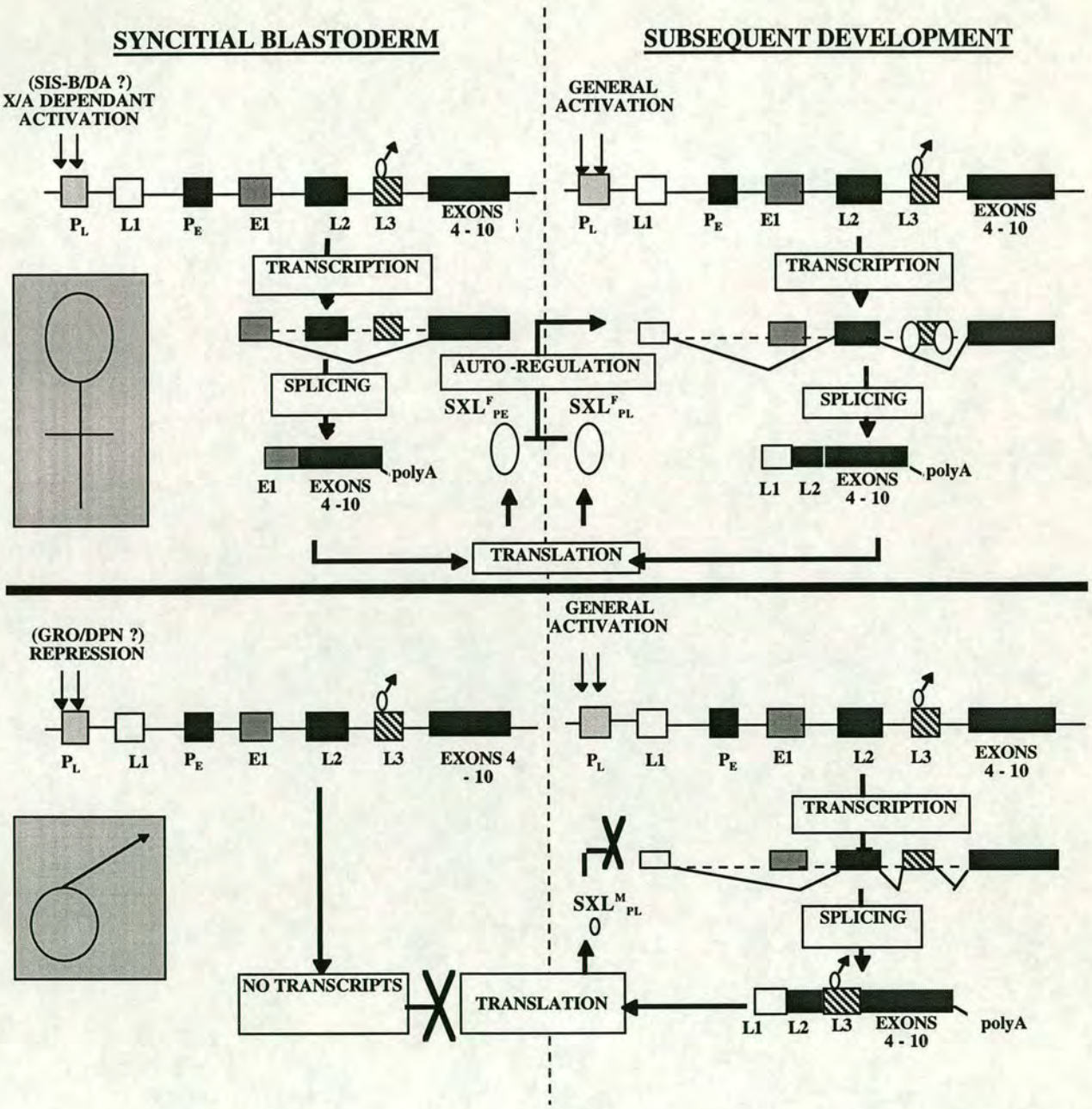


Figure 1.4. Activation and maintenance of *Sxl* expression throughout development. Exon E1 is present only in transcripts derived from the 'early' promoter (P_E). Exons L1, L2 and L3 are included in transcripts derived from the 'late' promoter (P_L). L3 is only present in transcripts spliced in the male-default mode and is therefore male-specific, and includes in frame stop codons which truncate the open reading frame. Proteins are indicated as being either male-specific (SXL^M) or female-specific (SXL^F) , and the relevant promoter utilised in its production is indicated as subscript (e.g. SXL^F_{PE} is female-specific protein derived from the 'early' promoter). Exons 4-10 are shown grouped together for simplicity, though the correct organisation is shown in Figure 1.3.

1.2.2.1.2.2 Dosage compensation

A second critical role for *Sxl* in development is the regulation of dosage compensation, which in *Drosophila* is dependant on hypertranscription of the X chromosome in haplo-X individuals (see Baker and Belote [1983] and references therein). Evidence that *Sxl* is involved in the repression of hypertranscription in diplo-X individuals comes from the analysis of transcription rates of several X-linked loci in *Sxl* mutants (Lucchesi and Skripsky, 1981; Gergen, 1987; Bernstein and Cline, 1994).

Critical to the process of dosage compensation, the genes collectively known as the '*male specific lethals*' (*msl's*) appear to be involved in marking the X chromosome for hypertranscription. The *maleless (mle)* mutation, originally isolated from a natural population, was reported by Fukunaga *et al.* (1975) as being a zygotic male-lethal which killed subsequent to the 3rd larval instar. In an EMS induced screen for autosomally encoded *male-specific lethal* mutations, Belote and Lucchesi (1980) also identified two new loci, *male-specific lethal-1 (msl-1)* and *male-specific lethal-2 (msl-2)*. The *mle*, *msl-1*, and *msl-2* genes have now been cloned (Kuroda *et al.*, 1991; Palmer *et al.*, 1993; Zhou *et al.*, 1995; Kelley *et al.*, 1995). Sequence analysis reveals homolgy between *mle* and RNA helicase proteins, *msl-1* encodes a novel polypeptide, whereas *msl-2* encodes a putative ring finger protein. A fourth gene, *male-specific lethal-3* has recently been isolated by Gorman *et al.* (1995).

In all cases, mutations in any of these *msl* genes result in male-lethality (death ensues in late larval/ early pupal development), while females remain unaffected. These proteins appear to be interdependent, since MLE, MSL-1 and MSL-3 require each other, and MSL-2 protein, for association with the X chromosome in haplo-X individuals (Gorman *et al.*, 1993; Hilfiker *et al.*, 1994). Additionally, acetylated histone H4 (H4Ac16) co-localises with MLE and MSL-1 proteins on the male X chromosome (Turner *et al.*, 1992; Bone *et al.*, 1994), and this acetylation is dependant on these proteins (Hilfiker *et al.*, 1994).

Of these four genes, only *msl-2* appears to be regulated in a sex-specific manner, although depressed *msl-1* transcript levels are apparent in females relative to males (Kelley *et al.*, 1995; Zhou *et al.*, 1995). In females, transcripts derived from *msl-2* are larger than those in males due to differential splicing of an 132bp intron in the untranslated leader sequence. This appears to affect expression, since MSL-2 protein is only detected in male individuals, where it is seen to be physically associated with the X-chromosome. Direct splicing regulation by SXL^{F} is inferred by the presence of poly-U sequences (already implicated in the regulation of *Sxl* and *tra* splicing) around the splice junctions of this intron (Zhou *et al.*, 1995). Consistent with the hypothesis that *msl-2* is a target of direct *Sxl* repression, ectopic expression of an *msl-2* transgene (male-specific splicing variant)

results in association of MLE and MSL-1 proteins with the X-chromosome in diplo-X individuals.

The finding that ectopic expression of MSL-2 causes the association of other MSL proteins on the X-chromosome suggests this protein may be the co-ordinator of complex formation. Additionally, since histone-H4 acetylation is dependant on these proteins, the formation of MSL complexes on the X-chromosome may be directly related to the acetylation of this residue by stabilising other factors at these sites.

1.2.2.2 transformer

Null mutations at the *transformer (tra)* locus, when homozygous, transform diplo-X (i.e. female) individuals into pseudomales but have no effect on germ line sex determination or male differentiation (Sturtevant, 1945; Brown and King, 1961). The *tra* locus has now been cloned (Butler *et al.*, 1986; McKeown *et al.*, 1987), and produces both female-specific (0.9kb) and non sex-specific (1.1kb) transcripts. These transcripts are most abundant during pupation (the stage at which the most overt sexual differentiation is occurring), consistent with observations suggesting *tra* is required during this period for sex determination (Baker and Ridge, 1980; Wieschaus and Nöthiger, 1982).

The presence of a non sex-specific transcript was initially intriguing, since *tra* appears to have no function in males. However, Boggs *et al.* (1987) have demonstrated that both female- and non sex-specific transcripts derive from a single hnRNA species, and arise by differential splicing of the first intron. Essentially, splicing removes a 78 base pair sequence to generate the non sex-specific transcript, whereas the female-specific transcript has a 248bp sequence removed. Differential splicing is now known to arise as a consequence of direct *Sxl* regulation, whereby SXL^F binds at a polypyrimidine tract adjacent to the default 3' splice acceptor site of the first intron. This binding inhibits the association of U2AF (an essential splicing factor) at this site, and a weaker 3' splice acceptor consensus site located 175bp downstream is utilised instead (Nagoshi *et al.*, 1988; Sosnowski *et al.*, 1989; Inoue *et al.*, 1990; Valcárcel *et al.*, 1993; Sosnowski *et al.*, 1994).

The female-specific transcript encodes a predicted protein of approximately 22kDa, due to the presence of a single long open reading frame. The non sex-specific transcript (where splicing takes place at the default splice acceptor) contains stop codons in all three reading frames, and presumably generates a truncated and therefore non-functional protein (Boggs *et al.*, 1987).

A model for the regulation of *tra* splicing is presented in Figure 1.5.

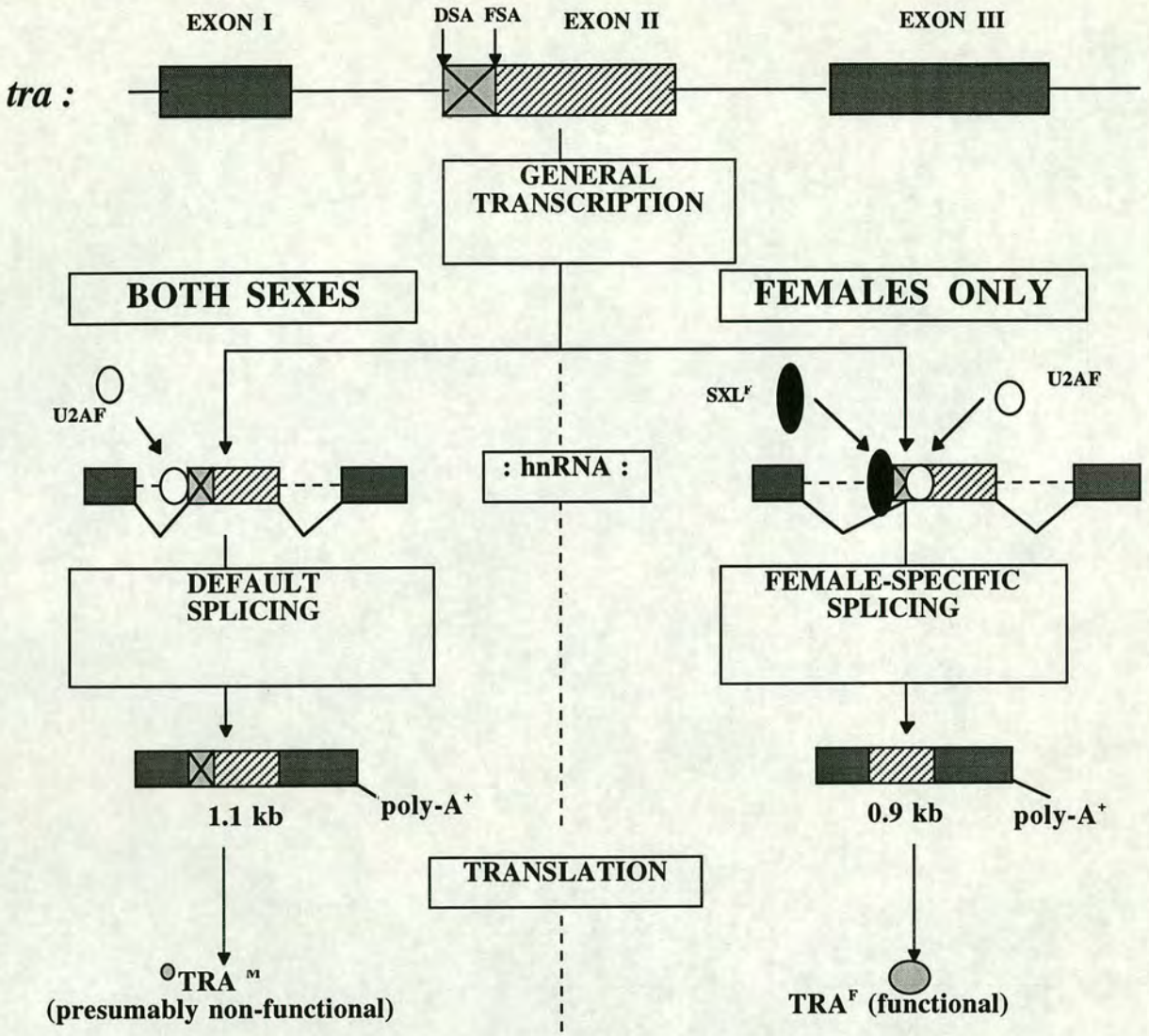


Figure 1.5. Regulation of sex-specific alternative splicing of transcripts derived from the *transformer* locus. DSA refers to the default 3' splice acceptor site, whereas FSA refers to the female-specific 3' splice acceptor site. TRA^M and TRA^F refer to the male- and female-specific *transformer* derived proteins respectively. SXL^F is female-specific SXL protein. The 'X' associated with Exon II (default splicing) represents an in frame termination.

1.2.2.3 transformer-2

An involvement of the autosomal *transformer-2 (tra-2)* locus in somatic sex determination is inferred since diplo-X individuals homozygous for a null *tra-2* allele develop as pseudomales, although these flies are sterile and only contain rudimentary gonads (Watanabe, 1975; Fujihara *et al.*, 1978, Baker and Ridge, 1980). Homozygous haplo-X individuals, in contrast, remain morphologically male but are sterile.

The presence of a male-fertile allele, *tra-2*^{OTF}, suggests the sterility associated with previous *tra-2* alleles is not a consequence of pleiotropy, but rather a differential role in both the soma and the germline (Fujihara *et al.*, 1978). Indeed, both pole cell transplantation (Schüpbach, 1982) and temperature shift experiments (Belote and Baker,

1982, 1983) reveal that *tra-2* is required at several stages during development for correct female differentiation, as well as spermatogenesis. In *tra-2^{ts}/ tra-2* heterozygotes the early stages of spermatogenesis appear normal, but later stages are defective (at 18°C and 29°C; Belote and Baker, 1983). Particularly, a frequent failure in sperm head elongation is observed, as well as a more dissociated nature than is observed in wild type individuals. Belote and Baker (1983) propose that these spermatogenic defects may arise due to failure of X-chromosomal inactivation in the male germline.

With respect to female sexual-differentiation, the temperature sensitive period (TSP) of *tra-2* extends from the second larval instar to the early middle pupal period. Also, a reduction in the relative dose of tra in individuals homozygous for *tra-2^{ts}* results in transformation to male morphology even at the normal permissive temperature of 16°C (Belote and Baker, 1982), suggesting an interaction between these two loci necessary for female differentiation.

The *tra-2* gene has now been cloned (Amrein *et al.*, 1988; Goralski *et al.*, 1989), and transcription generates complex multiple overlapping tissue-specific transcripts due to variable transcription initiation sites and alternative splicing (Amrein *et al.*, 1990; Mattox *et al.*, 1990; Mattox and Baker, 1991). Figure 1.6 shows the predicted organisation and localisation of transcripts identified to date. Two transcripts are non sex-specific, and differ in abundance and the presence (T_{min}) or absence (T_{maj}) of exon 3. The remaining two transcripts (msT_{maj} and msT_{min}) are male germline specific.

The major male germline specific transcript (msT_{maj}) was initially thought to encode the *tra-2* function required for spermatogenesis, but this can not be the case since *tra-2*$^{-}$ males carrying an msT_{maj} transgenic cDNA do not show restored viability (Amrein *et al.*, 1990). Intriguingly, this construct is able to partially rescue the sexual transformation observed in *tra-2^{ts}/ tra-2^{ts}* diplo-X individuals. It is clear however that T_{maj} can provide all *tra-2* function required for correct female differentiation and male fertility.

The transcripts derived from the upstream promoter (T_{maj} and T_{min}) encode polypeptides of 264 and 226 amino acids respectively, whereas msT_{maj} and msT_{min} transcripts derive from transcription initiated between exons 2 and 3, and encode polypeptides of 179 and 226 amino acids respectively. The msT_{min} and T_{min} encoded proteins are presumed to be identical. All proteins contain an arginine-serine rich domain and an RNA recognition motif similar to those found in several RNA-binding proteins, including SXL.

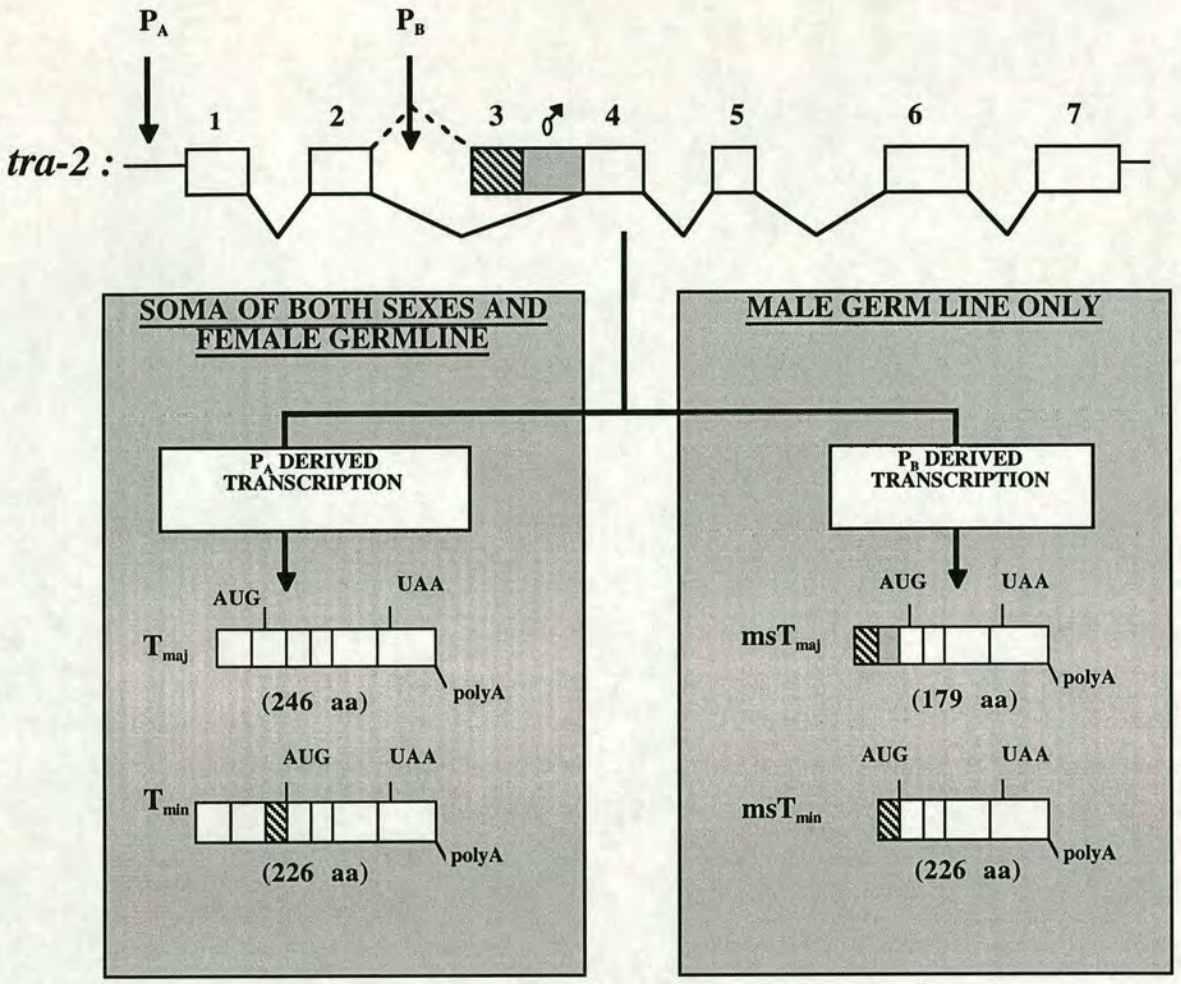


Figure 1.6. Alternative splicing of transcripts derived from *transformer-2* and their tissue-specific localisation. P_A and P_B refer to two different transcription initiation sites. The prefix 'ms' refers to the male-specific nature of these transcripts. Subscript of 'maj' or 'min' indicates major and minor transcripts respectively.

One hypothesis explaining the observation that an msT_{maj} transgenic cDNA can rescue the masculinising effects of *tra-2*^{ts} in diplo-X individuals is that *tra-2* has autoregulatory activity, and that protein derived from this construct is *trans*-activating the endogenous *tra-2* allele. To test this hypothesis, Mattox and Baker (1991) transformed an msT_{maj}-*LacZ* reporter gene fusion construct into *tra-2*^{B} / + and *tra-2*^{B} / *tra-2*^{B} flies (*tra-2*^{B} flies are deficient of msT_{maj} transcripts). In this construct, splicing which generates an msT_{maj} leader results in *LacZ* activity, whereas other splicing variants lead to frame shift mutations. Consistent with an autoregulatory function for *tra-2*, *LacZ* activity is only observed in *tra-2*^{B} / + individuals, suggesting msT_{maj} encoded protein is required to maintain its own expression. However, when a *tra-2* allele only capable of producing msT_{maj} transcripts is transformed into *tra-2*^{B} / *tra-2*^{B} individuals, no accumulation of msT_{maj} transcripts is observed. This suggests that proteins encoded by the other *tra-2* transcripts are involved in this autoregulatory activity. Mutations in the splice sites

flanking the M1 intron also disrupt the accumulation of msT_{maj} transcripts, suggesting autoregulation is a consequence of splicing regulation rather than transcript stability.

Mattox and Baker (1991) suggest the msT_{min} encoded polypeptide is likely to be responsible for the accumulation of msT_{maj} transcripts, since it is expressed in the male germline and elevated msT_{min} transcript levels are observed in flies showing reduced msT_{maj} transcript abundance levels (e.g. in *tra-2*^{B} homozygotes). Also, since the msT_{min} and T_{min} transcripts are presumed to encode the same polypeptide, T_{min} encoded protein would presumably have similar autoregulatory function in the female germline and soma.

1.2.2.4 doublesex

Hildreth (1965) reported that null mutations at the *doublesex (dsx)* locus transform both diplo-X and haplo-X individuals into intersexes in which individual cells are seen to express both male and female characteristics simultaneously (see Baker and Belote, 1983 for review). Since such mutations affect both sexes it can be inferred that *dsx* functions in both males and females.

The localisation of *dsx* to salivary gland chromosome band 84E1-2 (Belote *et al.*, 1985) was facilitated by the existence of two dominant alleles (*dsx*^{D} and *dsx*^{Mas}; both transform diplo-X individuals into morphological males). Further characterisation using the position of breakpoints and chromosomal rearrangements, as well as an extensive 107kb chromosome walk, has delimited the *dsx* locus to approximately 40kb (Baker and Wolfner, 1988). Northern analysis reveals a complex temporal array of transcripts, some of which are expressed sex-specifically. In larvae, non sex-specific transcripts of 2.8kb and 1.65kb are enriched, whereas the sex-specific transcripts (3.9kb in males, 3.5kb in females) are present at lower abundance. During pupation, this expression profile is reversed such that the sex-specific transcripts predominate. Adults express lower levels of sex-specific transcripts (analogous to the levels observed in larvae), and males also express an additional transcript of 2.9kb which differs from the larger 3.9kb transcript as a consequence of alternative polyadenylation sites. The non sex-specific transcripts enriched in larvae (2.8kb and 1.65kb) are not detected in adults (Baker and Wolfner, 1988; Burtis and Baker, 1989).

Burtis and Baker (1989) isolated cDNAs corresponding to the sex-specific transcripts derived from *dsx*, and demonstrated that they contain common 5' sequences, but differ at their 3' termini due to the presence of sex-specific exons. The first three exons are common to both male and female transcripts, but in females a single female-specific exon is then found, in contrast to the two male-specific exons evident in males. Thus, the encoded proteins comprise a common N-terminal 397aa region, and an additional sex-specific C-terminus of either 30aa (females) or 152aa (males). The predicted polypeptides are therefore 44.8kDa (female, DSX^{F}) and 57.4kDa (male, DSX^{M}) respectively.

Initial sequence comparisons (Burtis and Baker, 1989) failed to reveal any significant homology to known DNA-binding proteins, as would be expected if these proteins regulate downstream differentiation gene expression as predicted. However, using a combination of deletions and gel retardation assays, an atypical zinc finger related DNA binding domain has been localised to the common N-terminal portion of both proteins, in exon 2 (Erdman and Burtis, 1993). The importance of this domain in *doublesex* function is confirmed by the observation that several *dsx* null alleles are a consequence of missense mutations in this region, which abolish DNA binding activity (associated with a reduction in the level of associated zinc). The organisation of these related proteins suggest they may bind the same regulatory target sequences, but differ in function as a consequence of their sex-specific C-termini.

1.2.2.4.1 The regulation of *doublesex* splicing

Several lines of evidence suggest that *tra* and *tra-2* are involved in the regulation of sex-specific alternative splicing of transcripts derived from *dsx*. Genetic studies show that *dsx* is epistatic to both *Sxl, tra* and *tra-2* (Baker and Ridge, 1980; Nöthiger *et al.*, 1987; McKeown *et al.*, 1988), and molecular studies demonstrate that in either tra^- or $tra\text{-}2^-$ individuals *dsx* transcripts are spliced in a default (male) mode (Nagoshi *et al.*, 1988). Consistent with the hypothesis that *Sxl* (the primary sex-determining gene) regulates the expression of *tra*, which in turn regulates *dsx* expression, diplo-X individuals homozygous for a hypomorphic *Sxl* allele (Sxl^{2593}) produce both male- and female-specific *dsx* transcripts. Temperature shift experiments using a $tra\text{-}2^{ts}$ allele demonstrate that *dsx* expression is continuously regulated by *tra-2* (and by inference *tra*), and is therefore not irreversibly determined early in embryogenesis, as is the case for *Sxl*.

Nagoshi and Baker (1990) proposed that *tra* and *tra-2* directly regulate sex-specific alternative splicing of *dsx* hnRNA in diplo-X individuals such that female-specific transcripts are produced. In haplo-X individuals, which lack functional TRA protein, splicing occurs in a default (male-specific) manner. These authors also demonstrate that in diplo-X individuals hemizygous for four *dsx* dominant alleles (i.e. the allele over a deficiency of the *dsx* region), only male-specific *dsx* transcripts are ever produced. These mutations were further shown to result from chromosomal rearrangements in the vicinity of the female-specific splice acceptor, suggesting that they perturb the association of *trans*-acting factors in this region. Interestingly, diplo-X individuals heterozygous for such dominant *dsx* alleles (e.g. $dsx^M/+$) display an intersexual phenotype analogous to that observed in *dsx* null homozygotes, suggesting DSX^F and DSX^M proteins are mutually antagonistic. This is consistent with the observations of Erdman and Burtis (1993), who report that DSX^F and DSX^M contain an identical DNA binding domain, and may interact with the same target regulatory sequences.

The chromosomal rearrangements associated with all of the *dsx* dominant alleles analysed by Nagoshi and Baker (1990) all displace or delete a region of the female-

specific exon containing six repeats of an 13bp element. Tissue culture experiments demonstrate that both *tra* and *tra-2* encoded proteins are required to actively promote female-specific splicing of *dsx* hnRNA, and that the efficiency of this splicing correlates quantitatively with the presence of these 13bp repeat elements (Hoshijima *et al.*, 1991; Ryner and Baker, 1991). Furthermore, Hedley and Maniatis (1991) report that TRA-2 protein binds specifically to this 13bp repeat element, and that this element is also required for female-specific polyadenylation of transcripts spliced in a female-specific manner. Recent evidence (Tian and Maniatis, 1993; Heinrichs and Baker, 1995) suggests that TRA and TRA-2 proteins complex at this repeat element and stabilise the association of general splicing factors (such as the SR protein RBP1) in this region, thus enhancing the utilisation of the female-specific 3' splice acceptor.

The organisation of transcripts derived from the *doublesex* locus, and the model of their sex-specific alternative splicing regulation, is illustrated in figure 1.6.

1.2.2.5 intersex

intersex (ix) is implicated in somatic sex determination in *Drosophila* since the null *ix* phenotype specifically transforms diplo-X individuals into intersexes, whereas haplo-X individuals are indistinguishable from wild type (Baker and Ridge, 1980, and references therein; Chase and Baker, 1995). This transformation is morphologically similar to that observed in *dsx* mutants, although *ix* mutants are masculinised less extensively. Genetic studies (Baker and Ridge, 1980) suggest that *ix* encodes a protein necessary for the repression of male differentiation in diplo-X individuals, and acts either in parallel to, in concert with, or downstream of *dsx*. Molecular studies confirm this, since *ix* mutations have no effect on the sex-specific splicing of transcripts derived from *dsx* (Nagoshi *et al.*, 1988).

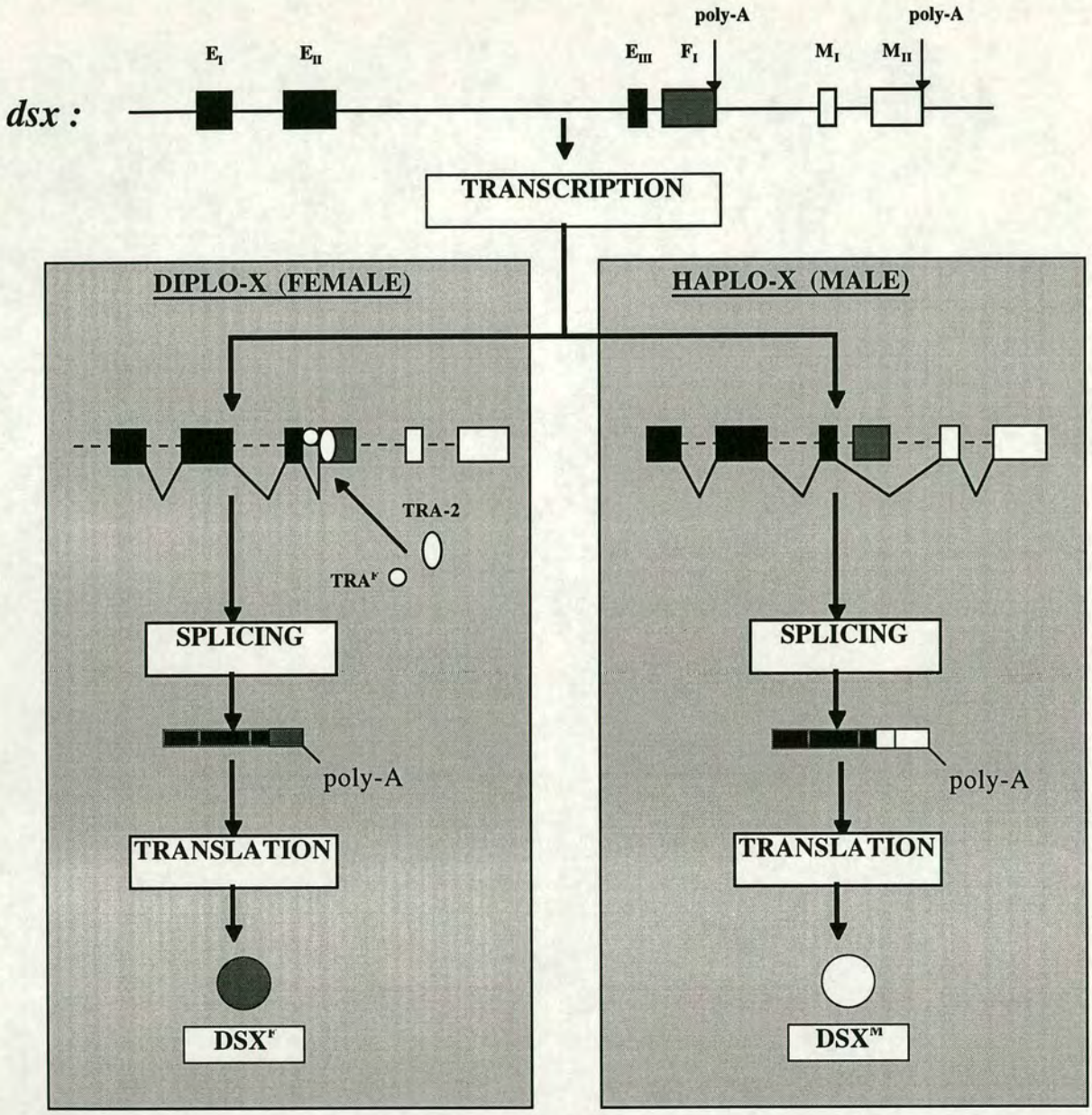


Figure 1.7. The organisation of the *doublesex* locus and the sex-specific alternative splicing of transcripts derived from it. TRA^F represents the female-specific protein derived from the *transformer* locus (*Sxl* needs to be in an 'ON' state for this protein to be produced). TRA-2 protein derives from the *transformer-2* locus. DSX^F and DSX^M refer to female- and male-specific *doublesex* proteins respectively. The zinc finger related DNA binding domain of the DSX proteins is derived from exon II. E_I to E_{III} refers to exons I to III. F_I is the female-specific exon, whereas M_I and M_{II} are male-specific exons.

1.2.3 THE INFLUENCE OF SEX DETERMINING LOCI ON DEVELOPMENT

We have seen how the sex-specific expression of several zygotic sex-determining loci is interrelated, but what effect does this have on the expression of downstream differentiation genes? Do all of the sex-determining genes influence such genes, or does one influence the expression of downstream genes more than any other ?

The extensive transformations observed in hypermorphic *dsx* mutants (e.g. dsx^F or dsx^M) suggests that the majority of differentiation is regulated by *dsx* encoded proteins. The presence of a DNA binding domain in both DSX^F and DSX^M further reinforces the idea that these proteins may act as transcriptional regulators. To date however, only the *yolk protein* genes have been shown to be directly regulated by *dsx* (see section 1.5.3.1.1.2), although direct regulation of glucose dehydrogenase expression (*Gld*) in the reproductive tract is suspected (Feng *et al.*, 1991). Many other genes are expressed differentially in males and females, but appear to be dependent on tissue-specific factors (which in turn are dependent on the sex of the tissue), and are not therefore likely to be regulated directly by *dsx* (see review by Burtis and Wolfner, 1992).

In contrast to the initial model proposed for *dsx* activity, whereby DSX^F and DSX^M act in a mutually antagonistic manner and repress differentiation of tissues found in the opposite sex (i.e. DSX^F represses male differentiation, thus allowing female differentiation, and vice-versa), recent evidence suggests these proteins may also act in a positive manner. For example, ectopic expression of DSX^M can induce partial transformation of bristles on all six legs to a more sex-comb like (male-specific) morphology. However, these results must be viewed with some scepticism since dsx^M hypermorphs do not show such transformations, suggesting this result is a consequence of the atypical ectopic expression. More conclusively, Taylor and Truman (1992) report that the male-specific division of abdominal neuroblasts is dependent on the presence of DSX^M, rather than the absence of DSX^F.

However, not all aspects of differentiation are regulated by *dsx*. In diplo-X individuals mutant at *Sxl, tra* or *tra-2*, male courtship behaviour is observed, whereas mutations at *dsx* or *ix* induce no such effects (McRobert and Tompkins, 1985; Tompkins and McRobert, 1989). Also, differentiation of the male-specific muscle (also known as the muscle of Lawrence), located in the 5^{th} abdominal segment in males is dependent on the sex of the innervating axons and is not affected by *dsx* mutations (Taylor, 1992; Currie and Bate, 1995). This muscle however does differentiate in diplo-X individuals deficient of *tra* or *tra-2* function, suggesting these genes are involved in determining the sex of the innervating axons.

Thus despite the observation that *dsx* controls the majority of differentiation, some aspects of this complex process are regulated by genes acting above *dsx* in the sex-determination hierarchy. Although it is likely that *Sxl* negatively regulates *msl-2* activity, and could theoretically regulate some aspects of differentiation, this seems

unlikely. This conclusion is drawn from the results of temperature shift experiments, which demonstrate that all aspects of adult somatic sexual differentiation can be related to *tra-2* (and by inference *tra*) activity(see Burtis and Wolfner, 1992, and references therein).

1.2.4 SUMMARY OF SOMATIC SEX-DETERMINATION IN *DROSOPHILA*

We have seen how somatic sex-determination in *Drosophila melanogaster* is defined at the level of transcription and transduced at the level of alternative splicing. Transcription of the primary sex-determining gene, *Sex-lethal*, at the syncitial blastoderm stage of embryogenesis only occurs in diplo-X individuals due to X/A ratio dependent activation. This 'early' burst of transcripts encodes constitutive female-specific SXL protein (SXL^F). Throughout subsequent development, transcription initiated at a non X/A ratio dependent *Sxl* promoter generates transcripts which are spliced in a default (male) mode unless SXL^F is present, in which case female-specific splicing occurs. Thus in either case *Sxl* expression is locked in one of two modes of expression, male- or female-specific.

Similarly, expression of the two major downstream sex-determination genes, *transformer* and *doublesex*, is regulated by alternative splicing. *transformer* female-specific splicing is dependent on SXL^F, and subsequently *doublesex* female-specific splicing is dependent on TRA^F. In males, splicing at all of these loci occurs in a default (male) mode. Only *dsx* transcripts are believed to encode functional proteins in both males and females. Ultimately, the majority of differentiation is regulated by DSX^F (females) or DSX^M (males), although some aspects appear to be regulated by TRA^F and TRA-2.

1.3 GERM LINE SEX-DETERMINATION IN *DROSOPHILA*

That sex-determination in the germ line of *Drosophila melanogaster* operates via a different pathway than somatic sex-determination was suggested by early pole cell transplantation experiments (Van Deusen, 1976; Marsh and Wieschaus, 1978; Schüpbach, 1982; for reviews see Pauli and Mahowald, 1990; and Steinmann-Zwicky, 1992). In these experiments it was shown that female gametes are only produced when the sex of the germ cells matched that of the surrounding soma, whilst male germ cells can be produced in either sex. Also, the sex of these germ cells is not altered in cells deficient of either *tra, tra-2, dsx* or *ix* function. Thus none of these genes are required autonomously in germ cells for sex-determination. Similar experiments involving *da*$^-$ germ cells yields identical results, whereby *da* has no effect on germ line sex-determination *per se*, but rather is required in the germ line for correct egg chamber morphology and the maternal effect on somatic sex-determination described previously (Cronmiller and Cline, 1987).

Later experiments involving pole cell transplantation's into host embryos lacking their own germ line (due to maternal *oskar* mutations) demonstrated that both inductive and cell-autonomous signals influence the sex of the germ cells (Steinmann-Zwicky *et al.*, 1989). XY germ cells become spermatogenic in an XX host, suggesting these cells determine sexual identity in a cell-autonomous manner. In contrast however, XX germ

cells do not become oogenic in an XY host, but rather become spermatogenic. Thus XX germ cells either require a positive inductive signal from XX somatic tissue directing them to become oogenic, or XY somatic tissue emanates a negative inductive signal which represses oogenic differentiation.

Factors identified to date which influence the determination of germ cell sex are discussed below.

1.3.1 Sex-lethal

Sex-lethal (Sxl) was first implicated in germ line, as well as somatic sex-determination, when it was observed that *Sxl*$^-$ germ cells transplanted into a wild type diplo-X host differentiate to form multicellular cysts in host ovaries (Schüpbach, 1985). Subsequently it was demonstrated that partial loss or gain of function *Sxl* mutations dictate spermatogenic or oogenic differentiation respectively, independent of the somatic sex (Steinmann-Zwicky *et al.*, 1989; Nöthiger *et al.*, 1989). Interestingly, consistent with the results of earlier experiments, individual XX germ cells undergo oogenic differentiation, whereas XY cells become spermatogenic, when transplanted into an ovary. In testis, both XX and XY cells, as well as XO cells, enter spermatogenesis. This suggests that, analogous to somatic sex-determination, germ cell sex-determination is influenced by autonomous signals dependent on chromosomal constitution (i.e. the X/A ratio). It is unlikely that the effects of *Sxl* are indirect (i.e. as a consequence of improper dosage compensation) since pole cells deficient of either *msl-1* or *msl-2* activity become spermatogenic and produce functional sperm. However, *mle* does appear to have some function in late spermatogensis (Bachiler and Sánchez, 1986).

Immunolocalisation of SXL protein reveals that in contrast to the expression of SXL^F protein in somatic tissue, no anti-SXL^F cross reacting antigens are detected in pole cells by gastrulation (Bopp *et al.*, 1991). However, western blot analysis demonstrates that several SXL isoforms are expressed in adult ovaries, and therefore *Sxl* must be activated at a somewhat later stage in the germ line compared to the soma. Interestingly, the sex-specific number of germ cells (at 8 hours after egg laying, female embryos possess on average 4 germ cells less than males) appears to be determined during migration of the germ cells to the sites where the gonads will form (Poirié *et al.*, 1995). This suggests germ cells acquire sexual identity prior to this stage, consistent with the results of Wei *et al.* (1991) who report that sexual dimorphism of germ cells is determined during embryogenesis. Collating this data suggests that the process which defines germ cell sex is activated shortly after invagination of the germ cells during gastrulation.

Despite the fact that activation of *Sxl* in somatic tissue requires the activity of *sis-a, sis-b (scute),* and *runt*, these genes do not seem to be required for *Sxl* activity in the germ line. Germ cells simultaneously heterozygous for *sis-a, sis-b , run* (i.e. *sis-a, sis-b,* and *run* over a wild type X chromosome) and a deficiency of *Sxl* (Sxl^{7M1}/Sxl^{fc} ; these females do not develop ovaries), when transplanted into a wild type diplo-X host, can develop into

functional oocytes (Granadino *et al.*, 1993). This suggest two things. Firstly, the lack of ovaries in Sxl^{7M1}/Sxl^{fc} females must be attributable to a somatic defect, and secondly, it supports the hypothesis that a positive feminising inductive signal emanates from diplo-X somatic tissue. These results are also consistent with the observation that *da* has no function in germ line sex-determination (Cronmiller and Cline, 1987), since in somatic tissue *da* and the aforementioned numerator elements interact to activate *Sxl*.

However, some genes involved in the regulation of *Sxl* expression in the soma do appear to be required in the germ line, and presumably have the same function in both tissues. Both *sans-fille (snf)* and *fl(2)d* generate an ovarian tumor phenotype analogous to that observed in germ cells deficient of *Sxl* (Steinmann-Zwicky, 1988; Salz, 1992; Granadino *et al.*, 1992). These genes are likely to function in a similar manner in both the germ line and soma, i.e. the maintenance of *Sxl* expression, since *snf* mutant females lack germ line specific *Sxl* transcripts (Salz, 1992) and germ cells homozygous for weak *fl(2)d* mutations (masculinising) can be rescued by Sxl^{M1}. Indeed, Bopp *et al.* (1993) report that in germ cells mutant for either snf^{1621} or *otu (ovarian tumor)*, *Sxl* is spliced in a male-specific manner and no anti-SXL^F cross reacting antigen is detected. These authors also suggest that SXL protein localisation may be critical for its function, since extensive changes in SXL^F subcellular localisation are apparent throughout oogenesis. This hypothesis is supported by the observation that several *Sxl* alleles (Sxl^{f4} and Sxl^{f5}), as well as the female-sterile loci *bag of marbles (bam*; McKearin and Spradling, 1980) and *fused* (*fu*; King, 1959) all affect the distribution of SXL^F protein.

Recently, Oliver *et al.* (1993) have implicated several loci in the process of germ line sex-determination, and in particular the regulation of *Sxl*, using RT-PCR to determine if several female-sterile mutants result in the production of male-specific *Sxl* transcripts in the female germ line. It was demonstrated that although not required autonomously in germ cells themselves, somatic *tra, tra-2,* or *dsx* mutations do nevertheless influence the splicing of *Sxl* transcripts in the germ line, and are therefore likely to regulate the production of an inductive signal. Also, the germ line ovarian tumor genes *sans-fille (snf), fused(fu), ovarian tumor (otu), ovo* and *Sxl* itself are all involved in the reception and/or interpretation of this inductive signal, and that *snf, fu, otu* and *ovo* all act upstream of *Sxl* in the germ line.

Consistent with these findings, Steinmann-Zwicky (1994) has demonstrated conclusively that somatic *tra* influences the sex of germ cells. Diplo-X germ cells developing in pseudomales lacking somatic *Sxl* sex-determining function are spermatogenic, but can be induced to become oogenic by ectopic expression of a hs-tra^F construct.

1.3.2 ovarian tumor

The X-linked recessive female-sterile *ovarian tumor (otu)* locus is involved in several processes in the germ line, particularly the determination of germ cell sex and

several aspects of oogenesis. Strong *otu* alleles result in a failure of female germ cell differentiation, whereas weak alleles allow female gonads to form, but such individuals are sterile due to defects apparent at several stages of oogenesis (Pauli *et al.*, 1993; and references therein).

A direct role for *otu* in germ line sex determination was first suggested by the observation that in the phenotypically characteristic tumorous ovaries generated in diplo-X otu^{1}/otu^{1} individuals, germ cells resemble small spermatocytes. That these cells are in fact sexually transformed was confirmed using two b-galactosidase enhancer trap lines which are specifically active in male germ cells. In wild type females, no germ cell staining is apparent with these markers, whereas *otu* deficient diplo-X individuals contain germ cells which express these markers strongly (Pauli *et al.*, 1993). Consistent with the results of Oliver *et al.* (1993), these authors also demonstrate that *otu* deficient diplo-X germ cells contain male-specific *Sxl* transcripts, suggesting *otu* acts upstream of *Sxl* in the germ line.

The *ovarian tumor* gene has now been cloned, and encodes two polypeptides of approximately 98 kDa and 104 kDa respectively, which arise by differential splicing (Mulligan *et al.*, 1988; Steinhauer *et al.*, 1989; Steinhauer and Kalfayan, 1992). The 104 kDa isoform can rescue all classes of *otu* mutants, whereas the 98 kDa isoform only rescues those *otu* alleles giving rise to a failure in germ cell differentiation (Sass *et al.*, 1995). This 98 kDa isoform, interestingly, is able to restore XY germ cell proliferation in agametic *otu* mutant pseudofemales, such that tumorous pseudo-ovaries are formed (Nagoshi *et al.*, 1995). This is particularly intriguing since XY germ cells in a wild type male soma do not require *otu* function, suggesting the soma of XY pseudofemales influences the requirement for *otu* in the germ line. Thus, Nagoshi *et al.* (1995) suggest *otu* may be involved in the reception of a female-specific signal emanating from the soma. However, the absence of homology to known signal transduction molecules, transcription factors, or RNA/ DNA binding motifs is inconsistent with this hypothesis.

1.3.3 ovo

Similar to the mutant phenotypes of *otu*, *ovo* mutations selectively affect the female germline and have no effect in the male germ line, although a non sex-specific function (under the *shavenbaby* pseudonym) is required for the elaboration of denticle belts (Oliver *et al.*, 1987, 1993). Diplo-X individuals homozygous for null *ovo* mutations lack functional gonads as a consequence of female germ cell degeneration, which initiates during the late blastoderm to early gastrulation stages. Additionally, when homozygous, several *ovo* alleles lead to the development of small ovarian tumours in which germ cells resemble spermatocyte morphology, suggesting an involvement in germ line sex-determination (Oliver *et al.*, 1987, 1990).

Several lines of evidence support this hypothesis. Firstly, analogous to several other genes implicated in germ line sex determination, *ovo* mutations perturb female-

specific splicing of *Sxl* transcripts (Oliver *et al.*, 1993), showing *ovo* acts upstream of *Sxl* in the germ line. Secondly, consistent with the involvement of chromosomal constitution (i.e. the X/A ratio) in germ line sex-determination, ovo^{+} function is dependent on the presence of an XX karyotype (Oliver *et al.*, 1994). Thirdly, individuals doubly heterozygous for ovo^{D1} and *snf* or *otu* display a synergistic interaction in which some 2X:2A germ cells follow a male pathway of differentiation (Pauli *et al.*, 1993), and in the case of *otu* this interaction appears to be dose dependent. In contrast to *otu*, the requirement for *ovo* function in the germ line does not seem to be dependent on somatic inductive signals, since diplo-X ovo^{D1rSI}/ovo^{D1rSI} homozygotes transformed into pseudomales (using somatic sex-determination mutants) develop pseudotestes which are either empty or contain undifferentiated germ cells (Nagoshi *et al.*, 1995). Recently, Pauli *et al.* (1995) have identified several regions which interact with ovo^{D2} in a screen covering approximately 58% of the *Drosophila* euchromatic genome. In particular, four potential suppressers and six enhancers of the ovo^{D2}/+ phenotype have been delimited. It will be interesting to see what function these genes may have in germ line sex-determination.

Consistent with the proposed role in germ line developmental programs and transcriptional regulation, *ovo* encodes a protein of at least 1209aa (131 kDa) and contains four putative zinc finger domains (Mevel-Ninio *et al.*, 1991).

1.3.4 SUMMARY OF GERMLINE SEX-DETERMINATION IN *DROSOPHILA*

Any model accounting for germ line sex-determination is inherently speculative at this time, since research in this field is in its infancy. However, collation of the present data supports a model whereby germ cells follow a male differentiation pathway unless instructed to do otherwise by a positive female-specific inductive signal emanating from the female soma. The receptivity of germ cells to this inductive signal is dependent on their karyotype, such that XX germ cells are competent whereas XY germ cells are not. This aspect of germ cell sex-determination is therefore dependent on cell-autonomous factors.

Current data indicates that otu^{+} activity renders germ cells competent, suggesting this gene is involved in transducing the inductive signal into an autonomous feminising activation signal. How *ovo* activity can be accommodated in this model is unclear. Dose-dependent synergistic interactions are apparent between *otu* and dominant *ovo* alleles (e.g. in ovo^{D}/+ individuals), such that increased *otu* dose suppresses the ovo^{D} phenotype whereas reduced *otu* dose has an enhancing effect. Furthermore, ovo^{+} activity is only required in diplo-X, and never in haplo-X (either XY or XO), germ cells. It is unlikely that *otu* regulates *ovo* activity since a b-galactosidase reporter gene construct under the control of an *ovo* germ line specific promoter is still active in *otu* deficient germ cells (Nagoshi *et al.*, 1994). Interestingly, this reporter gene construct is inactive in ovo^{D1}/+ germ cells, suggesting ovo^{D1} is antimorphic, and acts as a negative regulator of *ovo* activity (i.e. *ovo* has autoregulatory activity). Perhaps *otu* and *ovo* encoded proteins complex with one another to activate the female germ cell differentiation pathway, since this could

explain the synergistic interactions between these two loci. For example, increased OTU protein concentration (as a consequence of increased *otu* dose) could alleviate the negative autoregulation of *ovo* by sequestration of OVO^D protein. This hypothesis is of course merely speculation however, since there is no evidence of a molecular interaction between the proteins derived from these two loci.

Ultimately however, germ line sex-determination is dependent on *Sex-lethal* activity. Although neither *tra, dsx* or *ix* have an autonomous role in germ line sex-determination, the possibility that *Sxl* regulates germ line specific downstream sex-determination genes can not be excluded. Speculative models for male and female germ line sex-determination are illustrated in Figure 1.8.

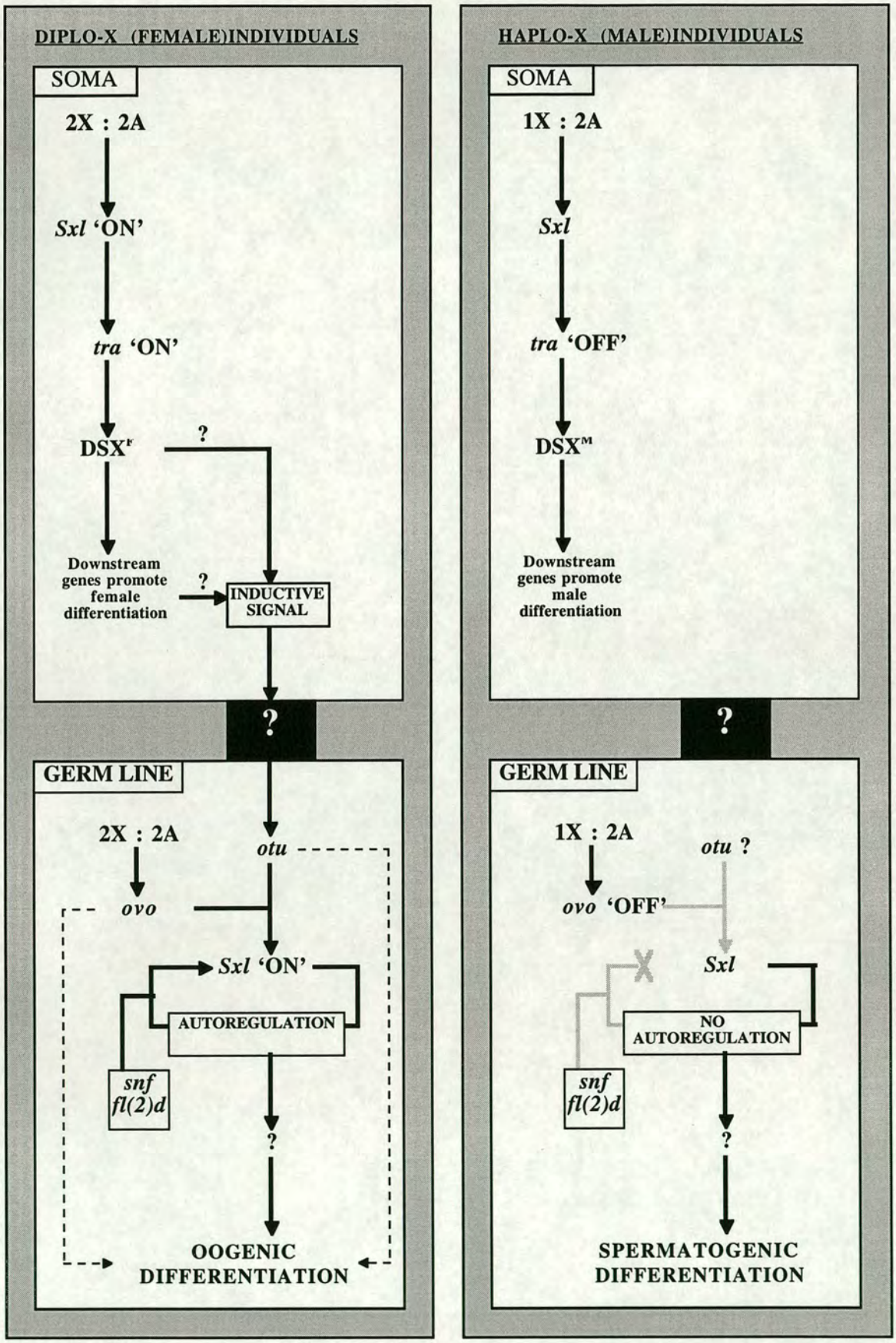


Figure 1.8. Germ line sex-determination in *Drosophila melanogaster*. Please refer to text for description of the genes implicated. Hatched lines indicate a function not related to sex-determination. Bold lines indicate activity. '?' as yet unidentified factor(s).

1.4 SEX-DETERMINATION IN OTHER DIPTERAN SPECIES

In contrast to our extensive knowledge of sex-determination in *Drosophila melanogaster*, relatively little is known about how this critical aspect of development is regulated in other Dipteran species. It does seem however that a wide variety of mechanisms are employed to determine sexual phenotype. In many species sex is determined by the presence or absence of a dominant male-determining allele (*M*), such as in some strains of *Musca domestica* (see section 1.4.1), several *Chironomus* species (Thompson and Brown, 1972l; Martin and Lee, 1984; Hägele, 1985), *Calliphora erythrocephala, Culex molestus,* and *Megaselia scalaris* (reviews: Nöthiger and Steinmann-Zwicky, 1985; Dübendorfer *et al.*, 1992; and references therein). In other species temperature (e.g. *Aedes stimulans*; Horsfall and Anderson, 1963), nutritional status (*Heteropeza pygmae*; Went and Camezind, 1980), or maternal factors (*Chrysomya rufifacies*; Ullerich, 1984) seem to be the major sex-determining factors.

However, in the majority of cases, relatively few loci implicated in sex-determination have been identified, and most certainly have not been characterised at a molecular level. For this reason, any functional comparisons between these loci and the *Drosophila* sex-determination genes must be treated with some scepticism. Despite this, Nöthiger and Steinmann-Zwicky (1985) have proposed a single unifying program for sex-determination to which all other insect sex-determination mechanisms can be related. In this system, a zygotic feminising master regulatory gene (analogous to *Sxl*) is activated to direct female differentiation. Regulation of this gene could occur by maternal factors, zygotic male repressors, temperature dependent factors, or a combination of all of these things. The presence or absence of functional activity at this *Sxl*-like locus would then determine which of two active states is expressed by a downstream double switch gene analogous to *dsx*, which in turn would dictate either male or female development respectively.

Since work reported in this thesis directly relates to *Musca domestica*, a more detailed discussion of sex-determination in this species is presented below.

1.4.1 Sex determination : the case of *Musca domestica*

In *Musca domestica,* having a karyotype of five pairs of autosomes and two sex chromosomes (i.e. females [10AXX], males [10AXY]), there appear to be a wide variety of sex-determination mechanisms throughout various natural populations (recently reviewed by Dübendorfer *et al.*, 1992), which are illustrated in Figure 1.9. Standard strains have a dominant male-determining allele (*M*) located on the Y chromosome, and therefore exhibit classical sex-linkage. In other strains however, *M* can be located on any of the autosomes, or indeed on an X chromosome, such that individuals homozygous (*M/M*) or heterozygous (*M/+*) develop as males, whereas wild type individuals (+/+) develop as females (Wagoner *et al.*, 1969; Franco *et al.*, 1982, Denholm *et al.*, 1985).

To further complicate this system, yet more *Musca domestica* strains have been discovered in which sex is determined by the presence or absence of a dominant female-determining factor (F^D)located on autosome IV (McDonald *et al.*, 1978; Denholm *et al.*, 1985). In such flies, both males and females are homozygous for *M*, but female development is dictated by the presence of F^D (i.e. F^D is epistatic to *M*). Using mitotic recombination to generate mosaic flies, Hilfiker-Kliner *et al.* (1993) have demonstrated that removal of F^D from *M/M;* F^D/+ cells prior to pupation results in sex-reversal. *M* on the other hand is required early during embryogenesis, since in those strains in which *M* is the primary determinant of sex (i.e. *M/M* or *M*/+ = male, whereas +/+ = female), mitotic recombination at the syncitial blastoderm stage of embryogenesis induces sex-reversal, whereas later mitotic recombination does not. This also indicates that once *M* induces male development, subsequent differentiation becomes independent of *M*, and that somatic sex-determination in *Musca domestica* is determined in a cell-autonomous manner.

Interestingly, pole cell transplantation experiments demonstrate that the genotype of the transplanted pole cells does not determine the sex of the germ line (i.e. neither *M* nor F^D has an effect), rather this is entirely determined by somatic induction (Hilfiker-Kliner, 1994). Somewhat surprisingly however, germ cell genotype does influence the sex of progeny individuals. During oogenesis of *M*/+ germ cells, *M* exerts a maternal effect such that all progeny develop as males irrespective of genotype (even those of a female +/+ constitution). Similar to the epistatic relationship observed in somatic tissue, F^D can override this maternal effect of *M*, such that +/+ progeny now develop as females.

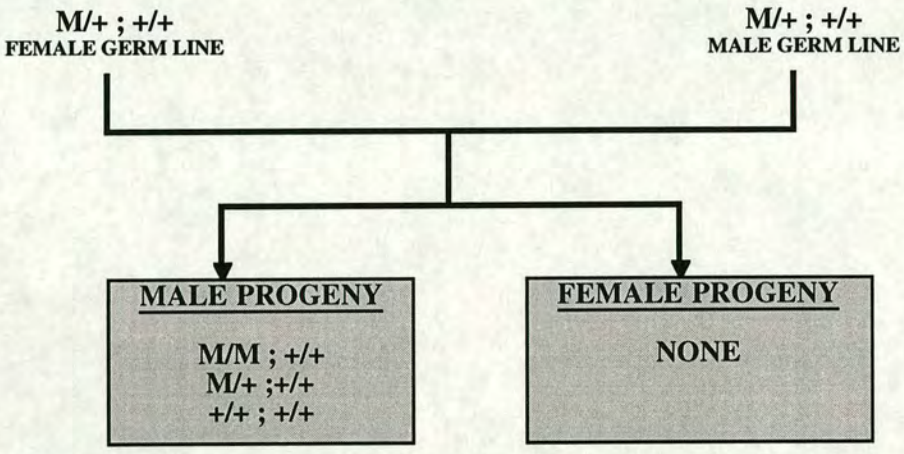
Two other mutations which induce sexual transformation by a maternal effect have been reported in *Musca domestica*. Both the *Arrhenogenic (Ag;* Vanossi Este and Ravoti, 1982) and *transformer (tra;* Inoue and Hiroyoshi, 1986*)* mutations, in the absence of *M*, transform genotypically female zygotes into fertile males and intersexes. Thus, in this case, these males lack a zygotic male-determining allele, and are therefore referred to as '*NOM*-males' (Dübendorfer *et al.*, 1992). It has therefore been possible to establish a *Musca domestica* strain in which sex is determined entirely by maternal effect. However, several lines of evidence suggest that *Ag* and *tra* may in fact represent hypomorphic *M* and *F* alleles respectively. Firstly, in one *Ag* strain an *M* factor arose spontaneously and mapped to the *Ag* locus. Secondly, *tra* could not be genetically separated from F^D (Hilfiker-Kleiner, 1994; and references therein).

A fuller understanding of how all of these loci interact in what at present appears to be an inherently complex manner will require the cloning and characterisation of the genes encoding *M* and F^D function.

A ; VARIOUS SEX-DETERMINING SYSTEMS

SEX	STANDARD STRAINS	STANDARD STRAINS	F^D STRAINS
FEMALE	X/X	+/+	M/M ; F^D/+
MALE	X/Y^M	M/+ ; M/M	M/M ; +/+

B : MATERNAL EFFECT OF 'M'



C: MATERNAL EFFECT OF F^D

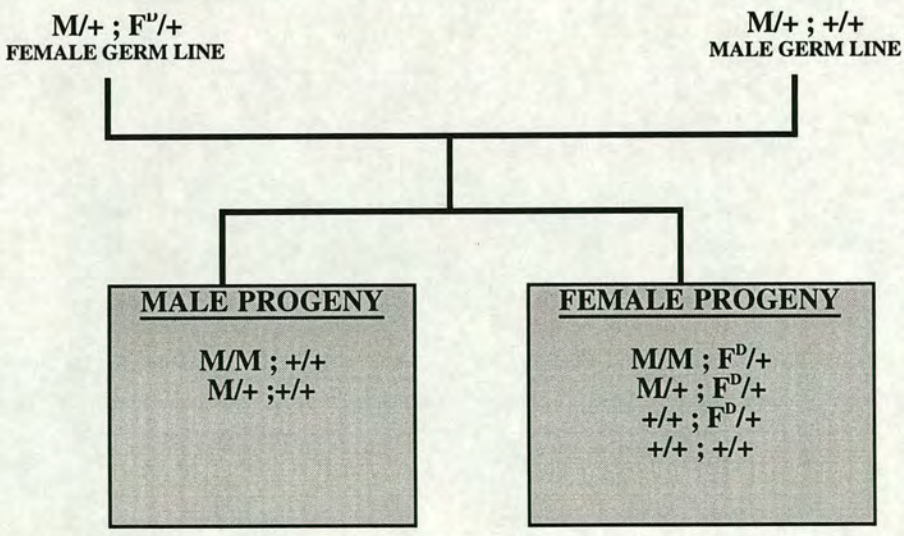


Figure 1.9. A. Sex determination mechanisms in various strains of *Musca domestica*. **B.** Maternal effect of *M* , the male-determining factor, as revealed by germ line transformation experiments. **C.** Maternal effect of F^D , the female-determining factor as revealed by germ line transformation experiments

1.5 VITELLOGENESIS

Vitellogenesis defines the process whereby yolk is synthesised and accumulated in developing oocytes, and has been extensively characterised in a wide variety of organisms (for reviews see Hagedorn and Kunkel, 1979; Bownes, 1986). Since the aim of the work reported here was to isolate and/or identify genes homologous to *Drosophila* sex determination genes, an extensive review of vitellogenesis in other species is unnecessary. However, since the *yolk protein* genes are the only known genes in *Drosophila* where a direct regulation by *dsx* proteins has been demonstrated (see section 1.5.3.1.1.2), such that differential expression occurs in a non sex-specific tissue (the fat body), an understanding of their functional significance and regulation would be beneficial. For this reason I include here an extensive review of *yolk protein* gene expression in *Drosophila*, and where appropriate discuss pertinent analogies to other insects in the relevant results chapters.

1.5.1 VITELLOGENESIS IN *DROSOPHILA MELANOGASTER*

In *Drosophila* there are three major proteins present in the yolk, which are encoded by three single copy X-linked genes collectively referred to as the *yolk protein (yp)* genes (Barnett *et al.*, 1980; Postlethwait and Jowett, 1980; for review see Bownes *et al.*, 1993). Synthesis of yolk proteins (YP's) occurs in the female fat body and the ovarian follicle cells (Bownes and Hames, 1977, 1978; Brennan *et al.*, 1980, 1982). YP's synthesised in the fat body are secreted into the haemolymph, whereas those synthesised in the ovarian follicle cells are secreted unidirectionally towards the oocyte via the interfollicular spaces. All YP's are taken into the developing oocyte by receptor mediated endocytosis (Gelti-Douka *et al.*, 1974; Bownes and Hames, 1977, 1978; Brennan *et al.*, 1980; Isaac and Bownes, 1982; , 1980; Isaac and Bownes, 1982; Yan and Postlethwait, 1990; Butterworth *et al.*, 1992; Giorgi *et al.*, 1993).

Post translational modification of these proteins was first suspected when it was noticed that yolk protein polypeptides derived from *in-vitro* translation systems were approximately 1kDa larger than their *in-vivo* counterparts (Postlethwait and Kaschnitz, 1978; Brennan *et al.*, 1980). Subsequently glycosylation (Minoo and Postlethwait, 1985), phosphorylation (Minoo and Postlethwait, 1985; DiMario *et al.*, 1987) and tyrosine sulfation (Baeuerle *et al.*, 1988; Friedrich *et al.*, 1988) of these yolk proteins has been demonstrated.

The genes encoding the major yolk polypeptides have now been cloned (Hung and Wensink, 1982, 1983; Garabedian *et al.*, 1987), and show extensive homology in their C-terminal regions (Yan *et al.*, 1987). The *yolk protein-1 (yp1)* and *yolk protein-2 (yp2)* genes are located at cytological location 8F-9A. Both *yp1* and *yp2* genes contain single introns, are divergently transcribed, and share common regulatory elements located in an 1225bp intergenic spacer (Hung and Wensink, 1983; Garabedian *et al.*, 1985; Tamura

et al., 1985). The *yolk protein-3 (yp3)* gene contains two introns, and is located some distance away from *yp1* and *yp2* (>1000kb) at cytological band 12BC (Garabedian *et al.*, 1987). The organisation of *yp1* and *yp2*, as well as *yp3*, is discussed in section 1.5.2, as are enhancer regions regulating their expression. Several factors known to influence the synthesis of yolk proteins are also discussed in section 1.5.3.

Initially, the sole function of the yolk proteins was thought to be as a nutritional source for the utilisation during embryogenesis. However, the high degree of homology between *yp* genes in a variety of species, and apparent homology to lipases, suggests a rather more positive role. Indeed, the yolk spheres present in developing oocytes have been shown to contain inactivated ecdysteroid conjugates (Bownes *et al.*, 1988). Thus Bownes *et al.* (1992) suggest that in addition to supplying the developing embryo with sufficient amino acids to complete embryogenesis, the controlled degeneration of the yolk spheres results in a timed and regulated release of ecdysone.

1.5.2 GENOMIC ORGANISATION OF THE *YP* GENES

The *yolk protein-1 (yp1)* and *yolk protein-2 (yp2)* genes have been extensively characterised, and much has been elucidated with respect to how the expression of these divergently transcribed genes is regulated. As discussed previously, these genes are closely linked at 8F-9A on the X-chromosome, being separated by a short intergenic spacer of 1225bp (Hung and Wensink, 1983). *yp1* generates a single 1.6kb transcript encoding a single polypeptide of 439aa (approximately 47kDa). In contrast transcription of *yp2* generates two transcripts of 1.60kb and 1.67kb which differ in size due to alternative transcriptional termination sites, although both proteins encode identical polypeptides of 442aa (approximately 46 kDa).

The *yolk protein-3 (yp3)* gene is however located some 1000kb away at cytological location 12BC (Garabedian *et al.*, 1987), contains two introns, and generates a single transcript of 1.54kb encoding a 420aa polypeptide (approximately 45kDa). In all cases the yolk proteins (YP's) synthesised contain a short signal peptide sequence (Brennan *et al.*, 1980; Minoo and Postlethwait, 1985), which is essential for secretion (Liddel and Bownes, 1991a). Transcription of all of the *yp's* is first detected shortly after eclosion, and plateaus some 24 hours later, consistent with observed protein titres (Isaac and Bownes, 1982). Both the ovarian follicle cells and female fat body cells seem to synthesise equivalent amounts of YP's, as judged by transcript levels, although reduced *yp3* transcript levels are evident in ovaries. This however is not a result of reduced *yp3* transcription or amplification of the *yp1* and *yp2* genes, but rather a consequence of reduced *yp3* transcript stability (Williams and Bownes, 1986). Figure 1.10 shows the schematic organisation of the *yp1*, *yp2*, and *yp3* loci.

1.5.3 THE REGULATION OF *YOLK PROTEIN* GENE EXPRESSION

The *yolk protein* genes are expressed in a sex-, stage-, and tissue-specific manner in *Drosophila melanogaster* (Bownes and Hames, 1978; Brennan *et al.*, 1982; Isaac and Bownes, 1982). Expression occurs in the female fat body cells and in the ovarian follicle cells during stages 8-10B of oogenesis, whereas no expression is observed in males. In addition to sex-specific regulation, both hormonal and nutritional signals influence *yp* gene expression (for review see Bownes *et al.*, 1993, Bownes, 1994). How all of these regulatory signals can be co-ordinated such that expression occurs correctly is an intriguing puzzle. Fortunately, the cloning of the *yp* genes has facilitated a comprehensive dissection of *cis*-acting regulatory elements involved in the regulation of expression. Figure 1.11 shows a schematic model for the regulation of *yp* gene expression.

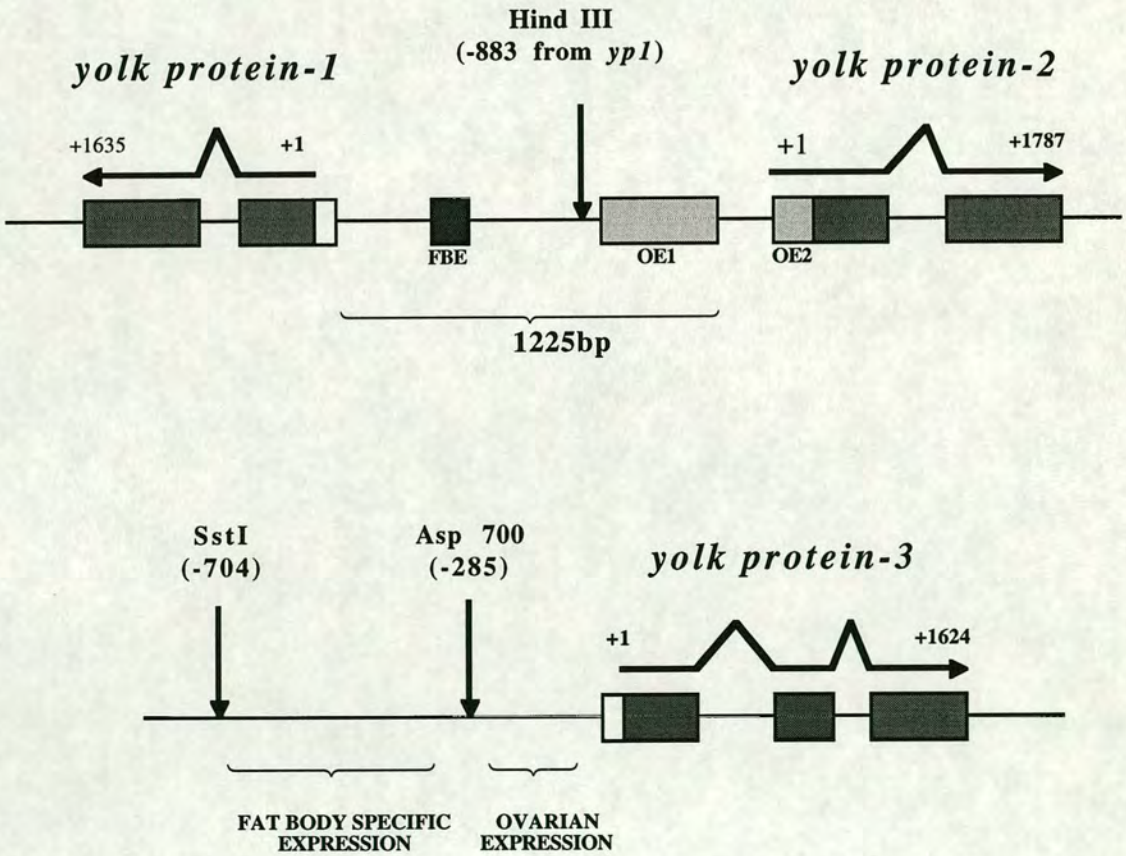


Figure 1.10. Genomic organisation of the *yp1*, *yp2* and *yp3* genes indicating identified regions involved in regulating their expression. FBE is the fat body enhancer, whereas OE refers to ovarian enhancer (i.e. OE1 is the ovarian enhancer-1).

1.5.3.1 IDENTIFICATION OF ENHANCERS ELEMENTS

1.5.3.1.1 REGULATION OF *YP-1* AND *YP-2* EXPRESSION

1.5.3.1.1.1 Fat body specific expression

Early experiments (Tamura *et al.*) demonstrated that all *cis*-acting regulatory sequences required for the correct sex-, stage-, and tissue-specific expression of *yp1* and *yp2* are present within a 5.0kb genomic fragment containing a 3' truncated *yp1* and complete *yp2* sequence separated by the 1225bp intergenic spacer. Further, separation of *yp1* and *yp2* sequences by *Hin* dIII restriction enzyme digestion (please refer to Figure 1.10) such that each gene has 883bp and 342bp of upstream regulatory sequences respectively demonstrates that ovarian and fat body specific enhancers are located in these regions (Garabedian *et al.*, 1985). The *yp1* construct (containing 883bp of upstream sequence) is expressed only in the female fat body, whereas the *yp2* construct (containing 342bp of upstream sequence) is expressed only in the ovarian follicle cells at the correct stages of oogenesis. Subsequent experiments using β-galactosidase reporter gene constructs have delimited a fat body specific enhancer element to an 125bp sequence located 195bp upstream of the *yp1* cap site (Garabedian *et al.*, 1986). However, despite the fact that this so called 'fat body enhancer' (FBE) confers correct sex- and tissue-specific expression on reporter gene constructs, it is not essential. An *Alcohol dehydrogenase (Adh)* reporter gene, regulated by the intergenic spacer in which the FBE has been deleted, is still expressed in the correct sex- and tissue-specific manner (Abrahamsen *et al.*, 1993).

1.5.3.1.1.2 Regulation of sex-specific expression

The observation that yolk proteins are only synthesised in the fat body of females, and not males, is particularly fascinating since it suggests those genes regulating sexual differentiation are also involved (either directly or indirectly) in the regulation of *yolk protein* gene expression. Indeed, early experiments demonstrated that sexual transformation (induced by mutations at several sex-determining loci) is co-ordinate with *yp* expression, such that individuals with any apparent female morphological traits are seen to synthesise YP's (Postlethwait *et al.*, 1980; Bownes and Nöthiger, 1981). That expression of the *yolk protein* genes is continuously dependent on the sex determination hierarchy was first demonstrated using temperature shift experiments. *tra-2*^{ts} homozygotes reared at the permissive temperature (16°C) develop as females, whereas those reared at the restrictive temperature (29°C) develop as pseudomales. At the permissive temperature, normal *yp* expression is observed, but at the restrictive temperature no expression is detected. Conclusively, pseudomales (individuals reared at 29°C for 2 days) show restored *yp* expression when moved to the permissive temperature even though they are morphologically male (Belote *et al.*, 1985). These results suggest that expression of the

yp's is continuously dependent on either *tra-2*, or a gene downstream of it which is expressed sex-specifically. Subsequent experiments demonstrated that direct regulation by *doublesex* proteins was likely to be responsible for the sex-specific expression of the *yp's*, since DSX^F and DSX^M were shown to bind specifically to identical sequences located within the FBE (Burtis *et al.*, 1991). These binding sites were delimited to three specific locations using DNA footprinting assays.

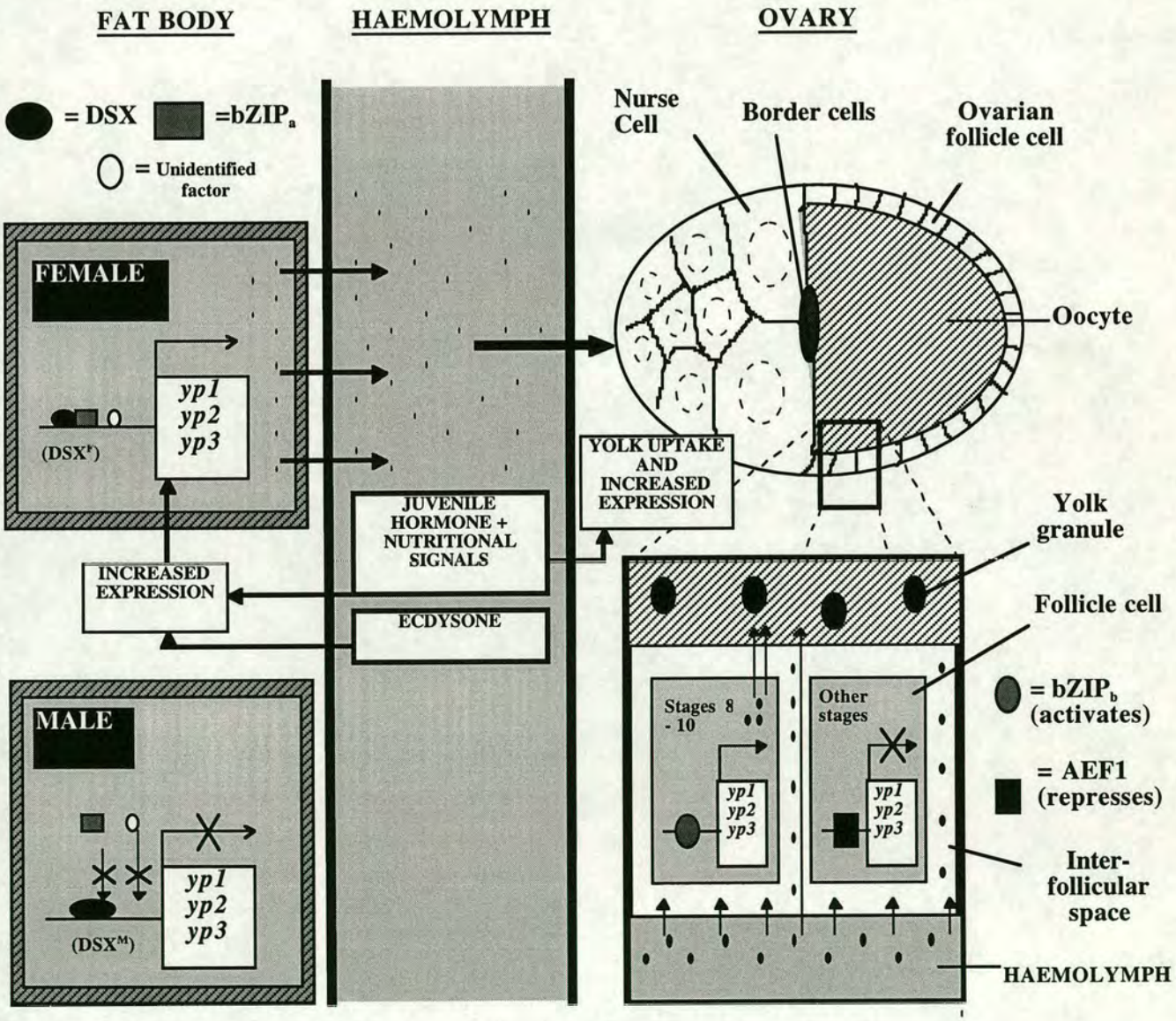


Figure 1.11. Schematic diagram illustrating the effect of various factors on *yolk protein* gene expression in *Drosophila melanogaster*.. The two fat body cells indicated are female (top) or male (male) and contain either DSX^F or DSX^M respectively. $bZIP_a$ and $bZIP_b$ represent putative basic Leucine zipper proteins.

These *dsx* proteins share identical N-terminal sequences, including an atypical zinc finger DNA binding domain, but differ at their 3' ends such that DSX^M has an extended C-terminus relative to DSX^F (see section 1.2.2.4). This suggests that, in the case of *yp* regulation, DSX^M may repress expression by steric hindrance of *trans*-acting factors at the FBE, whereas DSX^F would not. Recent experiments (Coschigano and Wensink, 1993) confirm this hypothesis. Reporter gene constructs under the control of FBE elements containing independent mutations in all three *dsx* footprint sites, which reduce DSX binding *in-vitro*, show expression to varying degrees (dependent on the mutation site) when transformed into male individuals. Simultaneous mutations in both of the strongest DSX binding sites located in the FBE result in reporter gene expression in males analogous to that observed with the native FBE in females. Interestingly, these experiments also demonstrate that DSX^F enhances expression slightly, presumably by stabilising any complexes formed at this site.

Further experiments have delimited a short regulatory element within the FBE (termed the o-r element) consisting of four protein binding sites, which is capable of directing reporter gene expression similar to that of the native *yp* genes (An and Wensink, 1995a, 1995b). The 'o' element (29bp) contains overlapping binding sites for DSX, AEF1 (adult enhancer factor-1; Falb and Maniatis, 1992), and an as yet unidentified basic Leucine zipper protein which binds to a site termed $bZIP_1$. Initially the bZIP protein was thought to be the CCAAT/enhancer binding protein (C/EBP), since this factor has been implicated in the regulation of *Alcohol dehydrogenase* (*Adh*) expression in fat body tissue, as is AEF1. However, this can not be the case since C/EBP is encoded by the *slow border cell* gene, and mutations at this locus have no effect on either reporter gene expression levels (An and Wensink, 1995b) or YP levels in the haemolymph (M. Bownes, unpublished). The 'r' element (11bp) contains a single protein binding site, ref1, which binds an as yet unidentified factor. With this o-r enhancer element, An and Wensink (1995b) have been able to confirm the hypothesis that DSX^M acts to repress *yp* transcription in the fat body by steric hindrance. Both DSX^F and DSX^M exclude AEF1 from its binding site and thus prevent repression by this protein, and the sex-specific activation of transcription becomes dependent on the synergistic interaction of DSX^F and the protein (referred to as $bZIP_a$) bound at $bZIP_1$. DSX^M represses activation by steric hindrance of $bZIP_a$ as a consequence of its extended C-terminal domain. Interestingly, ovarian expression is also evident with this construct, and is dependent on AEF1. An and Wensink (1995b) propose that since the concentration of AEF1 encoding transcripts are approximately three fold higher in ovaries relative to the fat body, and no *dsx* transcripts are detected in either ovaries or in male individuals, this apparent tissue-specific difference in AEF1 binding is likely to be dependent on the relative concentration of AEF1 and DSX proteins in each tissue.

The proposed model for *yp* gene regulation by these proteins is depicted in Figure 1.11.

1.5.3.1.1.3 Regulation of ovarian expression

In addition to the FBE, early experiments suggested a 2.8kb fragment encompassing 342bp of upstream sequence and the entire *yp2* coding region contained sequences enhancing ovarian-specific expression of *yp2* (Garabedian *et al.*, 1985). This region was in fact shown to contain two enhancer elements, termed ovarian enhancer-1 (OE1) and ovarian enhancer-2 (OE2), which are 301bp and 105bp in length respectively (Logan *et al.*, 1989). OE1 is located 343bp upstream of the *yp2* cap site, and can be subdivided into two regions which are necessary for the correct ovarian- and cell type-specific expression of both *yp1* and *yp2* (Logan and Wensink, 1990). Expression is apparent only in the columnar follicle cells and their descendants, and is not detected in the border cells and squamous epithelia covering the nurse cells. However, deletion of a 91bp region (-43 to -133) within OE1 results in strong expression in both the border cells and anterior pole cells. Logan and Wensink (1990) propose that the -134 to -282 region (149bp) directs border cell and anterior pole cell expression, but that the adjacent 91bp region (-43 to -133) somehow represses this, and enhances expression in the columnar main body cells covering the developing oocyte.

The OE2 region is located, surprisingly, within the first exon of the *yp2* gene (+1 to +105), and is required for normal levels *yp1* ovarian-specific expression. Although OE2 does direct the correct cell type-specific expression of reporter gene constructs, expression is not restricted to the correct stages of oogenesis. Thus, elements located within OE2 are not sufficient to direct the correct stage-specific expression of *yp1*, although the possibility that the OE2 enhancer extends beyond the boundaries defined in these analyses has not been excluded.

1.5.3.1.2 REGULATION OF *YP-3* EXPRESSION

In contrast to *yp1* and *yp2*, the promoter and enhancer elements regulating *yp3* expression are relatively poorly defined. Preliminary experiments demonstrated that a fragment spanning the *yp3* coding region, and encompassing 706bp of upstream and 825bp of downstream sequence is expressed in the correct sex-, stage-, and tissue-specific manner when transformed into a YP3 deficient background (Liddel and Bownes, 1991b). These results have recently been refined by Ronaldson and Bownes (1995), who report that sequences 3' of the *yp3* coding region do not regulate any aspects of *yp3* transcription. Further, subdivision of the 706bp upstream sequence reveals that, analogous to the results obtained with the *yp1/yp2* intergenic region, fat body and ovarian-specific enhancers are separable and located in this region. A 419bp fragment (-704 to -285) confers the correct sex- and tissue-specific expression of a reporter gene in the fat body, and the adjacent 328bp fragment (-285 to +43; termed the ovarian enhancer-3 [OE3]) confers the correct stage- and tissue-specific expression of the reporter gene in ovarian follicle cells.

Comparisons of these regions with the FBE, OE1, and OE2 sequences suggest factors regulating *yp1* and *yp2* expression (i.e. DSX and AEF1 etc.) may also regulate *yp3*. If this is the case, analysis of how enhancers and *trans*-acting factors regulate *yp3* transcription may provide insight into how *yolk protein* genes are regulated in general. The interpretation of results is also inherently less complex with yp3 since these factors only regulate the expression of one gene, whereas those present in the intergenic spacer between *yp1* and *yp2* must act divergently.

1.5.3.2 HORMONAL CONTROL OF *YP* GENE EXPRESSION

1.5.3.2.1 Ecdysone

Ecdysone was first implicated in *Drosophila* vitellogenesis when it was observed that isolated abdomens (which have reduced yolk protein levels) show increased YP concentrations after injection of 20-hydroxyecdysone (Jowett and Postlethwait, 1980). This effect must be a consequence of increased transcription, rather than transcript stability, since 20-hydroxyecdysone also induces YP synthesis in males (Postlethwait *et al.*, 1980; Bownes *et al.*, 1983; Kozma and Bownes, 1986). Inhibition of protein synthesis by treatment with cyclohexamide prevents 20-hydroxyecdysone induction of YP synthesis, suggesting its effects are mediated indirectly via a protein intermediate (Bownes *et al.*, 1987).

Despite the fact that the enhancer elements regulating sex-, stage-, and tissue-specific expression of *yp1* and *yp2* have been localised, ecdysone responsive regions have proved difficult to identify (Tamura *et al.*, 1985; Shirras and Bownes, 1987). Recent experiments (Bownes *et al.*, in press) using reporter gene constructs, however, suggest an 20-hydroxyecdysone responsive element is located in 345bp of 5' *yp2* flanking DNA, and perhaps a second 5' of *yp1*. In addition, computer searches identified several putative ecdysone responsive elements in *yp3*, localising to the coding sequence between introns I and II, 3' of the coding region, and also in the ovarian enhancer-3

1.5.3.2.2 Juvenile Hormone

Similar to the observed induction by ecdysone, topical application of a juvenile hormone analogue (ZR515) to isolated abdomens increases yolk protein synthesis in females (Jowett and Postlethwait, 1980). In contrast to ecdysone induction however, increased YP levels are also observed in ovaries, although this may be an indirect result of increased yolk uptake, since juvenile hormone is thought to induce progression of vitellogenesis (Wilson, 1982; reviewed by Bownes, 1986). No juvenile hormone responsive *yolk protein* regulatory elements have been identified to date, although Bownes *et al* (in press) have observed ZR515 (Methoprene, a juvenile hormone analogue) responsiveness with some *yp2/ yp3* reporter gene constructs in mutant backgrounds. It is unclear at this

stage, however, if this is a consequence of transcriptional activation, derepression, or transcript stability.

1.5.3.3 NUTRITIONAL CONTROL OF *YP* GENE EXPRESSION

Both vitellogenesis and oogenesis are dependent on the nutritional status of the fly. Egg-laying females transferred from a normal diet to one of only sugar and water show a rapid cessation of egg production, and the ovaries of such females only contain pre-vitellogenic oocytes (Bownes and Blair, 1986; Bownes *et al.*, 1988). Both juvenile hormone and ecdysone can induce *yp* transcription in these starved females, although only juvenile hormone can restore vitellogenesis. A rapid restoration of oogenesis is observed when flies are returned to a normal diet or treated with ecdysone, although vitellogenesis is never restored in those flies only treated with ecdysone.

Despite these observations, the nutritional response is unlikely to be mediated by either juvenile hormone or ecdysone since similar titres of these hormones are observed in starved and fed flies (Bownes, 1989; Bownes and Reid, 1990). Initial experiments using reporter gene constructs (Bownes *et al.*, 1988) have shown that a nutritional response element is located in the 883bp fragment of 5' *yp1* flanking DNA (i.e. the fragment derived from *Hin* dIII digestion of the intergenic spacer; refer to Figure 1.10). Recent experiments have enhanced these findings, and demonstrate that a nutritional response element is also contained within the 342bp upstream of *yp2*. Surprisingly, the FBE itself does not exhibit a good nutritional response (Søndergaard *et al.*, 1995). These authors propose that the nutritional response is mediated via a specific peptide produced by digestion of yeast extracts, and is independent of *dsx* mediated regulation (since *dsx* transcripts are present in similar levels in starved and fed flies). This hypothesis is based on the observation that starved females fed on a diet of amino acids do not elicit a nutritional response (Bownes *et al.*, 1988), even though treatment of flies with a proteinase inhibitor reduces the flies receptivity to nutritional induction.

1.5.4 SUMMARY OF *DROSOPHILA* VITELLOGENESIS

Yolk proteins are essential for embryogenesis, providing the developing embryo with a source of amino acids, as well as a hypothesised role as a protective store for inactivated ecdysteroid conjugates which are released and subsequently utilised upon yolk protein degradation. The yolk proteins, encoded by three X-linked genes (*yp1, yp2* and *yp3*) are synthesised in two tissues in *Drosophila melanogaster*, namely the female fat body and ovarian follicle cells. The sex-specific expression of the *yp* genes in fat body cells is regulated autonomously by the proteins encoded by the terminal sex-determining gene *doublesex*, and represent the first conclusive indication of direct sex-specific regulation of downstream differentiation gene expression in a non sex-specific tissue. The stage-specific expression of the *yp's* in ovarian follicle cells (during stages 8-10B of oogenesis) is thought be regulated by the relative concentrations of repressive (e.g.

AEF1) inductive (an putative bZIP protein) tissue-specific factors. Hormones, and an inductive signal correlated to the nutritional status of the fly, all influence the expression of the *yolk protein* genes.

2. MATERIALS AND METHODS

2.1 GENERAL

2.1.1 Suppliers and sterilisation techniques

Chemicals were obtained from SIGMA, BDH and Aldrich.

Restriction enzymes, modification enzymes, polymerases and reverse transcriptase were obtained from GIBCO BRL, USB, NBL and Pharmacia.

Redivue radioisotopes [^{32}P]dCTP and [^{35}S]dATP, Hybond-N, Hybond-N+, Hybond-C, Qiagen and Qiaex kits were obtained from Amersham.

Typically, solutions were made using sterile double distilled water in sterile baked glassware. Sterilisation was achieved either by autoclaving (15 psi/ 15 min) or by passing through a 0.22µm pore sized filter.

Glassware to be used in the preparation of RNA was rendered RNAase free by soaking for at least one hour in distilled water supplemented to 0.1% (v/v) DEPC. The water was then discarded and residual DEPC removed by autoclaving (15 psi/ 20 min).

Solutions required to be RNAase free were rendered such by supplementing to 0.1% (v/v) DEPC, incubating overnight at room temperature, and then autoclaving to remove the DEPC (15 psi/ 20 min). Solutions reactive with DEPC (and thus unsuitable for this treatment) were made using pre-DEPC-treated sterile double distilled water and then autoclaved as per normal.

All necessary plasticware was sterilised by autoclaving (15 psi/ 15 min), and then placed at 37°C until dry.

Sequence analysis was performed using Genejockey II and the Genetics Computer Group software (hereby referencing the 'Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711').

2.1.2 General solutions and buffers

Restriction enzyme buffer (GIBCO BRL) constitutions are detailed in Table 2.1

General solutions are listed in Table 2.2

The composition of Qiagen and Qiaex solutions is detailed in Table 2.3

All solutions were stored at room temperature unless otherwise stated.

Table 2.1. GIBCO BRL restriction enzyme buffers

BUFFER	COMPOSITION
10 x Reaction Buffer 1	50mM Tris-HCl (pH 8.0), 10mM Magnesium Chloride
10 x Reaction Buffer 2	50mM Tris-HCl (pH 8.0), 10mM Magnesium Chloride, 50mM Sodium Chloride
10 x Reaction Buffer 3	50mM Tris-HCl (pH 8.0), 10mM Magnesium Chloride, 100mM Sodium Chloride
10 x Reaction Buffer 4	20mM Tris-HCl (pH 7.4), 50mM Magnesium Chloride, 50mM Potassium Chloride
10 x Reaction Buffer 5	10mM Tris-HCl (pH 8.2), 8mM Magnesium Chloride
10 x Reaction Buffer 6	50mM Tris-HCl (pH 7.4), 60mM Magnesium Chloride, 50mM Potassium Chloride, 50mM Sodium Chloride
10 x Reaction Buffer 7	50mM Tris-HCl (pH 8.0), 10mM Magnesium Chloride, 50mM Potassium Chloride, 50mM Sodium Chloride
10 x Reaction Buffer 8	20mM Tris-HCl (pH 7.4), 10mM Magnesium Chloride
10 x Reaction Buffer 9	200mM Tris Acetate (pH 7.9), 100mM Magnesium Acetate, 50 mM Potassium Acetate
10 x Reaction Buffer 10	100mM Tris-HCl (pH 7.6), 10mM Magnesium Chloride, 150mM Sodium Chloride
10 x Reaction Buffer 11	10mM Tris-HCl (pH 9.0), 12 mM Magnesium Chloride, 100mM Potassium Chloride

Table 2.2. List of general solutions and their composition

SOLUTION	COMPOSITION
1 x TE	10 mM Tris-HCl, 1mM EDTA, adjusted to pH 8.0
10 x TBE	0.89M Tris-Borate, 0.89M Boric acid, 10mM EDTA
20 x SSC	3M Sodium Chloride, 0.3M Tri-Sodium Citrate, adjusted to pH 7.0
0.5M EDTA	0.5M Diaminoethanetetra-acetic acid, adjusted to pH 8.0
10 x MOPS	0.2M Sodium-MOPS, 50mM Sodium Acetate, 10mM EDTA, adjusted to pH 7.0
100 x Denhardts solution	2% (w/v) Bovine Serum Albumin, 2% (w/v) Polyvinylpyrrolidone (ml. wt. 400 000), 2% (w/v) Ficoll (ml. wt. 400 000)
Salmon sperm DNA (purchased from Sigma)	10 mg/ml stock solution. Supplied sonicated and denatured.
Oligo Labelling Buffer (OLB)	Solution O : 0.125M Magnesium Chloride, 1.25M Tris-HCl (pH 8.0) Solution A : 0.95ml solution O, 18µl 2-Mercaptoethanol, 25µl 20mM dATP, 25µl 20mM dTTP, 25 µl 20mM dGTP Solution B : 2M HEPES, adjusted to pH 6.0 Solution C: Hexadeoxyriboneucleotides suspended in 1 x TE. to 90 OD units/ml (purchased from SIGMA) OLB is made by mixing Solutions A,B and C in a ratio of 2:5:3 respectively. It is stored at -20°C
RNA Formaldehyde Sample Buffer (FSB)	50 % (v/v) Formamide, 25% (v/v) Formaldehyde (at 14.8% w/v), 25% (v/v) 10 x MOPS buffer

10 x DNA gel loading buffer	0.25% (w/v) Bromophenol Blue, 0.25% (w/v) Xylene Cyanol FF, 30% (v/v) Glycerol
10 x RNA gel loading buffer	0.25% (w/v) Bromophenol Blue, 0.25% (w/v) Xylene Cyanol FF, 1mM EDTA (pH 8.0), 50% (v/v) Glycerol
DNA extraction buffer	50mM Tris-HCl (pH 9.0), 0.1M EDTA (pH 8.0), 0.2M Sodium Chloride, 1 mg/ml Ribonuclease A
RNA denaturing buffer	4M Guanidine Thiocyanate, 42mM Sodium Citrate, 0.83% (w/v) Lauryl Sarcosine, 0.2mM β-mercaptoethanol
GTE. solution (plasmid mini-preps)	50mM Glucose, 10mM EDTA (pH 8.0), 25mM Tris-HCl (pH 8.0)
Phenol/Chloroform	Phenol was purchased from SIGMA (re-distilled and pre-equilibrated with 100mM Tris-HCl pH 8.0) and mixed with Chloroform and Isoamyl Alcohol in the ratio 25:24:1 respectively. To prevent oxidation, 8-Hydroxyquinoline was added to 0.1% (w/v) and the solution stored at 4°C in the dark
40mM dNTP stock solution (for PCR and Reverse transcription)	10mM dATP, 10mM dCTP, 10mM dGTP, and 10mM dTTP. Made by dilution of 100mM stock nucleotide solutions (pH 7.0) purchased from Pharmacia
10 x Polymerase Chain Reaction (PCR) buffer	200mM Tris-HCl (pH 8.4), 500mM Potassium Chloride
5 x Reverse transcriptase first strand synthesis buffer (GIBCO BRL)	250mM Tris-HCl (pH 8.3), 375mM Potassium Chloride, 15mM Magnesium Chloride
5 x Reverse transcriptase second strand synthesis buffer	94mM Tris-HCl (pH 6.9), 453mM Potassium Chloride, 23mM Magnesium Chloride, 750μM β-Nicotinamide Adenine Dinucleotide, 50mM Ammonium Sulphate

10 x T4 Polynucleotide Kinase buffer	600mM Tris-HCl (pH 7.8), 100mM Magnesium Chloride, 150mM 2-Mercaptoethanol, 33µM ATP
5 x T4 DNA Ligase buffer (GIBCO BRL)	250mM Tris-HCl (pH 7.6), 50mM Magnesium Chloride, 5mM ATP, 5mM Di-thio-threitol, 25% (w/v) Polyethylene glycol (ml. wt. 8000). Stored at -20°C
Deoxyribonuclease I (DNAase I)	20 mg/ml in 50 % (v/v) Glycerol, stored at -20°C
Ribonuclease A (RNAase A)	20 mg/ml in 50 % (v/v) Glycerol, boiled for 5 min to inactivate contaminating DNAase's, and stored at -20°C
Proteinase K	20 mg/ml in 50% (v/v) Glycerol, stored at -20°C
Lysozyme	8 mg/ml in sterile double distilled water, used fresh
1 x TSS. (Preparation of competent cells)	10 % (w/v) Polyethylene Glycol (ml. wt. 8000), 5% (v/v) Dimethyl Sulphoxide, and 50mM Magnesium Chloride in L-Broth. Adjusted to pH 6.5
10 % SDS	10% (w/v) Sodium Dodecyl Sulphate in sterile double distilled water. Filter serilised
10 x Shrimp Alkaline Phosphatase Buffer (USB)	200mM Tris-HCl (pH 8.8), 100mM Magnesium Chloride
Shrimp Alkaline Phosphatase enzyme dilution buffer (USB)	50mM Tris-HCl (pH 8.0)
80% (v/v) Ethanol	80ml Ethanol made up to 100ml with sterile distilled water
TM buffer (small scale phage DNA extraction)	50mM Tris-HCl (pH 7.5), 10mM Magnesium Sulphate
DEPC-water	Sterile distilled water rendered RNAase free by treating with Di-ethyl Pyrocarbonate as per section 2.1.1.

First strand buffer (reverse transcription)	250mM Tris-HCl (pH 8.3), 375mM Potassium Chloride, 15mM Magnesium Chloride
Ethidium Bromide	10mg/ml in sterile distilled water

Table 2.3. Composition of buffers used with Qiagen and Qiaex

QIAGEN SOLUTIONS	COMPOSITION
P1 (Re-suspension Buffer)	50mM Tris-HCl, 10mM EDTA 100µg/ml RNAase A. Adjusted to pH 8.0 and stored at 4°C
P2 (Lysis Buffer)	200mM Sodium Hydroxide, 1% (w/v) SDS
P3 (Neutralisation Buffer)	3.0M Potassium Acetate, adjusted to pH 5.5 and stored at 4°C
QBT (Column Equilibration Buffer)	750mM Sodium Chloride, 50mM MOPS, 15% (v/v) Ethanol, 0.15% (v/v) Triton X-100. Adjusted to pH 7.0
QC (Column Wash Buffer)	1.0M Sodium Chloride, 50mM MOPS, 15% (v/v) Ethanol. Adjusted to pH 7.0
QF (Column Elution Buffer)	1.25M Sodium Chloride, 50mM Tris-HCl, 15% (v/v) Ethanol. Adjusted to pH 8.5

QIAEX SOLUTION	COMPOSITION
QX1 (Solubilisation Buffer)	3M Sodium Iodide, 4M Sodium Perchlorate, 50mM Tris-HCl (pH 7.5), 0.1% (w/v) Sodium Sulphite. Stored at 4°C in the dark
QX2 (First Wash Buffer)	8M Sodium Perchlorate
QX3 (Second Wash Buffer)	70% (v/v) Ethanol, 100mM Sodium Chloride, 10mM Tris-HCl, 1mM EDTA, adjusted to pH 7.5

2.2 MAINTENANCE OF FLY STOCKS AND STRAINS

Table 2.4 summarises the fly strains used.

2.2.1 Drosophila melanogaster

Stocks were maintained at 25°C on 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, cooled to approximately 40°C and poured into bottles. Flies were introduced into bottles only when the food medium had completely set. A fungicide, Nipagin, was added to a final concentration of 4.5µg/L and occasionally antibiotics such as Gentamycin (to 40µg/L) were added. In the event of mite infections, strips of Whatman filter paper were soaked in 3% (v/v) Benzyl Benzoate (in Ethanol), air dried, and placed on top of the cornmeal food. Table 2.5 summarises the development of Oregon R Drosophila melanogaster at 25°C.

2.2.2 Calliphora erythrocephala

Populations of *Calliphora erythrocephala* were maintained at room temperature in mesh cages on a sugar and water diet during most of the adult life cycle, with a fresh protein meal (a meat feed) provided every day subsequent to 4 days after emergence. Eggs laid on this meat (usually on the underside) were transferred to sandwich boxes containing cubed larval media (80g Davis agar, 600g Millac, and 50g autolysed yeast in a total volume of 4L water), which was subsequently placed on a bed of sawdust in a large metal tin (at room temperature). Typically larvae feed on the media for 3 - 5 days before wandering into the sawdust (2 - 3 days prior to pupation), at which point pupae were transferred to jam jars (the sawdust was removed by sieving) and sealed with a paper towel and elastic band until fly emergence was observed. The jam jar containing the newly emerging flies was then placed in a new mesh cage, and the process repeated.

2.2.3 Musca domestica

Essentially the same procedure was used to maintain *Musca domestica* populations as has been described for the maintenance of *Calliphora erythrocephala* populations, except that emerging flies are fed with a proteinaceous meat meal from the first day of emergence.

Table 2.4. List of fly strains and relevant features

STOCK	RELEVANT FEATURES	REFERENCE/SOURCE
Drosophila melanogaster OrR	Wild type strain	Lindsley and Grell (1968)
Calliphora erythrocephala	Wild type strain	Local
Musca domestica	Wild type strain	Local

2.3 COLLECTION OF SEXED FLIES AND DEVELOPMENTAL STAGES

2.3.1 Collection of *Drosophila melanogaster* OrR late 3rd instar larvae and pupae

Egg laying flies were placed in fresh cornmeal food bottles for 4 hr then removed. The bottle was then incubated at 25°C for 96 hr (larvae) or 120 - 240 hr (pupae). Late 3rd instar larvae were picked from the sides of the bottles using a fine paintbrush, placed into an microcentrifuge tube, and either used immediately or frozen in liquid nitrogen and stored at -80°C.

2.3.2 Collection of *Drosophila melanogaster* OrR sexed adults

Adult flies ranging in age from newly eclosed to 10 days old were anaesthetised with di-ethyl ether and sexed according to external morphology (abdominal pigmentation and genital morphology). The sexed flies were then transferred to microcentrifuge tubes, used immediately or frozen in liquid nitrogen and stored at -80° C.

2.3.3 Collection of *Calliphora erythrocephala* sexed adults

Adult flies ranging in age from newly eclosed to 20 days old were anaesthetised with di-ethyl ether, sexed, and either used immediately or placed into a falcon bluemax tube and frozen in liquid nitrogen. Such flies were stored at -80°C until required. Flies were sexed according to eye morphology (eyes of males are adjacent whereas eyes of females are spaced approximately 3mm apart).

2.3.4 Collection of *Musca domestica* sexed adults

Similar procedures to those described for the collection of *Calliphora erythrocephala* sexed adults were utilised.

Table 2.5. Developmental stages of Drosophila melanogaster Oregon R strain at 25°C (Bownes and Dale, 1982)

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 in the previously syncytial blastoderm
3.5	0	Gastrulation begins
6 - 8	0	Segmentation visible
18	0	Larval differentiation nearly complete
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	Prepupal moult
132	5	Pupation. Eversion of imaginal discs
216 - 240	9 - 10	Emergence of adult from pupal case

2.4 MEDIA AND MICROBIAL STRAINS, VECTORS AND LIBRARIES

2.4.1 Media preparation

Table 2.6 lists all appropriate media, which was prepared in distilled water and sterilised by autoclaving (15 psi/ 15 min). Any further supplements, such as vitamins or sugars were dissolved in distilled water (unless otherwise stipulated by the supplier) to a suitable concentration, filter sterilised, and added to the media to obtain the required final concentration.

If necessary, antibiotics were added to the media at the relevant concentration (Table 2.7.). All antibiotics were stored at -20°C.

2.4.2 Relevant microbial strains, vectors and libraries used

Microbial strains used are listed in Table 2.8.

Vectors used for the construction of recombinant molecules are listed in Table 2.9.

Libraries used are listed in Table 2.10.

Bacterial stocks were maintained either on appropriate plates at 4°C, as stabs stored at room temperature out of direct sunlight, or as glycerol stocks at -80°C.

Lambda bacteriophage were maintained at 4°C in SM buffer with a few drops of chloroform added to prevent microbial growth, or as -80°C stocks by supplementing the SM buffer with Dimethyl Sulphoxide to 7% (v/v).

Genomic DNA, Plasmid DNA and λ Bacteriophage DNA were dissolved in 1 x TE buffer and stored at 4°C.

All RNA was stored at -80°C, either dissolved in DEPC-water or as ethanol precipitates. Samples were thawed on ice prior to use.

Table 2.6. Media recipes

MEDIA TYPE	COMPONENTS
L-Broth (Luria-Bertani Broth)	10g Difco Bacto-tryptone, 5g Difco Bacto-yeast extract, 5g Sodium Chloride, adjusted to pH 7.2
L- Agar	16g Difco Bacto-tryptone, 10g Difco Bacto-yeast extract, 5g Sodium Chloride, adjusted to pH 7.2
BBL-Bottom	10g Baltimore Biological Laboratories Trypticase, 10g Difco Bacto agar, 5g Sodium Chloride
BBL-Top	10g Baltimore Biological Laboratories Trypticase, 6.5g Difco Bacto agar, 5g Sodium Chloride
SM buffer	5.8g Sodium Chloride, 2g Magnesium Sulphate, 50mM Tris-HCl (pH 7.5), 0.01% Gelatine
2 x YT	16g Bacto-tryptone, 10g Bacto-yeast extract, 5g Sodium Chloride

Table 2.7. Antibiotic supplements and their working concentrations

ANTIBIOTIC	STOCK SOLUTION	WORKING CONCENTRATIONS
Ampicillin	100 mg/ml in SDW	50µg/ml
Kanamycin	10 mg/ml in SDW	50µg/ml
Tetracycline	5 mg/ml in Ethanol	50µg/ml

Table 2.8. Microbial strains and λ Bacteriophage used

BACTERIAL/ LAMBDA STRAIN	GENOTYPE	RELEVANT USE	REFERENCE
MRF' XL1-BLUE	Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, endA1, supE44, thi-1, recA1, gyrA96, relA1, lac, [F' proAB, $lacI^qZ\Delta M15$, Tn10 (tet^r), Amy, cam^r]^c	Propagation of λZAP II libraries and pBluescript phagemids. F' episome was maintained by plating onto tetracycline plates (12.5μg/ml)	Please refer to Stratagene catalogue
SOLR	$e14^-$(mcrA), Δ(mcrCB-hsdSMR-mrr)171, sbcC, recB, recJ, umuC::Tn5 (kan^r), uvrC, lac, gyrA96, relA1, thi-1, endA1, λ^r, [F' proAB, $lacI^qZ\Delta M15$], Su^- (non-suppressing)	Used to rescue pBluescript phagemids from λZAP II libraries. Maintained on Kanamycin plates (50μg/ml)	Please refer to Stratagene catalogue
NM767	Restriction negative host	Propagation of recombinant λgem-11 bacteriophage. P2 Lysogenic	Kind donation from Noreen Murray (University of Edinburgh)
DL709	Restriction negative host	Propagation of wild type and recombinant λgem-11	Kind donation from David Leach (University of Edinburgh)

EXASSIST		Used in the rescue of pBluescript phagemids from λZAP II libraries. EXASSIST is unable to infect SOLR cells and thus only cells containing pBluescript phagemids are rescued	Please refer to Stratagene catalogue

Table 2.9. Plasmids, phagemids and Lambda Bacteriophage vectors used

VECTOR	COMMENTS	REFERENCE
pBluescript phagemid	General cloning vector. Also used to generate single stranded templates from either + or - strand. Confers Ampicillin resistance in host bacteria	Short, J.M. et al (1988) Nucleic Acids Research 16 : 7583 - 7600 Alting-Mees, M.A. and Short, J.M. (1989) Nucleic Acids Research 17 : 9494
λgem-11	Vector for the construction of genomic libraries. Derivative of EMBL 3/4	Frischauf, A.M. et al (1983) J. Mol. Biol. 170 : 827
λZAP II	Vector used in the construction of cDNA libraries. Inserts can be rescued as pBluescript phagemids using the EXASSIST/SOLR procedure	Short, J.M. et al (1988) Nucleic Acids Research 16 : 7583 - 7600

Table 2.10. Genomic and cDNA Libraries used

LIBRARY	VECTOR	SOURCE
Calliphora erythrocephala genomic library	λgem-11	This Study
Musca domestica genomic library	λgem-11	Constructed by Claudia Tortiglione, our laboratory.
Musca domestica ovarian cDNA library	λZAP II	Kindly donated by D. Bopp (University of Zürich)

2.5 GENERAL MANIPULATIONS OF NUCLEIC ACIDS

2.5.1 Precipitation of nucleic acids

Sodium Acetate (pH 4·8) was added to a final concentration of 0.3M, the solution mixed gently, and either 2 volumes of 100% (v/v) Ethanol or 0·6 volumes of 100% (v/v) Propan-2-ol were added. Generally the nucleic acid was precipitated on ice for 15 min, pelleted by centrifugation (13 000 rpm, 10 min), washed with 80% (v/v) Ethanol and dried under vacuum. The sample was then dissolved in an appropriate volume of either 1 x TE or SDW depending on subsequent applications.

If small quantities of nucleic acid were precipitated (< 1μg/ml) the incubation on ice was replaced with an overnight incubation at -20°C.

2.5.2 Deproteinisation by Phenol/ Chloroform extraction

An equal volume of Phenol/ Chloroform was added and the solution mixed either by inversion (high molecular weight samples) or by vortexing. The sample was then centrifuged (13 000 rpm, 3 min), and the aqueous phase transferred to a fresh tube. This extraction was repeated until the interface after centrifugation was not visibly particulate. To remove traces of Phenol, the solution was finally extracted with Chloroform/Isoamyl alcohol (24:1 respectively).

2.5.3 Reverse transcription of total RNA and poly-A^+ RNA samples

Total RNA (10 - 100μg) or Poly-A^+ RNA (1-5μg) was annealed to Oligo-dT $p(dT)_{12-18}$ primers (1μg) by incubating at 70°C (10min; 11μl total volume) and allowing to cool to room temperature. Reverse transcription was then achieved by adjusting the

solution to 1 x First Strand Buffer, 10mM DTT, 200mM dNTP's, Superscript II (200U; Gibco BRL) in a total volume of 20µl, and incubating at 37°C for 60 min. Subsequently, 1 x TE (80µl) was added and the mixture phenol/chloroform extracted, Ethanol precipitated, and resuspended to an estimated concentration of 1µg/µl.

2.5.4 Estimation of nucleic acid concentrations

Spectrophotometry was used to assess the concentration and purity of nucleic acid preparations. Table 2.11. lists the typical concentrations of nucleic acids when reading the absorbence of the sample at λ_{260nm}.

Nucleic acid was judged to be free of contaminating protein if the A_{260}:A_{280} ratio was greater than or equal to 1·8.

RNA preparations were tested for Guanidinium salt contamination carried over from the isolation procedure, where an A_{260}:A_{230} ratio of greater than or equal to·2·0 was judged to be sufficiently pure.

Approximate nucleic acid concentrations were also deduced by comparing the relative fluorescence of samples to those of a known concentration during agarose gel electrophoresis.

Table 2.11. Absorbence of nucleic acid solutions and their inferred concentrations

NUCLEIC ACID	ABSORBENCE (A_{260})	CONCENTRATION(µg/ml)
Double stranded DNA	1.0	50
Single stranded DNA	1.0	40
Single stranded RNA	1.0	40
Oligonucleotides/ Primers	1.0	40

2.6 DNA EXTRACTION PROTOCOLS

2.6.1 Genomic DNA extraction

A modified version of the procedure described by Towner, P. (1991) was used.

Approximately 5 - 10 g of tissue was homogenised in liquid nitrogen using a mortar and pestle, and the resulting fine powder suspended in 17·5ml DNA extraction buffer. This solution was then adjusted to 1% (w/v) SDS and 0·5mg/ml Proteinase K, mixed gently by inversion, and incubated at 55°C for 16 hr. Subsequently the solution was transferred to a 500ml beaker containing an equal volume of Phenol/Chloroform , mixed gently by swirling, and the phases left to interact for 3 hr at room temperature with occasional agitation. The phases were then separated by centrifugation (3 000g, 10 min, 4°C) and the aqueous phase transferred to a falcon tube before adding an equal volume of fresh Phenol/Chloroform. The mixture was mixed gently by inversion, the phases separated as previously described, and the aqueous phase transferred to a 50ml beaker. Nucleic acid was then precipitated by the addition of 0·1 volume of 3M Sodium Acetate (pH 4·8) and 0·8 volume 100% (v/v) Propan-2-ol, followed by gentle swirling. High molecular weight DNA was seen as a clump of material, which was transferred to a falcon tube containing 10ml 80% (v/v) Ethanol (to remove contaminating salt) using a glass Pasteur pipette and allowed to rest in the solution for 5 min. The DNA was then transferred to an empty falcon tube where it was allowed to air dry (approximately 30 min), before re-suspending in 1-2ml of 1 x TE

2.6.2 Plasmid DNA mini extraction (Stephen *et al*, 1990)

A small scale liquid culture (2.9.1) of the transformed bacterial strain was established. The following day cells from 1·5ml of culture were harvested by centrifugation (5 000 rpm, 3 min, 4°C), the supernatant removed by aspiration, and the cells re-suspended in 200µl GTE solution. Cell lysis buffer [0·2M Sodium Hydroxide, 1% (w/v) SDS] was then added (400µl) before placing on ice for 5 min. Cellular protein and chromosomal DNA were co-precipitated by the addition of 300µl 3M Potassium Acetate (pH 4·8) followed by gentle inversion, and pelleted by centrifugation (17 000 rpm, 5 min). The supernatant was transferred to a fresh tube, Ethanol precipitated (2.6.1) without the addition of 0·1 volume 3M Sodium Acetate, and the plasmid DNA re-suspended in 20µl 1 x TE.

2.6.3 Midi plasmid DNA extraction

A 20ml culture of the transformed bacterial strain was established (2.8.1). The next day the cells were harvested by centrifugation (4 000 rpm, 5 min, 4°C), the

supernatant discarded, and the cells re-suspended in 400µl GTE solution. Cell lysis was achieved by the addition of 0·6M Sodium Hydroxide/ 3% (w/v) SDS (100µl) followed by gentle inversion to mix and a 10 min incubation on ice. Subsequently, cellular protein and chromosomal DNA was co-precipitated by adding 150µl 3M Potassium Acetate pH 4·8, mixing by gentle inversion, and incubating for 10 min on ice. The precipitated material was removed by centrifugation (13 000 rpm, 10 min, 4°C), and the supernatant transferred to a fresh tube. The plasmid DNA was precipitated by adding 1ml 100% (v/v) Ethanol followed by inversion to mix, pelleted by centrifugation (13 000 rpm, 5 min, 4° C), washed briefly in 500µl 80% (v/v) Ethanol, vacuum dried and re-suspended in 400µl 1 x TE. Ribonuclease A was added to 1µg/µl and the solution incubated at 37°C for 60 min to remove contaminating RNA. The sample was then Phenol/ Chloroform extracted (2.5.2), Ethanol precipitated (2.5.1) and finally re-suspended in 50µl 1 x TE.

2.6.4 Plasmid maxi DNA extraction

A 500ml culture of the transformed bacterial strain was established (2.8.1). Cells were harvested by centrifugation (3 000 rpm, 10 min, 4°C), the supernatant discarded, and the cells re-suspended in 5 ml GTE solution. Cell lysis was achieved in two stages, initially Lysozyme (0·5µg/µl) was added (and the suspension mixed by vortexing) before incubating on ice for 5 min, and subsequently 10 ml of 0·2M Sodium Hydroxide/ 1% (w/v) SDS was added (mixed in by gentle inversion) before a second incubation on ice for 5 min. Cellular protein and chromosomal DNA was co-precipitated and removed by adding 7.5ml ice cold 3M Potassium Acetate (pH 5·2), mixing by gentle inversion, incubating on ice for 10 min, and finally centrifuging (17 000 rpm, 30 min, 4°C). The supernatant was transferred to a fresh tube, 2M Potassium Chloride (2·5ml) added before incubating on ice for 10 min, the plasmid DNA precipitated by the addition of 0·6 volume of 100% (v/v) Propan-2-ol and subsequently harvested by centrifugation (13 000rpm, 10 min, 4°C). The supernatant was discarded before washing the DNA briefly with 80% (v/v) Ethanol, vacuum drying, and re-suspending in 4·5ml 1 x TE. Caesium Chloride (10g) and Ethidium Bromide (0·5ml of 15mg/ml solution) were added, the sample shaken until the salt dissolved, and the solution made up to a final volume of 10ml before loading into quick seal Ti50 rotor tubes (balanced to 0·1g). The samples were centrifuged (38 000rpm, 18°C, 48hr) and supercoiled plasmid DNA (lower band) harvested using a 21G syringe by visualising under UV. Ethidium Bromide was removed from the solution by extracting with an equal volume of Butan-2-ol (TE saturated) until the Butan-2-ol phase was clear, the sample volume made up to 5ml with 1 x TE, and the plasmid DNA finally Ethanol precipitated (2.5.1) before re-suspending in 500µl 1 x TE.

2.6.5 Qiagen midi DNA extraction

A 100ml culture of the transformed bacteria was established (2.8.1),the cells harvested by centrifugation (3 000 rpm, 10 min, 4°C), and subsequently re-suspended in 4ml buffer P1. Buffer P2 (4ml) was added and the mixture gently inverted, cell lysis being apparent by the solution becoming clear rather than turbid, before incubating on ice for 5 min. Cellular protein and chromosomal DNA was precipitated by adding 4ml buffer P3 (ice cold), mixing by inversion, and incubating on ice for 15 min before pelleting by centrifugation (17 000 rpm, 30 min, 4°C). During this centrifugation step a Qiagen tip-100 was equilibrated with 4ml buffer QBT by applying to the column and allowing to drain by gravity flow. The supernatant from the centrifugation step was then promptly loaded onto the column, allowed to drain by gravity flow, and the column washed with 2 x 10ml buffer QC. Plasmid DNA was harvested by collecting the eluate after applying 5ml buffer QF to the column, precipitating the DNA by adding 0·7 volume 100% (v/v) Propan-2-ol, washing with 80% (v/v) Ethanol (5ml), air drying (15 min) and finally re-suspending in 100µl 1 x TE.

2.6.6 Small scale λ bacteriophage DNA extraction

Lambda bacteriophage were added to 0·2ml plating cells (2.8.4) to give a multiplicity of infection less than or equal to 1·0 and the phage allowed to absorb by placing at 37°C for 15 min. These cells were then used to prime a 10ml L-Broth culture [supplemented to 10mM Magnesium Chloride and 0.1% (w/v) Glucose], and the culture incubated at 37°C with good aeration (250 rpm) until lysis was apparent by the presence of characteristic 'rope-like' structures. Cellular debris was removed by centrifugation (13 000 rpm, 10 min, 4°C) and the supernatant transferred to a fresh tube before adding DNAase I (1µg/ml) and RNAase A (1µg/ml) and incubating at 37°C for 60 min. Phage were pelleted by centrifugation in a TST41.14 swing out rotor (30 000 rpm, 60 min, 18° C), the supernatant discarded, and the phage re-suspended in TM buffer (190µl) before adding Proteinase K (1µg/µl) and incubating at 65°C for 60 min. The sample was then Phenol/Chloroform extracted (2.5.2), Ethanol precipitated (2.5.1) and re-suspended in 50µl 1 x TE.

2.7 ISOLATION OF TOTAL CELLULAR AND POLYADENYLATED RNA

2.7.1 Extraction of total cellular RNA

A modified protocol (Promega protocols and applications guide) as initially described by Chomczynski and Sacchi (1987) was employed in which DNA is removed by acidic Phenol extraction.

Essentially, tissues were homogenised in liquid nitrogen using a mortar and pestle (pre-chilled) until a fine powder was obtained. This powder was transferred to a falcon tube containing RNA Denaturing Solution (12ml per gram of tissue), 0·1 volume of 2M Sodium Acetate (pH 4·0) added, and the solution shaken vigorously for 1 min. An equal volume of Phenol/Chloroform (pre-equilibrated to pH 4·0) was added before shaking the solution again, the sample incubated on ice for 10 min with occasional inversion to mix, and cellular protein and DNA removed by centrifugation (18 000 rpm, 30 min, 4°C). The aqueous phase was transferred to fresh falcon tube, the Phenol/Chloroform extraction repeated, and the aqueous phase promptly transferred to a corex tube. RNA was then precipitated by adding an equal volume of 100% (v/v) Propan-2-ol, pelleted by centrifugation (13 000 rpm, 10 min, 4°C), washed briefly with 80% (v/v) Ethanol, dried under vacuum, and re-suspended in 5ml DEPC-water. The RNA was then re-precipitated (2.5.1), and re-suspended to approximately 2µg/µl in 1 x TE.

2.7.2 Isolation of polyadenylated RNA

Polyadenylated messenger RNA was isolated using the PolyATtract system (Promega) as per manufacturers instructions. Essentially, 0.1 - 1.0 mg total RNA was suspended in a total volume of 500µl DEPC-treated sterile distilled water, and incubated at 65°C (10 min). Biotinylated-Oligo(dT) probe (3µl; 50 pmol/µl) and 20 x SSC (13µl) were then added, and the solution allowed to cool to room temperature. The RNA solution was then added to pre-washed Streptavidin-Paramagnetic particles (washed three times in 0.5 x SSC [0.3ml/wash], finally resuspended in 0.1ml 0.5 x xSSC) and allowed to anneal at room temperature (10 min). The supernatant was then removed and retained for further analysis (paramagnetic particles are captured using a magnetic stand), and the particles washed (four times with 0.1 x SSC [0.3ml/wash]). The aqueous phase was then removed entirely before eluting the mRNA in DEPC-SDW (two washes of the particles, firstly with 0.1ml and then with 0.15ml), ethanol precipitated, and resuspended in RNAase free water to an estimated concentration of 0.5µg/ml.

2.8 GROWTH AND MAINTENANCE OF BACTERIAL STOCKS

Where necessary, growth media was supplemented with antibiotic at the concentrations described in Table 2.7.

2.8.1 Growth of bacteria in liquid culture

Small scale cultures were prepared by inoculating L-Broth (5ml) with an isolated bacterial colony and growing overnight at 37°C with good aeration (250 rpm).

Larger cultures were prepared by inoculating L-Broth (20-500ml) with 0·01 volume (relative to the volume of media being inoculated) of a small scale culture and growing overnight at 37°C with aeration (250 rpm).

2.8.2 Growth of bacteria on agar plates

Bacteria were either grown as 'streaks' or as a 'lawn' of colonies on L-Agar plates. In the case of streaks, a single colony was spread sequentially around the plate (flaming the inoculating loop between each streak) as per standard microbial technique. Lawns were prepared by spreading a small volume (less than or equal to 200µl) of bacterial culture over the surface of the L-Agar plates using a glass spreader sterilised in 80% (v/v) Ethanol (flamed before used to remove traces of Ethanol). In both cases plates were incubated overnight at 37 °C.

If, as in the case of XL1-BLUE bacterial cells transformed with pBluescript phagemids, a β-Galactosidase based recombinant selection strategy was employed, prior to spreading the bacterial cells on the plate IPTG and X-gal were added to the media. This was achieved by spreading 100µl 0·1M IPTG and 40µl 2% (v/v) X-gal over the surface of the plates (using the glass spreader) 5 min before spreading the bacteria.

2.8.3 Preparation of competent cells

The protocol described in Chung *et al* (1989) was employed for the preparation of competent cells.

A large scale culture of the required bacterial strain was established (2.8.1), which was grown at 37°C with good aeration (250rpm) until early exponential growth was obtained (OD_{600} between 0·3 and 0·4). Cells were then harvested by centrifugation (1 000 g, 10 min, 4°C) and re-suspended in 10ml ice cold 1 x TSS (i.e. 0·1 volume of the original culture). These cells were aliquoted into microcentrifuge tubes (100µl per microcentrifuge tube), frozen in a dry ice/ Methanol bath, and either used immediately or stored at -80°C until required. Stored cells were thawed on ice prior to use.

2.8.4 Preparation of plating cells for the propagation of *Lambda* bacteriophage

A small scale culture of the required bacterial strain was established (2.8.1) in L-Broth supplemented to 0·2% (w/v) Maltose and 10mM Magnesium Sulphate. The following day this culture was used to prime a large culture (100ml; same supplements), grown for 6hr at 37°C with good aeration (250rpm), and the cells harvested by centrifugation (4 000 rpm, 10 min, 4°C). The cells were re-suspended in 10mM Magnesium Sulphate, and remained viable for 2 weeks if stored at 4°C.

2.8.5 Long term storage of bacterial strains as stabs

An isolated colony was picked using a sterile straight wire and stabbed into L-Broth supplemented to 0.6% (w/v) agar. This culture was grown overnight at 37°C with the caps slightly loose, the caps then tightened, the tubes sealed with parafilm, and the stabs stored at room temperature out of direct sunlight. Stabs were revived by streaking some of the agar onto L-Agar plates and growing as per normal.

2.8.6 Long term storage of bacterial strains as glycerol stocks

A small scale overnight culture of the required bacterial strain was established (2.8.1.). The following day 1ml of the culture was transferred to an microcentrifuge tube, centrifuged (4 000 rpm, 3 min, 4°C), the supernatant discarded and the cells re-suspended in 1ml 50% (v/v) Glycerol (in L-Broth). This re-suspension was then frozen in a dry ice/ Methanol bath and the cells stored at -80°C. To rescue bacterial cells stored as glycerol stocks, the suspension was allowed to thaw slightly such that a few ice particles could be scraped from the surface of the suspension using a sterile loop, which were subsequently streaked onto L-Agar plates and grown as per normal.

2.9 GROWTH AND MAINTENANCE OF *LAMBDA* BACTERIOPHAGE

2.9.1 Growth of λ bacteriophage in liquid culture

Lambda bacteriophage were added to 0·01 volume bacterial plating cells (e.g. for a 100ml culture, 1ml plating cells were infected) to give a multiplicity of infection of less than 1·0, and allowed to adsorb by placing at 37°C for 15 min. These cells were then used to inoculate the required volume of L-Broth (5-500 ml), and the culture grown with good aeration (250 rpm) at 37°C until lysis was apparent by the presence of rope-like structures of bacterial cell debris. Chloroform was added (1ml per 25ml culture) and the culture incubated at 37°C for a further 15 min to obtain complete cell lysis. Cellular debris was removed by centrifugation (15 000 rpm, 10 min, 4°C), and the supernatant stored at 4°C until required, with the addition of a few drops of chloroform to prevent microbial growth.

2.9.2 Growth of λ bacteriophage on agar plates

Stocks of *lambda* bacteriophage were diluted to give the required number of pfu (e.g. 100pfu for a 50mm diameter plate), added to 100µl of relevant plating cells (2.8.4.), and adsorbed at 37°C for 15 min. Subsequently, 3ml BBL-Top Agar (pre-molten and cooled to 50°C) was added, the mixture poured over BBL-Bottom plates (pre-dried) and allowed to set, and the plates then incubated at 37°C until plaques of approximately 1mm diameter were obtained (typically 6 - 7hr).

A similar procedure was used when establishing large plates for screening (23 x 23cm). Essentially, approximately 300 000 pfu were mixed with 3ml plating cells and adsorbed at 37°C for 15 min. BBL-Top Agar was then added (to 30ml), the mixture poured a large BBL-Bottom plate (pre-dried) and allowed to set, and the plates incubated until near confluent lysis was achieved.

All plates to be screened were placed at 4°C for at least 1hr prior to taking lifts to prevent the Top-Agar from sticking to the membranes.

2.9.3 Storage of λ bacteriophage as liquid lysates

Typically a small aliquot (1ml) of a *lambda* liquid culture (2.9.1.) was transferred to an microcentrifuge tube (after the removal of cellular debris), a few drops of chloroform added to prevent microbial growth, and the lysate stored at 4°C until required.

Alternatively, a high density 50mm *lambda* plate (2.9.2.) was overlaid with 3ml SM buffer, incubated at 4°C with gentle agitation for 1hr, and the overlay collected. An aliquot of this overlay (1.5ml) was transferred to an microcentrifuge tube, centrifuged (13 000 rpm, 5 min, 4°C) to remove contaminants, the supernatant (1ml) transferred to a fresh microcentrifuge tube, and a few drops of chloroform added before storing at 4°C until required.

If long term storage of the phage lysate was required (greater than 1 year) the lysate was made 7% (v/v) Dimethyl Sulphoxide, frozen in a dry ice/ Methanol bath, and stored at -80°C. The stocks were revived by thawing on ice slightly and removing a small aliquot, which was subsequently used to generate a fresh *lambda* plate (2.10.2.).

2.9.4 Determination of λ bacteriophage stock titre's

Serial dilution's of the phage were established (in SM buffer) and plated (2.9.2.). Plates containing a suitable number of plaques were identified, the number of plaques on each plate counted, and an average titre (as pfu/ ml) was determined.

2.9.5 ExAssist rescue of pBluescript phagemids from *Lambda* ZAP II bacteriophage

A single plaque was cored and placed into SM buffer (500µl) supplemented with Chloroform (20µl). The sample was then vortexed (1min), and the liberated phage particles allowed to diffuse through the solution (1-2hr at room temperature, or overnight at 4°C). This phage stock was then adsorbed to XL1-Blue plating cells (200µl of an OD_{600}=1.0 stock), ExAssist helper phage added (>1 x 10^6 pfu stock; 1µl) and the mixture incubated at 37°C for 15 min. The mixture was then supplemented with 2 x YT

(3ml) and incubated at 37°C for 2.5hr with shaking, before heat inactivation (70°C, 20 min) and centrifugation (4000g, 15min). The supernatant was decanted to a sterile tube, and stored at 4°C (for up to 2 months). The phagemid particles were rescued by adsorption of this stock (1µl) with SOLR plating cells (200µl of an OD_{600}=1.0 stock; 37°C, 15min), and plated onto LB-Ampicillin agar plates.

2.10 GENERAL MANIPULATION OF DNA

2.10.1 Restriction enzyme digestion of DNA

The DNA to be digested (dissolved in 1 x TE) was placed into an microcentrifuge tube, and 10 x restriction enzyme buffer added to achieve a 1 x concentration in the final reaction volume. Sterile distilled water was added to the required volume (i.e. final reaction volume minus Xµl, where X is the amount of enzyme to be added), and restriction enzymes added (3U per µg DNA). The reaction was allowed to proceed as per manufacturers instructions (typically at 37°C for 1 hr), and either stopped by adding 0·1 volume of 10 x DNA gel loading buffer (for analysis by electrophoresis) or by Phenol/ Chloroform extracting (2.5.1.) and Ethanol precipitating (2.5.2.).

2.10.2 De-phosphorylation of DNA

The DNA was restriction digested (2.10.1.) in a volume of 25µl, but rather than stop the reaction 19µl SDW and 5µl 10 x de-phosphorylation buffer were added. Phosphatase was then added at the manufacturers recommended concentration, and the tubes incubated at 37°C for a further 60 min. The samples were then Phenol/ Chloroform extracted (2.5.2.) and Ethanol precipitated (2.5.1.), and re-suspended in SDW prior to ligation.

Both Calf Intestinal (Boehringer Mannheim) and Shrimp (USB) Phosphatases were used.

2.10.3 Ligation of DNA molecules

Several ratio's of Insert DNA:Vector DNA were established (e.g. 1:1, 3:1, and 6:1) using equimolar DNA solutions. T4 DNA Ligation buffer (5x) and SDW were added to achieve a final 1 x concentration (minus 1µl for enzyme), and 1µl Ligase (0·1 U/µl for sticky ends, 1 U/µl for blunt ends) added. Ligation was then allowed to proceed by either incubating at room temperature for 1 hr (sticky ends) or at 4°C overnight (blunt ends).

2.10.4 Transformation of *E. coli* competent cells with plasmid DNA

Plasmid DNA (10-100ng in a volume of less than 10µl) was added to 100µl competent cells (2.9.3.) and the mixture incubated on ice for 30 min. L-Broth (0.9ml) was then added, the cells incubated at 37°C for 1 hr, and 200µl grown on suitable L-Agar plates (2.8.2.).

2.10.5 Packaging of λ bacteriophage DNA

Lambda DNA was packaged using Gigapack II packaging extracts as per manufacturers instructions (Stratagene). Essentially an equal number of sonic and freeze/thaw extracts (stored at -80°C) were placed on ice (each λ DNA sample requires one of each tube type for packaging). The freeze/thaw extract was then quickly thawed (between fingers; until just prior to the loss of ice crystals) before adding the λ DNA sample (0.1 - 5µg) and returning to ice. A 15µl aliquot of the sonic extract was then immediately added to this tube, pipetting gently to mix, and the mixture incubated at room temperature for 2 hours. SM buffer was then added, and cellular debris precipitated by the addition of Chloroform (20µl). This debris was sedimented by brief centrifugation, and the supernatant transferred to a new microcentrifuge tube before storing at 4°C.

2.10.6 QIAquick purification of DNA fragments from agarose gels

The fragment of interest was excised from the gel with a clean sharp scalpel, placed in an microcentrifuge tube, and 3 gel volumes of QX1 buffer added before incubating at 50°C (10 min, gentle agitation every 2-3 min) to dissolve the agarose particles. 100% (v/v) Propan-2-ol was then added (1 gel volume) and the sample loaded into a QIAquick spin column prior to centrifugation (13 000rpom, 1 min). The effluent was discarded, buffer PE added to the QIAquick spin column reservoir (0.75ml), the sample re-centrifuged (13 000rpm, 1 min), and the new effluent also discarded. The column was then centrifuged (13 000rpm, 1min) to remove traces of buffer PE, the QIAquick spin column transferred to a new microcentrifuge tube, and the DNA fragment eluted by the addition of 50µl 10mM Tris-HCl (pH 8.5) followed by centrifugation (13 000rpm, 1 min).

2.11 AGAROSE GEL ELECTROPHORESIS OF NUCLEIC ACIDS

2.11.1 Separation of DNA molecules

Table 2.12. details the separation range of DNA molecules at various percentages of agarose. DNA samples were made 1 x with respect to DNA gel loading buffer (10x). Gels were prepared by dissolving agarose (in a microwave) in 1 x TBE to yield the

required percentage gel, cooling the solution to approximately 65°C, adding Ethidium Bromide (to 0·5µg/ml), and finally pouring into gel trays and allowing to set. Gels were run in 1 x TBE gel buffer within the apparatus manufacturers specifications for maximum current, although for good resolution gels were run at low current (e.g. 20 mA for midi apparatus) overnight.

2.11.2 Separation of RNA molecules

Typically, gels with a final percentage agarose concentration of between 1·3 and 1·5 were employed. Gels were prepared by dissolving the agarose in 1 x MOPS (in a microwave), cooling to approximately 50°C, and adding Formaldehyde to 17.3% (v/v) such that the final volume gave the desired percentage agarose concentration. Gel mixes were then mixed by swirling, poured into gel trays, and allowed to set for a minimum of 30 minutes. Gels were run in 1 x MOPS buffer at similar voltages to those employed for electrophoresis of DNA.

Samples were prepared for loading by adding DEPC-SDW to 10µl, incubating at 65°C for 10 min, and adding an equal volume of Formaldehyde sample buffer (FSB). Ethidium Bromide (5 ng per sample) and RNA gel loading buffer (to 1x concentration) were added prior to loading on the gel.

Table 2.12. Characteristic separation range of nucleic acids at various percentage agarose concentrations (from Andrews, A.T. [1991] in Ess. Mol. Biol. vol. I)

PERCENTAGE AGAROSE	RELATIVE SIZE OF NUCLEIC ACID MOLECULES SEPARATED (Kb)
0· 3	1· 0 - 70
0· 5	0· 7 - 45
0· 8	0· 4 - 20
1· 0	0· 3 - 10
1· 2	0· 2 - 8
1· 5	0· 2 - 6
2· 0	0· 1 - 5

2.12 POLYMERASE CHAIN REACTION PROCEDURES

2.12.1 Optimisation of PCR procedures

In all cases, optimisation was initially performed using positive control template DNA samples, and then further optimised (if necessary) using other templates in the analysis. Typically, melting temperatures (T_m) of the primers were estimated using the formula ([4 x (G + C content)] + [2 x (A + T content)]), and an initial T_m five degrees Celsius below the minimal primer T_m was used as a starting annealing temperature. Magnesium Chloride titrations (0.5 mM to 8.0 mM) were then tested at incremental temperatures until a defined product was obtained. It should be noted that when using redundant oligonucleotide primers, T_m values were estimated on the basis of the primer with the highest (A+T) content (relative to those possible primer permutations).

2.12.2 PCR amplification using redundant *dsx* oligonucleotides

Cross-species amplification using redundant primers designed to *dsx* sequences was performed in a reaction mixture (1 x PCR Buffer [Gibco BRL], approximately 20 picomoles per µl each primer [Primers A_1, A_2, and B; see results section I), 'x' µl template (please refer to description in section I, figure 3.1.8), 1mM $MgCl_2$, 200nM each dNTP, 2.5U TAQ DNA Polymerase (Gibco BRL), and sterile distilled water to a total volume of 50µl) overlaid with approximately 30µl mineral oil. These reaction mixtures were then subjected to four successive cycling stages (STAGE I : 94°C [3 min], 43°C [1 min], 72°C [1 min], 1 x cycle; STAGE II : 94°C [1 min], 43°C [1 min], 72°C [1 min], 4 x cycle; STAGE III : 94°C [1 min], 65°C [1 min], 72°C [1 min], 29 x cycle; STAGE IV : 94°C [1 min], 65°C [1min], 72°C [10 min], 1 x cycle). Products from the amplification were analysed by agarose gel electrophoresis.

2.12.3 PCR amplification using redundant *yp* oligonucleotides

Cross-species amplification using primers designed to conserved *yp* sequences was performed in a reaction mixture (1 x PCR Buffer [Gibco BRL], approximately 20 picomoles per µl each primer [Primers F^{YP} and R^{YP}; see results section II), 'x' µl template (please refer to description in section II, figure 3.2.3), 4mM $MgCl_2$, 200nM each dNTP, 2.5U TAQ DNA Polymerase (Gibco BRL), and sterile distilled water to a total volume of 50µl) overlaid with approximately 30µl mineral oil. These reaction mixtures were then subjected to four successive cycling stages (STAGE I : 93°C [3 min], 58°C [30 sec], 72°C [20 sec], 1 x cycle; STAGE II : 94°C [45 sec], 58°C [30 sec], 72°C [20 sec], 4 x cycle; STAGE III : 93°C [45 sec], 70°C [30 sec], 72°C [20 sec], 29 x cycle; STAGE IV : 93°C

[45 sec], 70°C [30 sec], 72°C [10 min], 1 x cycle). Products from the amplification were analysed by agarose gel electrophoresis.

2.13 CONSTRUCTION OF NESTED DELETIONS

Nested deletions were constructed using an Exonuclease III/ Mung bean nuclease deletion kit (Stratagene) as per manufacturers instructions. Approximately 30µg plasmid DNA (i.e. the recombinant *M. domestica yp* cDNA clones) was digested to completion with *Sal*I restriction endonuclease in a total volume of 500µl, and the reaction mixture incubated at 75°C (15 min) to inactivate the endonuclease. The 5' overhang termini were then filled in with deoxythioderivatives by the addition of 2µl of a thio-dNTP mixture (1mM stock) and Klenow polymerase (5U) followed by a 10 minute incubation at room temperature. The reaction mixture was then phenol/chloroform extracted, ethanol precipitated, and re-suspended in 1 x *Sma*I buffer. The site from which deletion would proceed was then established by *Sma*I restriction endonuclease digestion to completion, followed by phenol/chloroform extraction and ethanol precipitation. The sample was then re-suspended to a concentration of 1µg/µl before proceeding with the deletion.

Exonuclease III deletions were performed at 30°C (resulting in an approximate deletion of 230 bases per minute) using 6 time points (i.e. approximately 1.4kb was deleted). Essentially, per time point the reaction mixture contained 5µg DNA, 1 x ExoIII buffer and 10mM β-mercaptoethanol (in a 24µl volume). Thus since 6 time points were required, a reaction mixture containing 30µg DNA (144µl volume) was established and Exonuclease III added (100U/5µg DNA). Time point aliquots (25µl) were removed at 1 minute intervals into an new microcentrifuge tube containing diluted Mung Bean nuclease buffer (20µl of a 10 x buffer diluted into 155µl SDW per time point), and this mixture was then quick frozen in a dry ice/methanol bath until all time points had been collected. The Exonuclease III contained within these samples was then heat inactivated by incubating at 68°C (15 min), and the tubes placed on ice. Mung Bean nuclease (15U) was added to each time point reaction mixture, the samples incubated at 30°C (30 min), and 1 M Tris-HCl (pH 8.0; 10µl), 8M LiCl[8] (20µl) and 20% (w/v) SDS (4µl) added prior to Phenol/Chloroform extraction and Ethanol precipitation (by the addition of two volumes of 100% Ethanol). The pellet was re-suspended to an estimated concentration of 0.3µg/µl in 1 x TE, and a small aliquot ligated and transformed into XL1-Blue competent cells.

2.14 CONSTRUCTION OF GENOMIC LIBRARY

A *Calliphora erythrocephala* genomic library was constructed by *Sau* 3AI partial digestion of genomic DNA, followed by dephosphorylation and subsequent cloning into

Bam HI sticky ended (an isoschizomer of *Sau* 3AI) λgem-11 vector (note : protocols were derived from the Promega Protocols and Applications guide).

Initially, high molecular weight *C. erythrocephala* genomic DNA was isolated. A small scale titration of *Sau* 3AI restriction endonuclease was then tested to determine optimal enzyme concentrations for the digestion of a constant amount of genomic DNA (i.e. one yielding a high density of fragments in the 15 - 23kb size). To minimise the possibility of co-ligation events representing a significant proportion of the library, a large scale *Sau* 3AI partial digestion was then established (identical conditions to the small scale titration experiments were employed) using an enzyme concentration half that empirically determined as optimal in the small scale analysis (in this case, 0.00425U/μg genomic DNA; 30 minute digestion at 37°C).

This *Sau* 3AI digested genomic DNA was alkaline phosphatase digested (2hr incubation at 37°C; 0.1U alkaline phosphatase/μg genomic DNA) prior to ligation with *Bam*HI/ *Eco*RI double restriction enzyme digested λgem-11 vector DNA (note : the double digestion minimises the background of non-recombinant phage). Ligation was performed at 16°C overnight an approximate 2:1 ratio of insert DNA to vector DNA (i.e. approximately 2μg genomic DNA and 1μg λgem-11 vector DNA) before packaging the ligated molecules. Titrations of the packaged phage determined this primary library had a non-recombinant titre 6.3 x 10^5 pfu/ml, and a recombinant titre of 2.6 x 10^6 pfu/ml. The average insert size was determined (by random selection of 6 phage) as being approximately 15kb.

2.15 SEQUENCING OF DOUBLE STRANDED TEMPLATES

All sequencing reactions were performed using PRISM™ Ready Reaction DyeDeoxy™ terminator cycle sequencing kits (Perkin Elmer) and analysed using an ABI automated sequencer. Sequencing mixtures (terminator premix [9.5μl], dsDNA template [1.0μg], sequencing primer of choice [3.2 pmol], and SDW to adjust to a 20μl total volume; overlayed with approximately 20μl mineral oil) were placed in a pre-heated (96°C) Hybaid Omnigene PCR thermal cycler and cycled as per manufacturers instructions (96°C [30 sec], 50°C [15 sec], 60°C [4min];25 x cycle). Subsequent to cycling, unincorporated dNTP's were removed by the addition of 80μl SDW, Chloroform extracting (100μl), Phenol extracting (100μl of a 68:18:14 Phenol:Chloroform:SDW mixture), and ethanol precipitation.

2.16 TRANSFER OF NUCLEIC ACIDS TO HYBRIDISATION MEMBRANES

Table 2.13 details membranes used and denaturation, neutralisation and fixation protocols employed with these various membranes.

2.16.1 Southern Blotting (Southern, 1975)

Following electrophoresis gels were treated as described in Table 2.X13. and then washed for approximately 5 min in SDW. A reservoir of transfer buffer was established and both ends of two strips of Whatman filter paper (overlaid upon one another) suspended on a plastic tray allowed to rest in this buffer. The gel containing the DNA to be transferred was placed inverted onto the portion of Whatman filter paper resting on the plastic tray, and any bubbles trapped in the system removed by gently rolling a pipette over the gel. The selected membrane (cut to size) was then overlaid carefully onto the gel, two sheets of Whatman filter paper (cut to size) placed on top of the membrane, and approximately 3 inches of absorbent paper placed on top of the stack. To prevent short circuiting between the transfer buffer in the reservoir and the absorbent paper, saran wrap was placed over the exposed areas around the base of the gel. A small weight was then placed on top of the stack so as to ensure even contact and pressure throughout the system, and transfer allowed to proceed for a minimum of 4 hr.

After transfer the stack was disassembled (ensuring to mark the position of the gel wells on the membrane with a pencil) and the filter washed briefly in 5 x SSC before drying. If necessary (Table 2.13.) the DNA was fixed to the membrane by exposing to UV light (λ_{254}) for 3 min.

2.16.2 Northern blotting

Transfers involving RNA used the same protocol as described for DNA. Hybond-N membrane was preferentially used for the transfer of RNA since Hybond-N^+ has been shown to give higher background signals (manufacturers notes). Formaldehyde denaturing RNA gels require no denaturation or fixation, however the formaldehyde was extracted from the gel by washing gently in SDW for 30 min prior to blotting.

2.16.3 Colony/Plaque lifts

Plates were chilled at 4°C for a minimum of 1 hr before taking lifts. Duplicate lifts were often taken to confirm any signals seen in subsequent hybridisations. In such cases the first lift was left on the surface of the plate for 3 min and the second lift for 5 min. Denaturation, neutralisation and fixation of these lifts is described in Table 2.13.

Table 2.13. Nucleic acid blots and their associated membranes, transfer buffers, denaturation/ neutralisation and fixation steps

TYPE OF BLOT	MEMBRANE EMPLOYED	TRANSFER BUFFER	DENATURATION/ NEUTRALISATION	FIXATION
Southern	HYBOND-N$^+$ HYBOND-N	0· 4M Sodium Hydroxide 10 x SSC	None required 1 x 30 min in Denaturing solution 2 x 30 min in Neutralising solution	None required 3 min (λ_{254})
Colony/ plaque lifts	HYBOND-N	Not applicable	1 x 3 min in Denaturing solution 1 x 5 min in Neutralising solution	3 min (λ_{254})
Northern	HYBOND-N	20 x SSC	None required	3 min (λ_{254})

2.17 LABELLING NUCLEIC ACIDS

2.17.1 Generation of probes by random priming

The required amount of DNA to be labelled (hybridisations were usually performed with a probe concentration of 10ng per ml) was made up to a volume of 31µl with SDW and incubated at 95°C for 2 min before snap cooling on ice. Oligo labelling buffer (10µl), 20 mg/ml Bovine Serum Albumin (1µl), 10µCi/µl [^{32}P]dCTP (3µl) and 1U/µl Klenow enzyme (5µl) were then added before incubating at 37°C for 1 hr. Unincorporated nucleotides were removed by passing the labelled mixture through a Pharmacia Nick column. Probes were denatured (95°C, 2 min before placing on ice) prior to adding to hybridisation solutions.

2.17.2 Removal of unincorporated nucleotides using Pharmacia Nick columns

Columns were prepared by discarding the storage solution contained within the column, washing briefly with 1 x TE, and then allowing 1 x TE (3ml) to drain through the column. The probe was then applied (in a volume of 100µl or less) and allowed to

enter the resin before applying 1 x TE (400µl), the eluate from this being discarded. A second aliquot of 1 x TE (400µl) was applied to the column, and the collected eluate inferred to contain the purified probe (as per manufacturers instructions).

2.18 HYBRIDISATION AND AUTORADIOGRAPHY

All hybridisations and washes were carried out in Hybaid hybridisation ovens. Hybridisations were typically carried out at 42°C in 50% (v/v) Formamide hybridisation solution, whereas washes were often performed at 65°C. Table 2.14. details hybridisation solutions and wash solutions employed.

2.18.1 Pre-hybridisation and hybridisation of membranes

Membranes to be hybridised were rolled (separated by gauze sheets) and placed inside the hybridisation bottles. To ensure the membranes were rolled tightly against the walls of the bottles and to pre-wet the filters, the tube was half filled with 5 x SSC and the contents rolled in the solution until they ceased moving. The SSC was discarded and the tubes allowed to drain before pre-hybridising by adding hybridisation solution (typically 20ml) and incubating in the Hybaid oven (42°C) for a minimum of 1 hr. The denatured probe (2.14.1.) was either added directly to the pre-hybridisation solution (new probes) or the pre-hybridisation solution discarded and replaced with a previous hybridisation solution (denatured by placing at 65°C for 10 min) and the hybridisation allowed to proceed overnight (42°C).

2.18.2 Washing of hybridised membranes

The hybridisation solution was transferred to a small conical flask (screw cap) and retained for further hybridisations (stored at room temperature for a maximum of one week). Medium wash solution (pre-warmed to 65°C) was added and the solution rolled briefly in the tube before discarding and replacing with fresh medium wash solution. The tubes were then placed back into the Hybaid oven (65°C) and incubated for 30 min with slow rotation, the solution discarded and the wash repeated. If a high emission was still detected emanating from the top of the uncapped tube, then the filters were washed in High wash solution (65°C, 20 min). The tubes were drained of wash solution prior to removing the membranes for autoradiography. In all cases, tubes were approximately half filled with wash solution.

2.18.3 Autoradiography

The membranes were placed into heat sealed bags, fixed into autoradiography cassettes, and overlaid with X-ray film before placing at -70°C. The length of storage was dependant on the signal emanating from the membranes, and is detailed with the relevant

results. Cassettes were warmed to room temperature before developing the film in an X-OGRAPH compact X2 automated film processor.

Cassettes containing sequencing gels were stored at room temperature since [^{35}S] signals are not enhanced at -70°C.

Table 2.14. Hybridisation and wash solutions (Dyson, N.J. [1991] in Ess. Mol. Biol. vol. II)

SOLUTION	COMPONENTS
45% (v/v) Formamide hybridisation solution	5 x SSC, 5 x Denhardts, 45% (v/v) Formamide, 1% (w/v) SDS, 50mM Sodium Phosphate (pH 6·8), 10% (w/v) Dextran Sulphate (ml. wt. 500 000), 100 µg/ml salmon sperm DNA
50% (v/v) Formamide hybridisation solution	As above except 50% (v/v) Formamide
Medium Wash Solution	0·5 x SSC, 0·1% (w/v) SDS
High Wash Solution	0·1 x SSC, 0·1% (w/v) SDS

Note : 1M Sodium Phosphate (pH 6·8) is 1M Sodium Di-Hydrogen Orthophosphate (25.5 ml) and 1M Di-Sodium Hydrogen Orthophosphate (2·45 ml) adjusted to pH 6·8 and filter sterilised.

2.19 ANALYSIS OF PROTEIN EXTRACTS

2.19.1 SDS-Polyacrylamide gel electrophoresis of proteins

Carefully washed protein electrophoresis plates were assembled and the bottom edge of the plates placed in freshly prepared 'sealer solution' (3M Tris-HCl pH8.0 [1.26ml], 10%[w/v] SDS [100µl], 30:8 acrylamide:bis-acrylamide [6.1ml], 80% [w/v] Sucrose [1.8ml], SDW [0.7ml], 15% [w/v] AMPS [100µl] and TEMED [8µl]) until set. 'Separation mix' (3M Tris-HCl pH8.8 [5ml], 10% [w/v] SDS [0.4ml], 30:8 acrylamide:bis-acrylamide [13.8ml], SDW [20.8ml], 15%[w/v] AMPS [133µl] and TEMED [25µl]) was then poured between the plates, quickly overlayed with a small amount of overlay buffer (3M Tris-HCl pH 8.8 [25ml], 10% [w/v] Lauryl Sulphate [2ml], SDW [173ml]), and allowed to set for a minimum of 30min at room temperature. Once set the overlay buffer was poured off, the gel surface was rinsed with ddH_2O, and

fresh 'stacker solution' (0.5M Tris-HCl pH6.8 [1.25ml], 10%[w/v] SDS [100µl], 30:8 acrylamide:bis-acrylamide [1.25ml], SDW [7.3ml], 15%[w/v] AMPS [75µl] and TEMED [7µl]) poured between the plates (to a depth of approximately 1cm from the top edge of the plates) before inserting the comb and allowing to set. The gel and plates were placed in the electrophoresis apparatus, the sample wells rinsed in running buffer, and the gel pre-run for 5min before loading any samples (note : 1 x running buffer [Glycine (144g), Tris-base (30.2g), Lauryl Sulphate (10g); per Litre] was used throughout). Samples were then applied and electrophoresis allowed to proceed overnight at 60V.

2.19.2 Preparation of protein samples for SDS-PAGE

The relevant number of flies was homogenised extensively in Laemmle buffer (*Anopheles* and *Drosophila* samples 60µl, *Calliphora* samples 1000µl) and stored at -20°C until required. All protein samples were denatured (10 min, 95°C), centrifuged briefly, and applied quickly to the pre-run SDS-PAGE gel.

2.19.3 Western transfer of SDS-PAGE protein samples to Hybond-C supports

The Western transfer tank was filled with transfer buffer (stock solution : Glycine [57.6g], Tris-base [12.0g], SDS [4g], Methanol [800ml], and SDW [3200ml]), and a sandwich construct (containing the gel to be transferred and the nitro-cellulose Hybond-C membrane between blotting paper/brillo pads) was immersed in the buffer reservoir in the apparatus (the nitro-cellulose orientated so it was closest to the +ve electrode). Transfer was then allowed to proceed at 0.3 amps (40-50V) for 4 hr.

2.19.4 Hybridisation and detection of antibodies to Western immobilised proteins

The nitro-cellulose membrane was blocked for 1hr at room temperature (4% [w/v] BSA in TBST [10mM Tris-HCl pH8.0, 150mM NaCl, 0.05%[v/v] Triton X-100]; 25ml), and this solution then replaced with TBST solution containing the primary antibody (α-SXL, final concentration was 1:10) before incubating overnight at 4°C. The unbound primary antibody was then removed by washing the membrane (TBST [25ml/wash], 5min at room temperature, 5 replica washes), and secondary antibody (anti-mouse IgG alkaline phosphatase conjugated antibody [Promega]; 1:7500 final concentration in TBST) hybridised to the membrane for 30min at room temperature. Unbound antibody was then removed (as per removal of unbound primary antibody), and the Protoblot (Promega) detection system employed to detect phosphatase conjugated antibodies on the membrane. Essentially, the membrane was incubated in 1 x alkaline phosphatase buffer (100mM Tris-HCl pH 9.5, 100mM NaCl, 5mM $MgCl_2$; 15ml), Nitro-Blue Tetrazolueum (50mg/ml in 70% [v/v] dimethylformamide; 99µl) and BCIP (50mg/ml in dimethylformamide; 49.5µl) added, and the colour reaction allowed

to proceed to the desired intensity. The membrane was then washed in TBST to stop the reaction, and membranes were stored in the dark indefinitely.

2.20 WHOLE MOUNT *IN-SITU* HYBRIDISATION OF OVARIES

This protocol is based on the procedure described by Tautz and Pfeiffle (1989) with slight modification.

2.20.1 Synthesis of Digoxigenin (DIG) labelled DNA probes

The DNA template sample (0.3-3.0µg) was denatured (95°C, 10min, 15µl total volume; snap cooled on ice), and 2µl Hexanucleotide mixture (0.5M Tris-HCl, 0.1M $MgCl_2$, 1mM Dithioerythritol, 2 mg/ml BSA, 62.5 A_{260} U/ml hexanucleotides; pH 7.5; DIG DNA labelling kit [Boehringer Mannheim]), 2µl DIG DNA labelling mixture (1mM dATP, 1mM dCTP, 1mM dGTP, 0.65mM dTTP, 0.35mM DIG-dUTP; pH7.5), and 1µl Klenow polymerase (5U/µl) added prior to incubation at 37°C (60 min). The reaction was then stopped by the addition of 1µl EDTA (0.5M), and DNA precipitated by the addition of 1µl tRNA (10mg/ml), 2.5µl LiCl (4M), 75µl 100% (v/v) Ethanol and either incubation overnight (-20°C) or for 30 min (-70°C). The sample was centrifuged (13 000rpm, 10 min), the pellet washed briefly in 80% (v/v) Ethanol, vacuum desiccated, and finally resuspended in 500µl DNA-Hybrix solution (50% [v/v] Formamide, 5 x SSC, 100µg/ml sonicated/denatured salmon sperm DNA, 50µg/ml Heparin, 0.1% (v/v) Tween-20)

2.20.2 Fixation of ovaries

Ovaries were dissected in Ringer's solution, fixed in 4% (w/v) Paraformaldehyde (in 1 x PBS) for 20 min at room temperature, and washed in 1 x PBT (1 x PBS, 0.1% [v/v] Tween-20; 3 x 5 min) on a Catherine wheel rotor. The tissue was then rinsed in 9 : 1 Methanol/ EGTA (0.5M, pH8.0; 3 x 5 min; the tissue can be stored at this point at -20°C), and re-washed in 1 x PBT (3 x 5 min).

2.20.3 Preparation of tissue for *in-situ* hybridisation

The ovaries were incubated at room temperature (60 min) in 1 x PBT supplemented with Proteinase K (100µg/ml), washed in 1 x PBT supplemented to 2mg/ml Glycine (1 x 5min, °C), and subsequently rinsed in PBT (3 x 5 min, room temperature). The tissue was then re-fixed in 4% (w/v) Paraformaldehyde (in 1 x PBS; 20 min at room temperature), washed in 1 x PBT (3 x 20 min), and either used to pre-adsorb the anti-DIG sheep alkaline phosphatase conjugated secondary antibody, or used in the *in-situ* hybridisation.

2.20.4 *In-situ* hybridisation procedure

The fixed ovaries were equilibrated in 1:1 PBT/ DNA-Hybrix (10 min, room temperature), and subsequently pre-hybridised in DNA-Hybrix (45°C, 60 min). The prepared DIG labelled DNA probe (in DNA-Hybrix) was then denatured (95°C, 10 min, snap cooled on ice), and this probe solution used to replace the pre-hybridisation solution, before incubation overnight at 45°C. Unbound probe was then removed by incubating in fresh DNA-Hybrix (45°C, 20 min), 1:1 PBT/ DNA-Hybrix (45°C, 1hr), 1 x PBT (45°C, 20 min) and finally 1 x PBT (5 x 5 min, room temperature, on Catherine wheel).

2.20.5 Detection of DIG labelled molecules using an alkaline phosphatase conjugated secondary antibody

The PBT solution containing the hybridised ovaries was replaced with 1 x PBT supplemented with pre-adsorbed anti-DIG sheep alkaline phosphatase conjugated secondary antibody (1:1000 final dilution), and incubated at room temperature (60 min). Unbound antibody was removed by washing in 1 x PBT (3 x 20 min), and the tissue prepared for staining by incubating in TLMNT (100mM Tris-HCl pH9.5, 50mM $MgCl_2$, 100mM NaCl, 0.1% [v/v] Tween-20, 1mM Levamisole; 3 x 5 min). Staining was then achieved by incubation in TLMNT supplemented with 4.5µl Nitro-blue tetrazoleum (75mg/ml in 70%[v/v] dimethylformamide) and 3.5µl X-phosphate (50mg/ml in 100% [v/v] dimethylformamide) until the desired intensity of staining was obtained (typically 15 - 30min for *yp* samples). The staining solution was then removed by rinsing in 1 x PBT (2 x 5 min), 1 x PBS (3 x 5 min), and the tissue mounted on slides in aquamount.

3. RESULTS

3.1 SECTION I

The process of sex determination in *Drosophila melanogaster* has been extensively studied, and a great deal has been elucidated with respect to how various genes interact to bring about male or female differentiation. However, little is known about how sex is determined in other Dipteran insects, although there do seem to be a wide variety of sex-determining mechanisms in evidence (see section 1.4). An understanding of these processes could lead to wide ranging beneficial effects, not just with respect to the increased knowledge and understanding of a critical aspect of development, but in both economical and medical terms. For example, malaria (which is only transmitted by female mosquitoes, such as *Anopheles stephensi*, during a blood meal) is a cause of high mortality in many developing and third world countries. An understanding of Dipteran sex determination mechanisms could provide insight into how the relative sex ratio's of natural mosquito populations could be manipulated. In this case a relative decrease in the number of females in a population should be concurrent with a reduction in the incidence of malaria. Alternatively, population control strategies could help reduce the incidence of diseases transmitted by houseflies and other flesh flies.

Since at the onset of the work reported here, little genetical research had been implemented in a broad spectrum of Dipteran species, a molecular approach seemed the most appropriate way to characterise sex-determinants in some of these species. The mosquito *Anopheles stephensi* , the housefly *Musca domestica*, and the bluebottle *Calliphora erythrocephala* were.

3.1.1 Preliminary analysis using antibody probes

In order to first establish whether a molecular analysis was feasible, preliminary characterisation was carried out to determine if any homology to the protein encoded by the *Drosophila* sex determination gene *Sex-lethal* was evident in *C. erythrocephala* and *A. stephensi* whole adult protein extracts. This protein is detected at all stages of female development in *D. melanogaster* subsequent to the formation of the cellular blastoderm, and is the primary sex-determining gene in this species (see Introduction).

Whole body protein extracts were isolated from a variety of developmental stages, size fractionated by polyacrylamide gel electrophoresis, and transferred to nitro-cellulose membranes by Western blotting. These membranes were then probed for the presence of *Sex-lethal* protein homologues by hybridising with a mouse anti-SXL^F monoclonal antibody (kindly donated by Peter Lawrence, University of Cambridge). Antibody protein complexes were then detected by incubation with an anti-mouse IgG alkaline phosphatase conjugated secondary antibody followed by biochemical assays for alkaline

phosphatase activity. It should be noted however that the use of a monoclonal, as opposed to polyclonal antibody in this analysis would not be considered optimal, since it relies upon conservation of tertiary structure within the single epitope detected by the monoclonal antibody. In addition, this monoclonal exhibits cross-reactivity with faster migrating unidentified proteins (Peter Lawrence, pers. comm.), and is therefore not specific to SXL proteins.

As can be seen in figure 3.1.1, the anti-SXL^F monoclonal antibody readily detects sex-specific SXL isoforms in *D. melanogaster* samples, but cross-reacting protein of a similar size is only detected in *C. erythrocephala* unsexed pupae. The fact that no protein is apparent in adults of this species suggests this protein, even if it is expressed sex-specifically in pupae, is unlikely to represent a functional homologue of SXL since it does not appear to be active in adults. No cross-reacting proteins of a similar size are detected in any of the *A. stephensi* samples. Thus it was concluded that a homologue with a similar sex-determining function to *Sex-lethal* is unlikely to be detected in either of these species. This does not, however, exclude the possibility that as yet unidentified functions of SXL^F are conserved between *D. melanogaster* and *C. erythrocephala*. The presence of cross-reacting polypeptides present in *C. erythrocephala* pupae may therefore reflect a conserved role in an aspect of development unrelated to sex-determination.

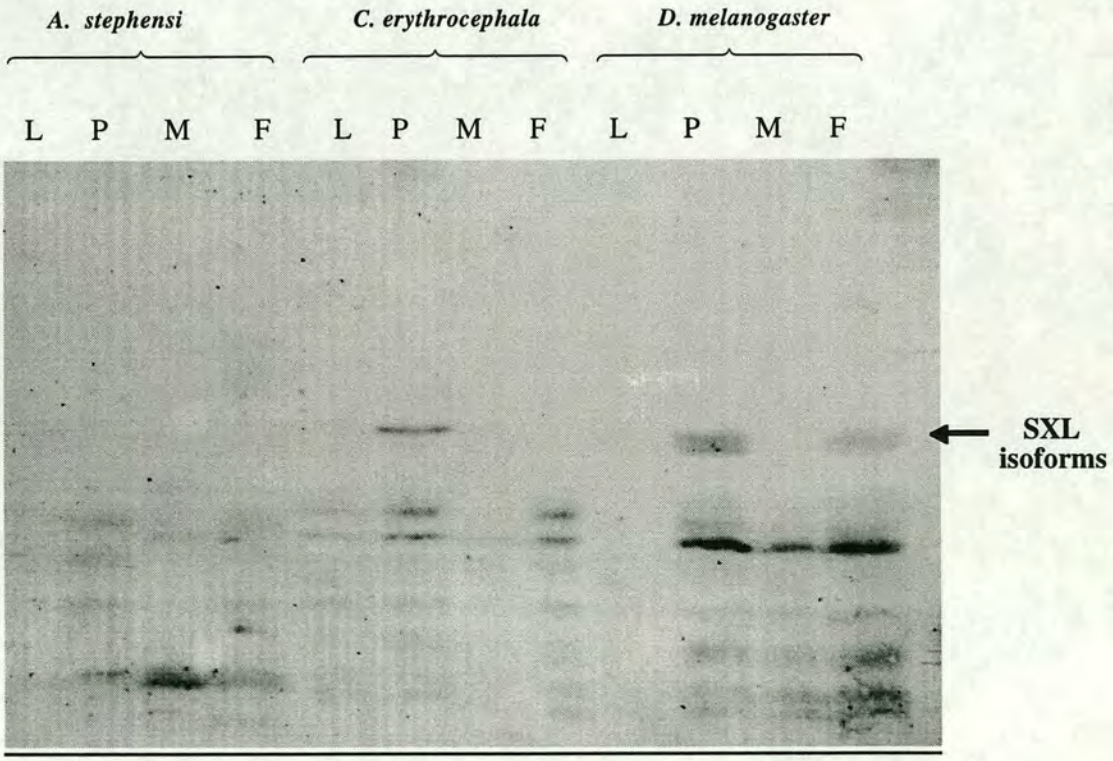


Figure 3.1.1. Western zooblot probed with anti-SXL^F monoclonal antibody. 'L', 'P', 'M' and 'F' refer to whole body protein isolates from unsexed larvae, unsexed pupae, adult males and adult females respectively. *Drosophila* and *Anopheles* samples contain 3 (larvae and pupae) and 5 (adult males and females respectively) samples respectively. The *Calliphora* samples contain protein extracted from the equivalent of 1/20th of each developmental stage.

3.1.2 Preliminary analysis using DNA probes

In parallel to the western analysis of protein extracts, cross-hybridisation analysis using cloned cDNA's representative of *Drosophila doublesex* transcripts was employed to detect any homology at the DNA level in *A.stephensi* and *C. erythrocephala* (the use of DNA rather than antibody probes was necessitated by the lack of an anti-DSX antibody). Genomic DNA (isolated from *A. stephensi* and *C. erythrocephala*) digested with a suitable restriction enzyme was fractionated on a 1% (w/v) agarose gel, Southern blotted, and hybridised to a high specific activity radioactively labelled female-specific *dsx* cDNA probe (termed DXP1; the entire linearised plasmid was used, please see figure 3.1.2). Since the correct stringency to use in these hybridisation's was unknown, a combination of both low stringency hybridisation and post-hybridisation washing was employed to minimise the chance that these conditions would be prohibitive to the analysis. To lower the stringency of the hybridisation (relative to the stringency which would normally be employed for the detection of homologous sequences), the formamide concentration was reduced (to 45% [v/v] instead of 50% [v/v], which has the effect of increasing the *Tm* of DNA duplexes) whilst the temperature of hybridisation was kept constant (42°C). Post-hybridisation washing was carried out using medium wash solution (1 x SSC; 0.1% [w/v] SDS) for 2 x 30 minutes at 42°C. As can be seen in figure 3.1.3, a single band of weak cross-reactivity (of approximately 8.0kb) is apparent in both *A. stephensi* and *C. erythrocephala* *Xho*I restriction enzyme digested genomic DNA samples. It should also be noted that the use of different restriction enzymes in this analysis does yield variation in the size of the cross-reacting fragments between the different species as expected (data not shown).

Collation of the Western and Southern hybridisation data suggested experimentation based upon DNA sequence similarity using *Drosophila dsx* cDNA probes would be more productive than an analysis based upon anti-SXL^F monoclonal antibody cross-reacting polypeptides. Since *C. erythrocephala* is more closely related to *D. melanogaster* than *A. stephensi,* any divergence between related sequences in *C. erythrocephala* and *D. melanogaster* are predicted to be less extensive than the equivalent sequence present in *A. stephensi*. Therefore a decision was made, on the basis of phylogeny, to concentrate on the analysis of *C. erythrocephala* cross-reactive sequences and return to those present in *A. stephensi* if time would allow.

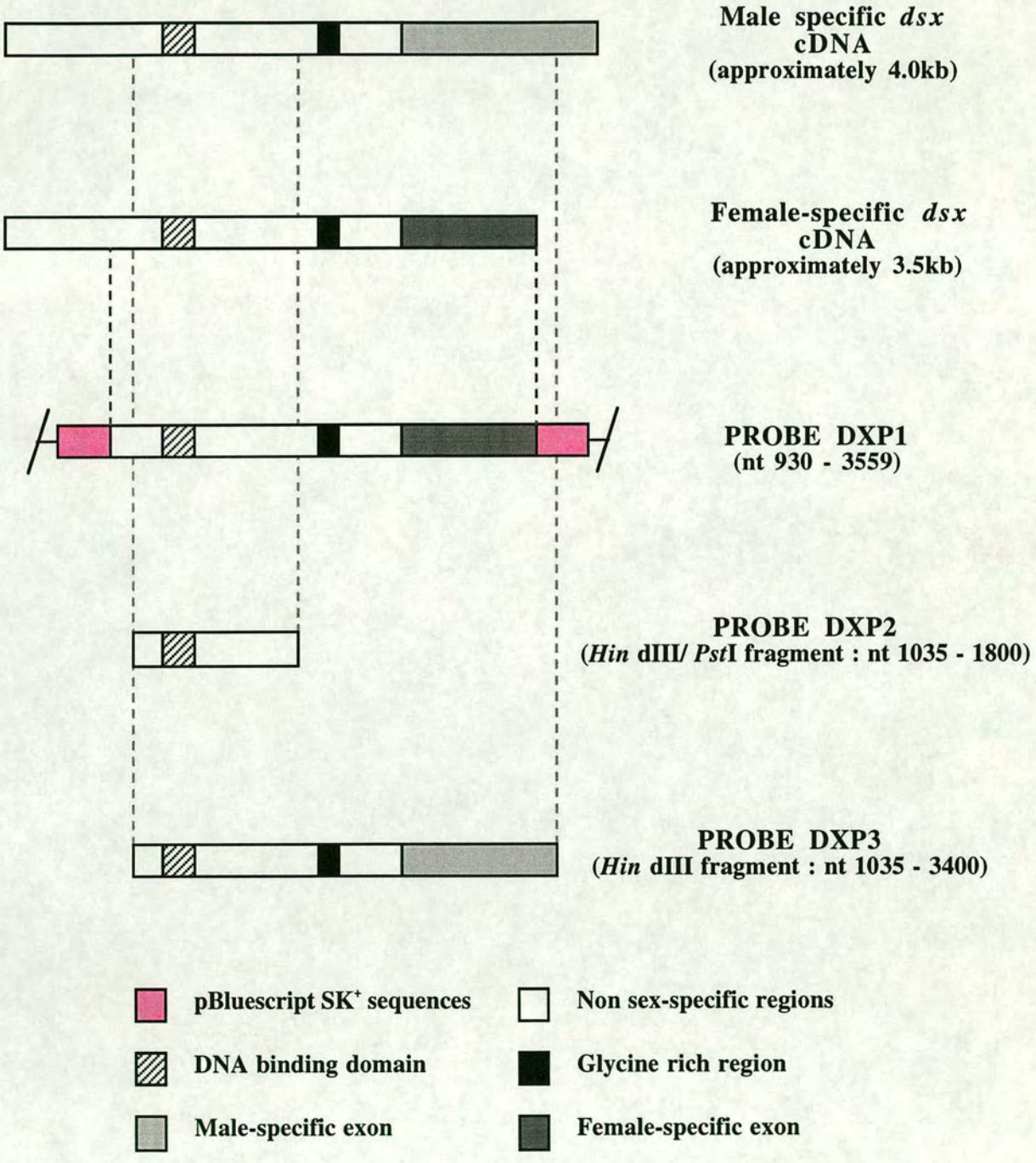


Figure 3.1.2. Schematic diagram showing the organisation of male- and female-specific *Drosophila dsx* cDNA's and the relevant probes used in cross-hybridisation studies.

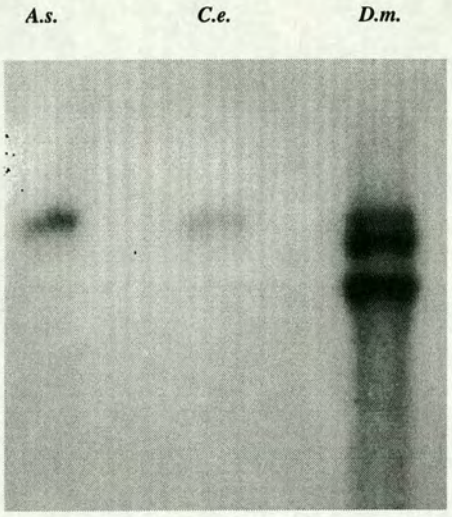


Figure 3.1.3. Southern zooblot hybridised to probe DXP1. '*A.s.*', '*C.e.*' and '*D.m.*' refer to *Anopheles stephensi, Calliphora erythrocephala* and *Drosophila melanogaster* (OrR) *Xho*I digested genomic DNA respectively. Each track contains approximately 5µg DNA . An overnight exposure is shown.

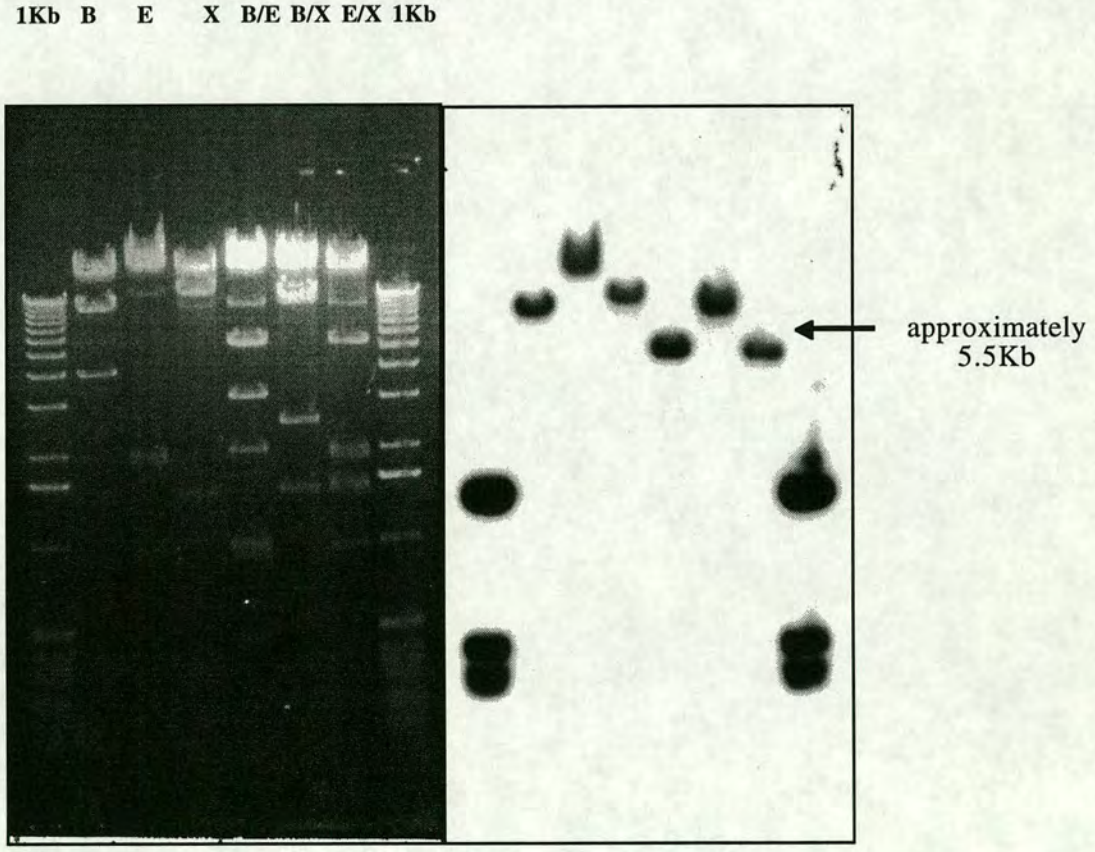


Figure 3.1.4. Cross hybridisation analysis of recombinant phage CedsxA with probe DXP1. 'B', 'E', and 'X' refer to restriction enzyme digestions using *Bam* HI, *Eco* RI, and *Xho* I respectively (note each track contains approximately 5µg phage DNA). 1Kb refers to Gibco BRL 1Kb ladder marker. Clone CedsxA.55 (referred to in main text) derives from subcloning of the 5.5Kb *Bam* HI/ *Eco* RI cross-hybridising fragment.

3.1.3 Isolation and analysis of *Calliphora* genomic sequences with apparent homology to *Drosophila doublesex* cDNAs

Although a *Calliphora erythrocephala* genomic library was available in our laboratory, this library contained small genomic inserts, and was known to be contaminated with plasmid derived sequences (Mary Bownes, pers. comm.). Thus, since no other library was available to our knowledge, a *C. erythrocephala lamda*-gem11 genomic library was constructed. A screen of this *C. erythrocephala* genomic library (approximately 1.5 x 10^5 pfu) using identical conditions to those used in the original Southern hybridisation analysis resulted in the isolation of two putative positive recombinant phage. Lambda DNA was isolated from these two phage, and both were subsequently shown to contain identical inserts (hereafter referred to as phage CedsxA) of approximately 14kb by restriction enzyme digestion followed by agarose gel electrophoresis (data not shown). To determine the location of those sequences with apparent homology to probe DXP1, CedsxA phage DNA was restriction enzyme digested, Southern blotted, and finally hybridised to probe DXP1 as per the original screens (see figure 3.1.4). This localised the putative homologous sequences to a 5.5kb *Bam*HI/ *Eco*RI fragment adjacent to the right arm of the lambda phage, which was subsequently isolated by gel purification and subcloned into the pBluescript SK^+ vector (hereafter referred to as clone CedsxA.55).

Sequence analysis (data not shown) of the insert contained within clone CedsxA.55 however revealed an extensive bias towards GGN nucleotide triplet repeats (the Glycine codon), and showed no significant homology to *dsx* cDNA sequences other than in the small Glycine rich domain located in the C-terminal portion of the non sex-specific region of the *dsx* encoded proteins (15 residues out of 18 are glycine, see figure 3.1.8). The fact that genomic library screens using probe DXP1 resulted in the isolation of sequences containing GGN nucleotide triplet repeats suggests this region of the probe is interfering with the cross-species hybridisation analysis. In order to determine if this hypothesis was correct, Southern blots containing genomic DNA isolated from *Calliphora erythrocephala* and *Musca domestica* were hybridised (using the same conditions as previously described for probe DXP1) to an 0.8kb *Hind* III/ *Pst*I *dsx* cDNA fragment (termed DXP2). This probe, which derives from the non sex-specific region of both male- and female-specific *dsx* cDNA's, contains the atypical DNA binding domain present in the encoded proteins (which is likely to be conserved), and excludes those codons encoding the glycine rich repeat element (see figure 3.1.2). As can be seen in figure 3.1.5, no cross hybridisation is observed in any sample of *C. erythrocephala* genomic DNA. Surprisingly however, strong hybridisation is detected in all samples of *M. domestica* DNA.

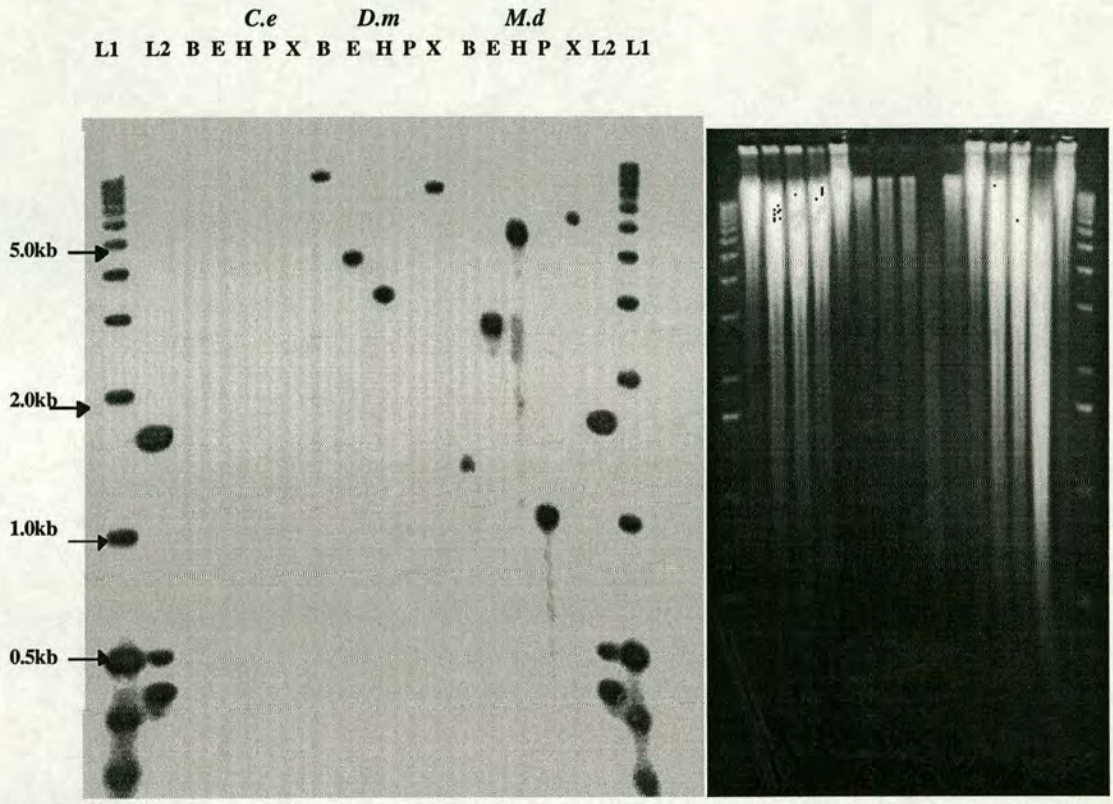


Figure 3.1.5. Southern zooblot hybridised to probe DXP2. 'B', 'E', 'H', 'P' and 'X' refer to *Bam*HI, *Eco*RI, *Hin* dIII, *Pst*I and *Xho*I restriction enzyme digested DNA samples respectively. The original gel is included to show that DNA is present in all tracks. Both *C. erythrocephala* (*C.e*) and *M. domestica* (*M.d*) samples contain approximately 10μg per lane, whereas the *D. melanogaster* (*D.m*) samples contain 3μg (to approximately account for differences in genome size). Note that the *D. melanogaster* track marked 'P' contained degraded DNA. An overnight exposure is shown. L1 refers to end labelled 1kb ladder marker (Gibco BRL), and L2 refers to the same marker but unlabelled.

Since no cross-hybridising sequences are evident in *C. erythrocephala* DNA using probe DXP2, and library screens using probe DXP1 yield sequences containing GGN triplets, it is likely that the cross-reactivity observed in hybridisations using probe DXP1 is dependent on the presence of nucleotides encoding the glycine rich repeat element. Taking this into consideration, it also seems likely that the cross reacting sequences present in *Anopheles stephensi* genomic DNA (see figure 3.1.3) will represent GGN triplet repeat elements and not *dsx* homologues, although this possibility can not be excluded. To determine this conclusively *Anopheles stephensi* genomic DNA should be tested for cross-reactivity with probe DXP2.

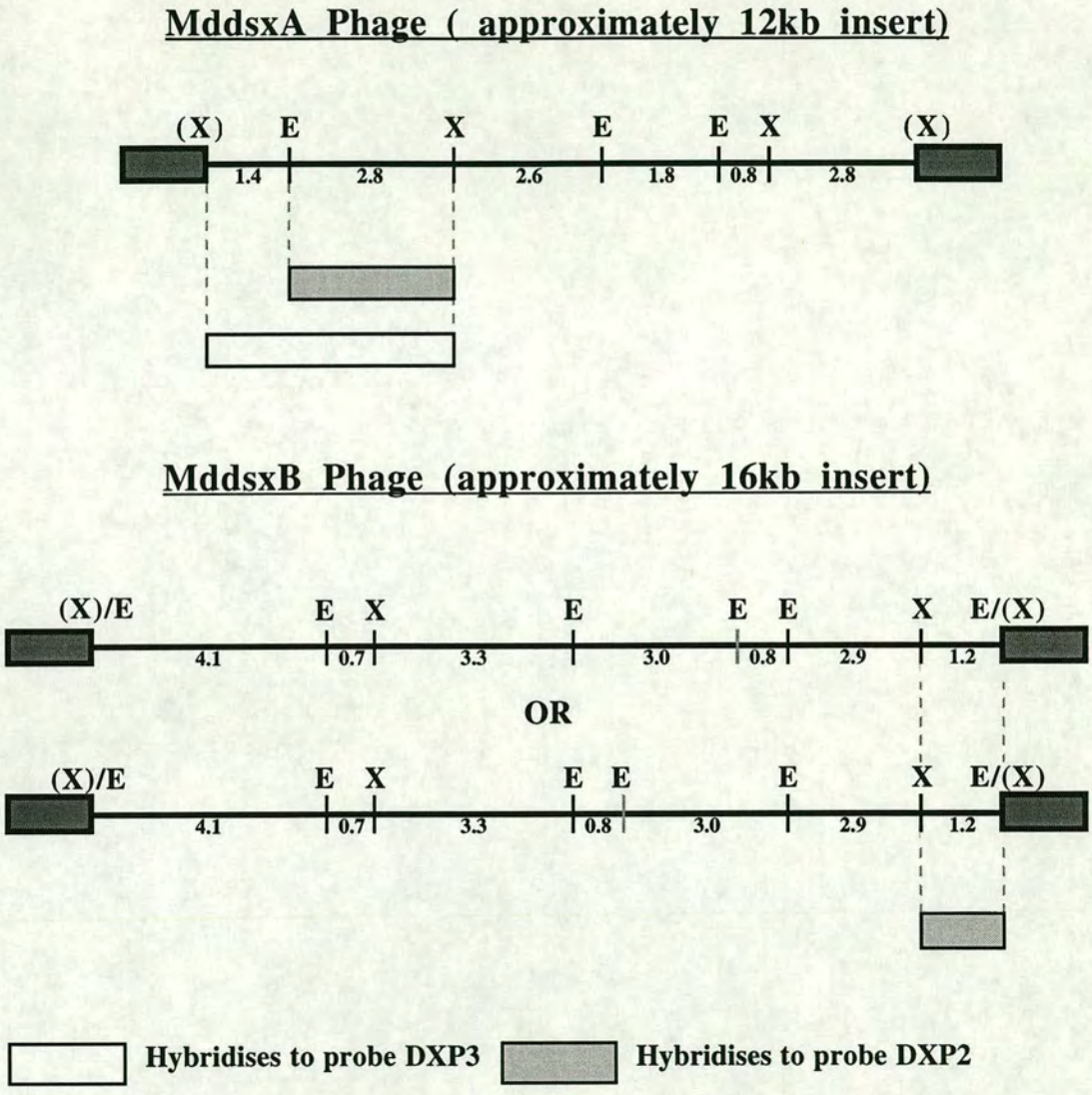


Figure 3.1.6. Organisation of phage MddsxA and MddsxB and the relevant regions hybridising to probes DXP2 and DXP3. Diagrams are not to scale. 'E' and 'X' refer to *Eco*RI and *Xho*I restriction enzyme sites respectively. An *Eco*RI site ,which could not be placed conclusively, is marked lightly. All sizes refer to kilobases.

3.1.4 Isolation and analysis of *Musca* genomic sequences showing homology to *Drosophila dsx* sequences

In order to characterise those *M. domestica* genomic DNA fragments cross-hybridising to probe DXP2 (see figure 3.1.5), a *M. domestica lambda*-gem11 genomic library (kindly donated by Claudia Tortiglione, our laboratory) was screened (approximately 1.5 x 10^5 pfu) using this probe. Two independent positive recombinant phage were isolated containing inserts of approximately 12kb and 16kb respectively, termed MddsxA and MddsxB, and are depicted in figure 3.1.6. Cross-hybridisation analysis of restriction enzyme digested *lambda* DNA isolated from phage MddsxA and MddsxB was used to delimit the regions showing homology to the DXP2 probe. Since probe DXP2 does not contain either the entire male or female *dsx* cDNA sequence, an additional hybridisation was carried out using a 2.4kb *Hind* III male-specific *dsx* cDNA probe (termed DXP3; see figure 3.1.2) which contains both the glycine rich repeat element and a large proportion of the male-specific exon. Figure 3.1.7 shows the relevant cross-hybridisation analysis of phage MddsxA and MddsxB with either probe DXP2 or probe DXP3. It is apparent that homologous sequences are more extensive when using probe DXP3 compared to probe DXP2, and supports the hypothesis that these phage contain *dsx* homologues, since in this case homology is evident in regions outwith the limitations of the original probe (see figure 3.1.2).

To characterise further those fragments derived from phage MddsxA and MddsxB which cross-hybridise to probe DXP2 (i.e. the 2.8kb and 1.2kb *Eco*RI/ *Xho*I fragments present within phage MddsxA and MddsxB respectively), they were gel purified and subcloned into the pBluescript SK^+ vector. These subclones are referred to as MddsxA.28 and MddsxB.12 respectively. Surprisingly, partial sequence analysis of these subclones revealed that the DNA present within these recombinants was essentially identical (> 99.8% identical) to the original *D. melanogaster dsx* cDNA sequence reported (Burtis and Baker, 1989b). When compared to the original *dsx* sequences, MddsxA.28 contains terminal sequences which map to the end of exon II and the start of the female-specific exon of the *Drosophila dsx* gene respectively, whilst MddsxB.12 in contrast maps entirely to exon II. The organisation of these fragments, and the apparent lack of introns, suggests those sequences contained within the MddsxA and MddsxB phage which show homology to probes DXP2 and DXP3 are derived from cDNA contamination of the library introduced prior to library construction. If this is the case, those *M. domestica* genomic fragments cross-hybridising with probe DXP2 in the original Southern analysis (see figure 3.1.5) would also constitute contaminating sequences, since the *M. domestica* genomic DNA used to construct the library was identical to that used in this Southern analysis.

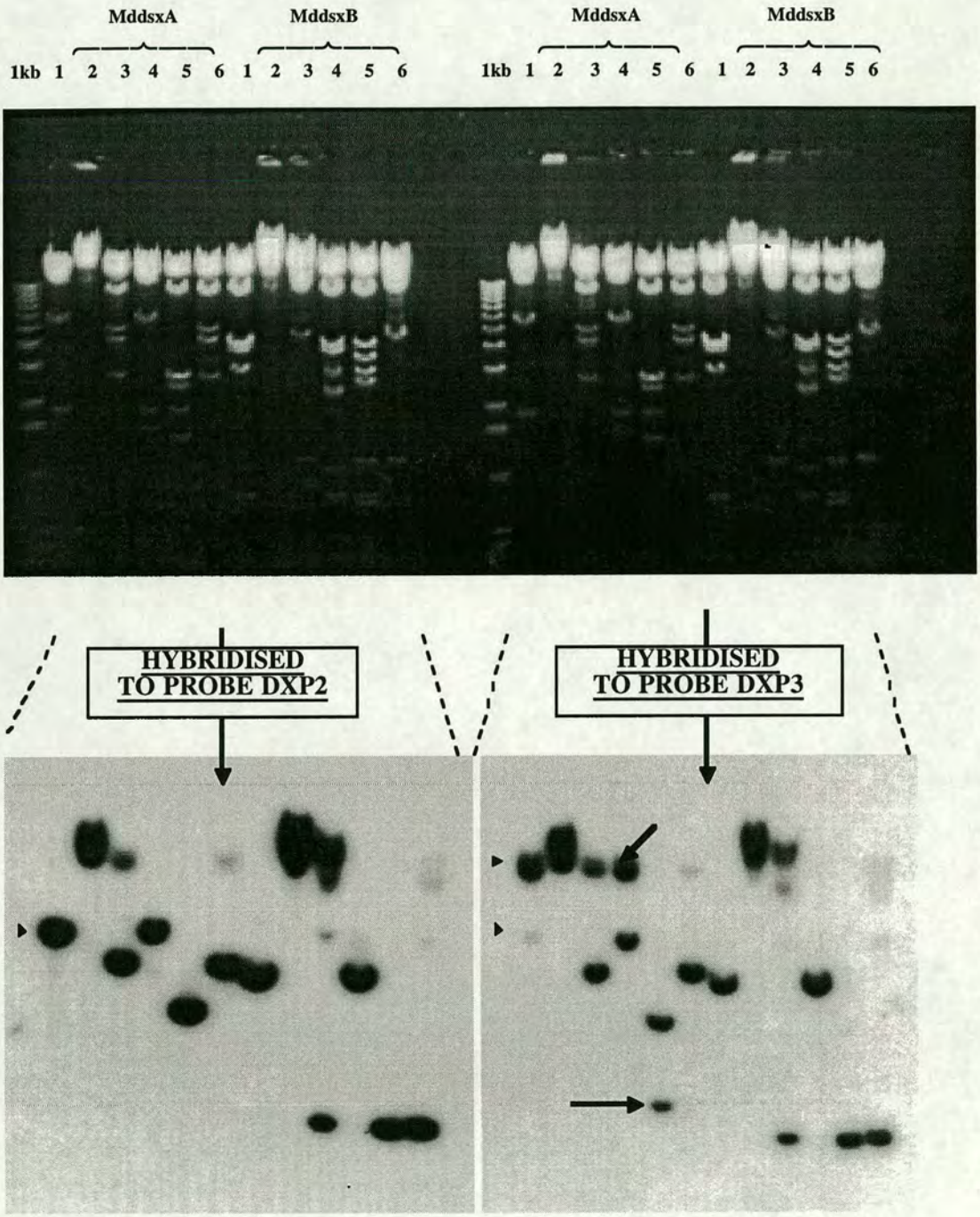


Figure 3.1.7. Cross-hybridisation analysis of recombinant phage MddsxA and MddsxB. Replica filters of restriction enzyme digested phage DNA were generated and probed with either probe DXP2 or probe DXP3 respectively. Each track contains approximately 5µg phage DNA. Major differences apparent in cross-hybridisation with the two probes is indicated (arrows). Arrowheads (MddsxA ;track 1) indicate differences which in our opinion are artefacts of the hybridisation, since in MddsxA track 4 (which contains identical fragments) cross-hybridisation is observed in the manner anticipated (i.e. the probe hybridises to the short 1.4kb *Eco* RI / *Xho* I fragment apparent in track 5, which in track 1 would be attached to the left arm of the lambda phage).'1', '2', '3', '4', '5', and '6' refer to restriction enzyme digestions using *Eco*RI, *Xba*I, *Xho*I, *Eco*RI / *Xba*I, *Eco*RI / *Xho*I, and *Xba*I / *Xho*I respectively. 1kb refers to Gibco BRL 1kb ladder marker.

3.1.5 RT-PCR amplification of *dsx* sequences

To test the hypothesis that the *M. domestica* genomic library is contaminated with *Drosophila dsx* cDNA sequences, reverse transcriptase PCR (RT-PCR) using redundant oligonucleotides was used to determine if transcripts containing sequences homologous to *dsx* were present in *Musca domestica* poly(A)$^+$ RNA.

Since only one *dsx* sequence has been reported so far (Burtis and Baker, 1989b), it is impossible to design oligonucleotides to functionally conserved domains of the encoded proteins. However, the DNA binding domain is crucial for *dsx* function in *Drosophila*, and would presumably be present and conserved in any functional *dsx* homologue in *Musca domestica*. Also, the presence of sex-specific exons in the C-terminal regions of the male- and female-specific *dsx* cDNA's suggests these exons may be intrinsic to *dsx* function. Additionally, comparison of *D. melanogaster* and *D. virilis* (recently isolated) *dsx* sequence reveals extensive homology in the female-specific exon (K. Burtis, pers. comm.). It was therefore reasoned that the best sites for oligonucleotide design would be in those regions defining the DNA binding domain (particularly those codons encoding the Cystine residues implicated in metal ion binding; see Erdman and Burtis, 1993) and the female-specific exon. Since it was predicted that the sequences evident in clones MddsxA and MddsxB would be absent from *M. domestica* total RNA, redundant oligonucleotides based on the codon bias of the *M. domestica* and *C. erythrocephala* genomes were designed (see figure 3.1.8). However, the redundancy of these oligonucleotides would not prohibit amplification of such sequences if they were present, as revealed by amplification of positive control template DNA.

Figure 3.1.9 shows the amplification products from a variety of templates using oligonucleotides A_2 and B (please refer to the materials and methods for a description of the reaction conditions and the optimisation procedure). Control template DNA (i.e. the *Drosophila* female-specific *dsx* cDNA) shows extensive amplification, whereas no product is detected when either *C. erythrocephala* or *M. domestica* first strand cDNA is used as a template (amplification was only ever observed on one occasion and was not reproducible). However, *Musca domestica* ovarian cDNA library (kindly donated by D. Bopp, University of Zurich) template DNA does generate a product of the anticipated size. Since it could be envisaged that the different template DNA samples have variable EDTA concentrations (which chelate Magnesium ions and reduce the effective Magnesium concentration in the reaction), a range of Magnesium Chloride concentrations (1-5mM) was tested. No product of the expected size was detected in any

reaction using either *C. erythrocephala* or *M. domestica* first strand cDNA as a template (note this cDNA derived from reverse transcription of adult RNA). Similar results were obtained when these analyses were repeated using the combination of oligonucleotides A_1 and B.

A. PEPTIDE SEQUENCE OF THE FEMALE-SPECIFIC *DSX* PROTEIN

1 MVSEENWNSDTMSDSDMIDSKNDVCGGASSSSGSSISPRTPP**NCARC**RNHGLKITLKGHK 60

MINIMAL DNA BINDING DOMAIN

61 RYCKFR**YCTCEK**CRLTADRQRVMALQTALRRAQAQDEQRALHMHEVPPANPAATTLLSHH 120

121 HHVAAPAHVHAHHVHAHHAHGGHHSHHGHVLHHQQAAAAAAAAPSAPASHLGGSSTAASS 180

181 IHGHAHAHHVHMAAAAAASVAQHQHQSHPHSHHHHHQNHHQHPHQQPATQTALRSPPHSD 240

GLYCINE RICH REGION

241 HGGSVGPATSSSGGGAPSSSNAAAATSSNGSSGGGGGGGGGSSGGGAGGGRSSGTSVITS 300

301 ADHHMTTVPTPAQSLEGSCDSSSPSPSSTSGAAILPISVSVNRKNGANVPLGQDVFLDYC 360

FEMALE-SPECIFIC

361 QKLLEKFRYPWELMPLMYVILKDADANIEEASRRIEE**GQYVVN**EYSRQHNLNIYDGGELR 420

EXON

421 NTTRQCG 427

B. OLIGONUCLEOTIDES USED FOR PCR AMPLIFICATION

FORWARD PRIMERS

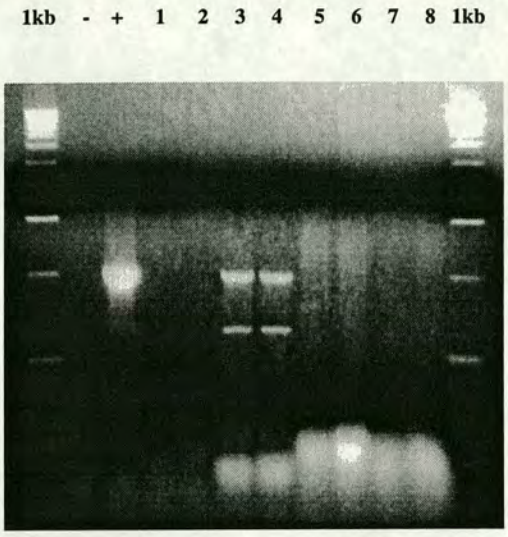
PRIMER A_1 : $^{5'}$ GGAATTC AAY TGY GCY CGY TGY CG $^{3'}$
*Eco*RI N C A R C
(minimum Tm = 50^{O}C, maximum Tm = 60^{O}C, redundancy = 36 fold)

PRIMER A_2 : $^{5'}$ GGAATTC TAY TGY ACN TGY GAR AAR $^{3'}$
*Eco*RI Y C T C E K
(minimum Tm = 44^{O}C, maximum Tm = 56^{O}C, redundancy = 128 fold)

REVERSE PRIMER

PRIMER B : $^{5'}$ GGAATTC RTT NAC NAC RTA YTG NCC $^{3'}$
*Eco*RI N V V Y Q G
(minimum Tm = 46^{O}C, maximum Tm = 58^{O}C, redundancy = 512 fold)

Figure 3.1.8. A. Peptide sequence of the female-specific *dsx* protein. The DNA binding domain, glycine rich domain, and female-specific exons are all highlighted. Bold type in the peptide sequence indicates sites used for PCR primer design. **B. Redundant oligonucleotides designed for cross species *dsx* PCR amplification.** Sequences under the DNA strand refer to the peptide sequence. With respect to the DNA strand only, Y = T/C, R = A/G and N = A/C/G/T.



Key to figure 3.1.9

- = negative control (i.e. no template)

+ = positive control (10ng dsx^F cDNA)

1, 2 = *C. erythrocephala* 1st strand cDNA (sample one, approx. 20ng)

3, 4 = *M. domestica* cDNA library (2µl ; approx. 1 x 10^{13} pfu/ml)

5, 6 = *C. erythrocephala* 1st strand cDNA (sample two, approx. 20ng)

7, 8 = *M. domestica* 1st strand cDNA (sample two, approx. 20ng)

Figure 3.1.9. PCR amplification from a variety of templates using redundant *dsx* oligonucleotides A_2 and B. Templates are indicated in the figure key, where first strand cDNA refers to the products derived from reverse transcription of polyA$^+$ RNA from the species indicated. 1kb refers to Gibco BRL 1kb ladder marker.

The apparently contradictory positive (library) and negative (first strand cDNA) results obtained using the *M. domestica* cDNA templates was initially intriguing, since identical amplification products would be anticipated. This result could be most simply explained by either mispriming of the oligonucleotides on template DNA only present in one of the reactions (i.e. vector DNA), or a difference in template quality [for example due to degradation of the poly(A)$^+$ RNA prior to first strand cDNA synthesis]. Two lines of evidence strongly support the former hypothesis (i.e. mispriming). Firstly, the product derived from amplification of the *M. domestica* ovarian cDNA library, when subcloned and sequenced, was shown to contain sequences identical to lambda bacteriophage (data not shown). Secondly, first strand cDNA generated using similar cDNA synthesis procedures was successfully used as a template in other RT-PCR applications (see results section II).

Mispriming of the oligonucleotides, in many respects, is not unexpected since the redundancy introduced into the oligonucleotides is inherently detrimental, as it allows binding at a large variety of target sequences. It is impossible to characterise, or predict, all possible target binding sites for oligonucleotide combinations present within such a

mixture (especially when designing high redundancy primers). However, it is anticipated that such redundant oligonucleotides would preferentially associate with target sequences closely resembling the original conserved sequence, and would rapidly become the major product of the reaction. The fact that this has not occurred in any of the reactions described here strongly supports the hypothesis that target sequences encoding *dsx* function are either absent or extensively diverged in the species analysed. However, we can not exclude the possibility that these oligonucleotides are not located in functionally conserved domains of the protein, although it is likely that sequences around the DNA binding domain would be conserved. Thus we favour the view that the oligonucleotide designed to sequences present within the female-specific exon would be most open to suspicion.

3.1.6 SECTION I : DISCUSSION

Despite using a wide variety of techniques, our attempts to isolate homologues of *Drosophila* sex determination genes in other Dipteran species using molecular approaches has proved unsuccessful. However, the results presented here could be viewed as being dependent on the limitations of the techniques. Cross-hybridisation analysis using heterologous probes is problematic, since the determination of the correct hybridisation conditions is intrinsically difficult. Additionally, as observed with those hybridisations using probe DXP1, apparently innocuous repetitive elements located within probes when hybridised under low stringency conditions can generate misleading results. The observation that these hybridisations yielded defined bands with low background suggested the conditions of hybridisation were suitable. Only subsequent characterisation of the homologous sequences revealed the limitations of the original probe. This analysis, therefore, serves as a cautionary note when considering the isolation of genes based purely on the use of heterologous probes under low stringency conditions.

Despite this, the Southern hybridisations using probe DXP2 demonstrate that, under the low stringency conditions employed, no homologous sequences can be detected in *Calliphora erythrocephala* genomic DNA. Additionally, homologous sequences evident in the initial cross hybridisation analysis of *M. domestica* genomic DNA are most likely derived from contaminating *Drosophila dsx* cDNA sequences. However, cross-hybridising fragments differing in size to those apparent in the initial *M. domestica* analysis are detected in newly isolated *M. domestica* genomic DNA (see figure 3.1.10). The fact that no product could be generated using *dsx* gene-specific redundant oligonucleotides in RT-PCR analyses however suggests that transcripts homologous to *dsx* are either not present in *M. domestica* or *C. erythrocephala,* or are highly diverged. It is of course conceivable that these oligonucleotides (particularly the primer designed to the female-specific exon) do not reside in functionally conserved

domains of the encoded proteins, and would therefore generate misleading results. Alternatively in contrast to *dsx* function in *Drosophila*, if a homologue does indeed exist it may not be required in adult stages at all, but rather acts at an early stage of development to direct male or female development irreversibly (such as in the larval or pupal stages). In this case, the RT-PCR based analysis reported here should be repeated using RNA derived from a variety of developmental stages.

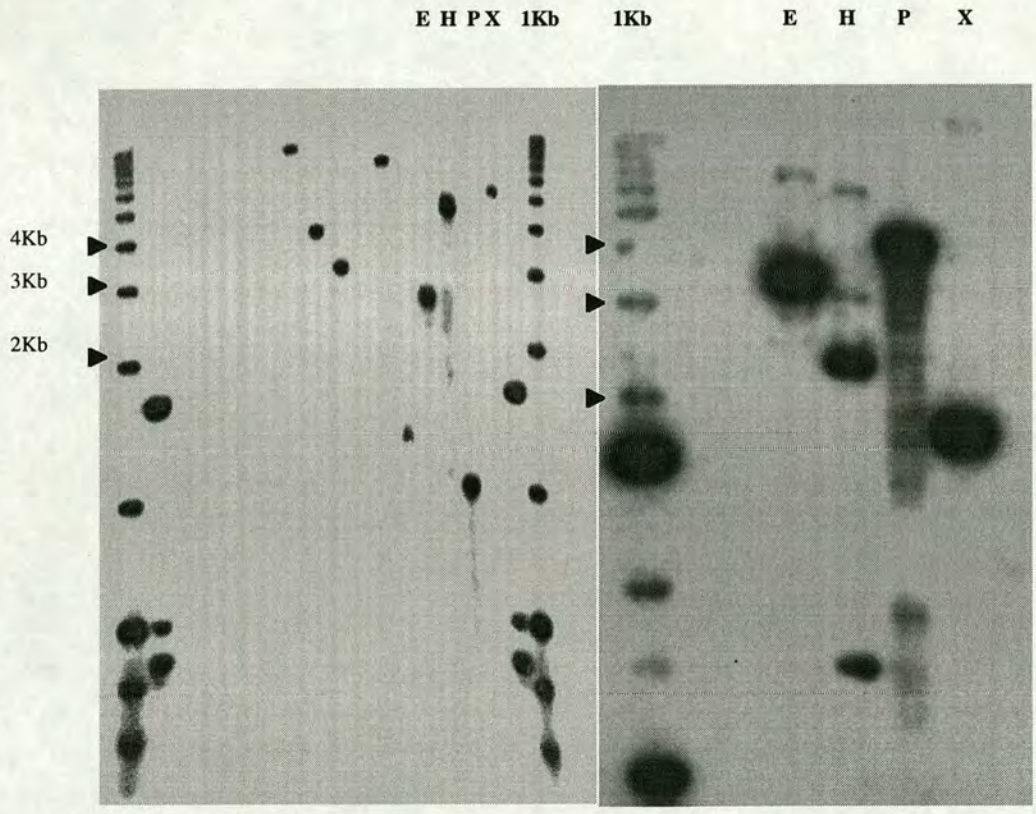


Figure 3.1.10. Comparison of Southern Hybridisations to different *Musca domestica* genomic DNA isolates using probe DXP2. The panel on the left derives from genomic DNA used in library construction procedures (refer to main text and figure 3.1.5). The panel on the right derives from newly isolated *Musca domestica* genomic DNA. 'E', 'H', 'P', and 'X' refer to *Eco* RI, *Hin* dIII, *Pst* I and *Xho* I restriction enzyme digestions respectively, and only those tracks containing *Musca domestica* genomic DNA are indicated. Arrowheads mark the positions of the 2Kb, 3Kb, and 4Kb fragments present within the marker in each instance.

Thus, either of two interpretations can be envisaged. Either the cross-hybridising sequences present in the *M. domestica* genome do in fact represent a *dsx* homologue (and the RT-PCR analysis is unreliable), or the reduced stringency of the hybridisation is facilitating hybridisation of the probe to sequences of only limited homology (such as unrelated DNA binding proteins). Conclusive determination of the correct interpretation would require the isolation of the newly identified cross-reacting sequences present in the

M. domestica genome. However, the fact that independent research (K. Burtis, University of Davis, pers. comm.) has not resulted in the isolation of an *M. domestica dsx* homologue despite using a similarly wide variety of techniques supports the hypothesis that either a gene functionally equivalent to *dsx* is absent in *M. domestica*, or that selective pressure is minimal resulting in rapid divergence of the gene sequence.

In summary, although none of these results are conclusive in themselves, taken together they strongly suggest that genes functionally equivalent to *dsx* are unlikely to be identified in *Calliphora* or *Musca* using the presently limited techniques. The cloning and characterisation of *dsx* homologues from Dipteran species more closely related to *Drosophila melanogaster* should clarify these results, and make the design of oligonucleotides for PCR based evolutionary comparisons more reliable.

3.2 SECTION II

The results presented in the previous section were inconclusive as to whether functional equivalents of *dsx* are present in either *C. erythrocephala* or *M. domestica*. In order to further investigate this, a strategy was devised to detect polypeptides functionally equivalent to DSX^M or DSX^F in either of these species by inference. As discussed in the introduction (see section 1.5.3.1.1.2), expression of the *yolk protein-1* and *yolk protein-2* genes in the female fat body of *D. melanogaster* is directly regulated by the sex-specific proteins encoded by *doublesex*. Male-specific *dsx* protein (DSX^M) represses transcription of both *yp1* and *yp2* by stearic hindrance of *trans*-acting factors, whereas DSX^F helps to stabilise the association of these factors, thus enhancing transcription.

Martinez and Bownes (1994) have demonstrated that two *yolk protein* genes in *C. erythrocephala* (*CeypA* and *CeypB*) are expressed in a manner analogous to that observed in *D. melanogaster* (i.e. they are expressed only in the female fat body and the ovarian follicle cells during the correct stages of oogenesis). The observation that the expression profiles of *yolk protein (yp)* genes is conserved in *D. melanogaster* and *C. erythrocephala* suggests that functionally equivalent *trans*-acting factors regulating their sex-, stage-, and tissue-specific expression will also be conserved. Indeed, in support of this hypothesis, factors involved in the selective uptake of vitellogenin into developing oocytes are conserved. This is apparent since yolk proteins present in radioactively labelled *C. erythrocephala* donor haemolymph, when injected into host *D. funebris* females, accumulate in the host ovaries (Martinez, 1991; Martinez and Bownes, 1992).

Since *C. erythrocephala yp* expression is clearly regulated in a sex-specific manner in the fat body, it is not unreasonable to suggest that this sex-specific expression is dependent on the presence of a functionally equivalent *dsx* gene homologue. In this case, it would be expected that the sex-specific expression of *D. melanogaster* and *C. erythrocephala yp*'s would be effected by similar regulatory elements. To test this hypothesis, we intend to generate reporter gene constructs under the control of species-specific *yolk protein* gene promoter elements. Germ line transformation would then be used to introduce these constructs into the *D. melanogaster* genome. Sex-specific expression of the reporter gene in the transgenic *D. melanogaster* fat body would infer that a gene functionally equivalent to *dsx* exists in the species from which the promoter element was derived (since the host *dsx* proteins are able to regulate expression of the reporter gene).

Although *C. erythrocephala* genomic *yolk protein* sequences have been isolated (Martinez and Bownes, 1994), only one of the phage characterised contains sequences upstream of the transcription start site. In order to make this type of analysis reliable, reporter gene constructs should be generated using a minimum of two promoter elements derived from independent endogenous *yolk protein* genes. Since only one

C. erythrocephala yp promoter was available, more sequences were required. In addition, the experiments reported in the previous section focused primarily on the analysis of both *C. erythrocephala* and *M. domestica*, and therefore an analysis of *M. domestica yp* promoter elements was of interest. However, the genes encoding yolk proteins have not yet been identified in *M. domestica*, necessitating their isolation and a characterisation of their expression profiles.

3.2.1 Isolation of *M. domestica* and *C. erythrocephala yolk protein* sequences

The availability of *yolk protein* gene sequences from several Dipteran species facilitated the design of redundant oligonucleotides to evolutionary conserved domains of the encoded proteins (Note : These oligonucleotides were designed by Dr. C. Simpson, our laboratory). Alignments of yolk protein polypeptide sequences derived from *D. melanogaster* (Hung and Wensink, 1982, 1983; Garabedian *et al.*, 1987), *C. erythrocephala* (Martinez and Bownes, 1994), and *C. capitata* (Rina and Savakis, 1991) reveals extensive homology in the C-terminal regions of these proteins (see figure 3.2.1). Since homology in several domains is also apparent with vertebrate lipases (Bownes, 1992), care was taken to design oligonucleotides which were only likely to amplify yolk protein encoding sequences. The location of the forward (F^{yp}) and reverse primers (R^{yp}) is illustrated in figure 3.2.1, and the oligonucleotide sequences are presented in figure 3.2.2.

Alignment of Dipteran yolk protein polypeptides

(Figure 3.2.1)

```
      1                                                            50
CcYP2   MNPLTIFCLV AVLLSAATAH ....RGSNAI RNNL....QP SGXLSPRELE
CeYPB   .......... .......... .......... .......... ..........
CcYP1   MNPLKIFCFL ALVIAVASAN KHGKNKDNAG PNSL....KP TDWLSVEELQ
CeYPA   MNPLRIVCVA ALLLAAGSA. ....NGNLSG LNKL....RP SQWLSSSQLD
DmYP1   MNPMRVLSLL A.CLAVAALA KPNGRMDNSV NQAL....KP SQWLSGSQLE
DmYP3   MMSLRICLLA TCLLVAAHAS K......DAS NDRL....KP TKWLTATELE
DmYP2   MNPLRTLCVM ACLLAVAMGN PQSGNRSGRR SNSLDNVEQP SNWVNPREVE
IDENT   ---------- ---------- ---------- ---------- ----------

      51                                                           100
CcYP2   DMPAINEITF EKLQEMPAEE AADLVNKIYH LSQMSRNIEP SYAPSPNQIP
CeYPB   .......... .......... .........H LSQAGRAIEP SFVPKASEIP
CcYP1   SMTAIDDITL QQLENMSVED AERKIEKIYH LSQINHALEP SYVPSPSNVP
CeYPA   KLPRSMEISL QKLESMSVEK GAELMQKLYH LSQINNDLKP SFVPSSSNVP
DmYP1   AIPALDDFTI ERLENMNLER GAELLQQVYH LSQIHHNVEP NYVPSG..IQ
DmYP3   NVPSLNDITW ERLENQPLEQ GAKVIEKIYH VGQIKHDLTP SFVPSPSNVP
DmYP2   ELPNLKEVTL KKLQEMSMEE GATLLDKLYH LSQFNHVFKP DYTPEPSQIR
IDENT   ---------- ---------- ---------* --*------* ---*------

     101                                                           150
CcYP2   AYTYTPTGQR VNFNLNQLVA TAQQQPNFGK QEVTVFITGL PNKSSAMLTA
CeYPB   AFLITPDNKK VNFKLSELPK IAKEEKSFGD EEVTVYITGL PQKTETVKKA
CcYP1   VMLMKPNGQS QQTNHNELVE AAKQQPNFGD EEVTIFITGM PQTSSAVLKA
CeYPA   CYIVKPNGKK VSTSLDKLAS ACKQQPNFGE EEVTILITGL PATTETVRKA
DmYP1   VYVPKPNGDK TVAPLNEMIQ RLKQKQNFGE DEVTIIVTGL PQTSETVKKA
DmYP3   VWIIKSNGQK VECKLNNYVE TAKAQPGFGE DEVTIVLTGL PKTSPAQQKA
DmYP2   GYIVGERGQK IEFNLNTLVE KVKRQQKFGD DEVTIFIQGL PETNTQVQKA
IDENT   ---------- ---------- -------**- -***----*- *--------*
```

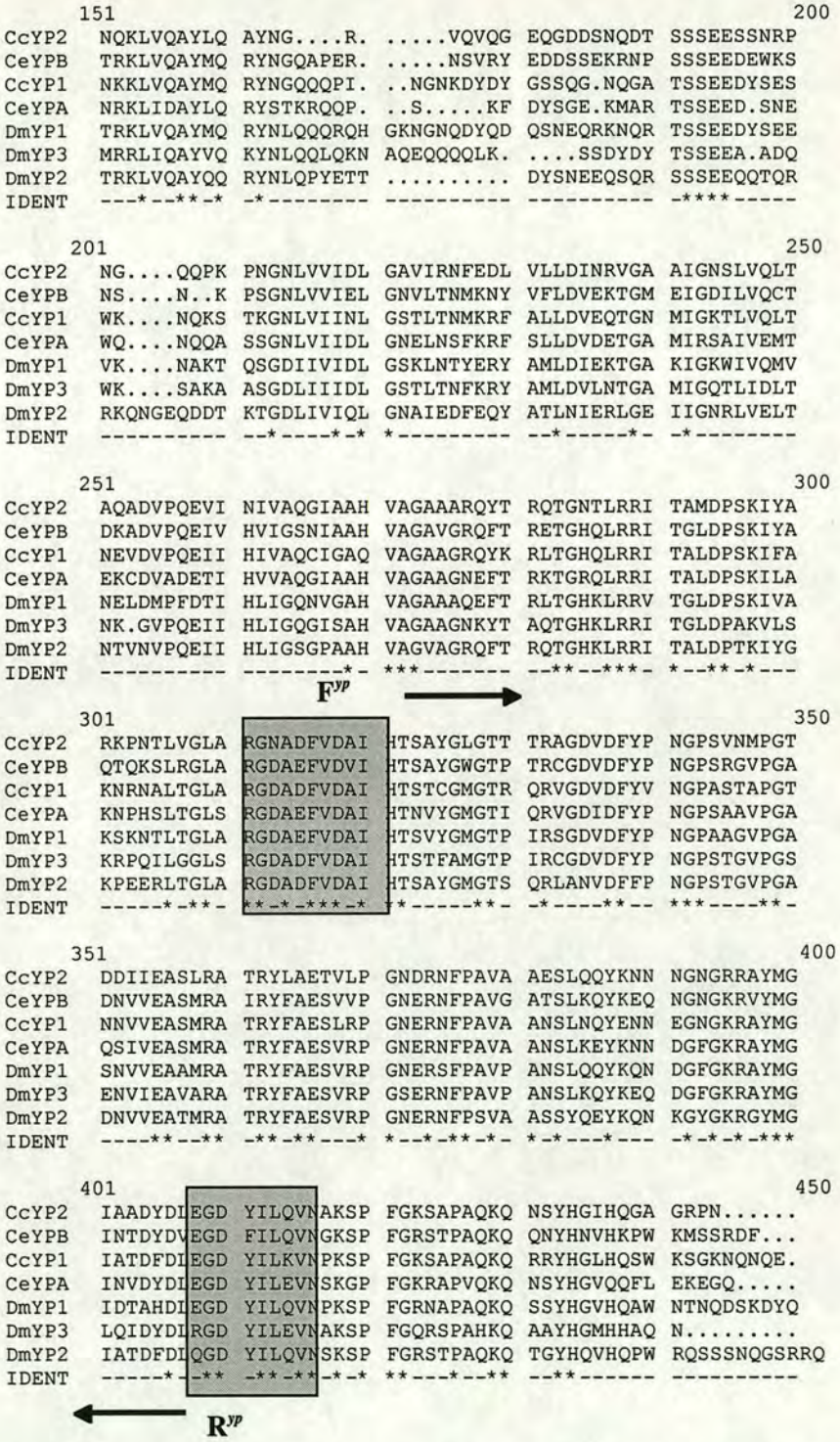


Figure 3.2.1. Alignment of Dipteran yolk protein polypeptides and relevant regions used to design oligonucleotides for cross-species PCR applications. Regions used for the design of primers (F^{yp} and R^{yp}) are boxed. The prefix to each YP sequence (e.g. **Cc**YP1) refers to the species from which the polypeptide derives (Cc = *Ceratitis capitata*, Ce = *Calliphora erythrocephala*, and Dm = *Drosophila melanogaster*) Please refer to text for references. 'IDENT' refers to a conserved amino acid in all of the polypeptides aligned. F^{yp} and R^{yp} refer to the oligonucleotides used in the cross-species PCR (see figure 3.2.2)

Forward primer

Primer F^{yp} : $^{5'}$**GGAATTC GGN GAY GCN GAY TTY GTN GAY GCN**$^{3'}$

*Eco*RI Gly(G) Asp(D) Ala(A) Asp(D) Phe(F) Val(V) Asp(D) Ala(A)

(Minimal *Tm* = 68°C, Maximal *Tm* = 84°C, Redundancy = 4096 fold)

Reverse primer

Primer R^{yp} : $^{5'}$**GGAATTC RTT NAC YTG RAN DAT RTA RTC NCC**$^{3'}$

*Eco*RI Asn(N) Val(V) Gln(Q) Leu(L) Ile(I) Tyr(Y) Asp(D) Gly(G)

(Minimal *Tm* = 56°C, Maximal *Tm* = 76°C, Redundancy = 6144 fold)

Figure 3.2.2. Redundant oligonucleotides used in cross-species PCR amplification procedures. For a description of where these primers are located relative to YP sequences please refer to the text and figure 3.2.1

The anticipated product from PCR amplification of yolk protein encoding cDNA sequences using the F^{yp} and R^{yp} oligonucleotides is approximately 300bp. This is indeed found to be the case, as shown in figure 3.2.3, which shows the products of PCR amplification from a variety of cDNA templates (Note : the optimisation of the RT-PCR procedure, the reaction conditions used, and the PCR cycles used are all detailed in the materials and methods). Product is clearly detected in all positive control template samples (i.e. cloned *D. melanogaster ypl, yp2* and *yp3* gene sequences, as well as *D. melanogaster* first strand cDNA carried out as a control for the cDNA synthesis procedure). In addition, product of the expected size is also amplified from both *C. erythrocephala* and *M. domestica* first strand cDNA templates (the cDNA was synthesised from adult female whole body total RNA extracts).

Since two of the genes encoding *C. erythrocephala* yolk proteins have already been characterised (Martinez and Bownes, 1994), product amplified from *C. erythrocephala* first strand cDNA was not characterised further, and attention was focused on the analysis of products derived from *M. domestica* first strand cDNA templates. These products were gel purified, restriction enzyme digested with *Eco*RI (an *Eco*RI restriction enzyme site was incorporated into the oligonucleotides), and subcloned into the pBluescript SK vector. The efficiency of this procedure was poor, since only ten recombinant colonies were obtained, even though the transformation efficiency of the competent cells was reasonable (approximately 1 x 10^6 cfu/ µg DNA). The most likely

cause of this low efficiency is incomplete digestion of the PCR product prior to ligation with the vector.

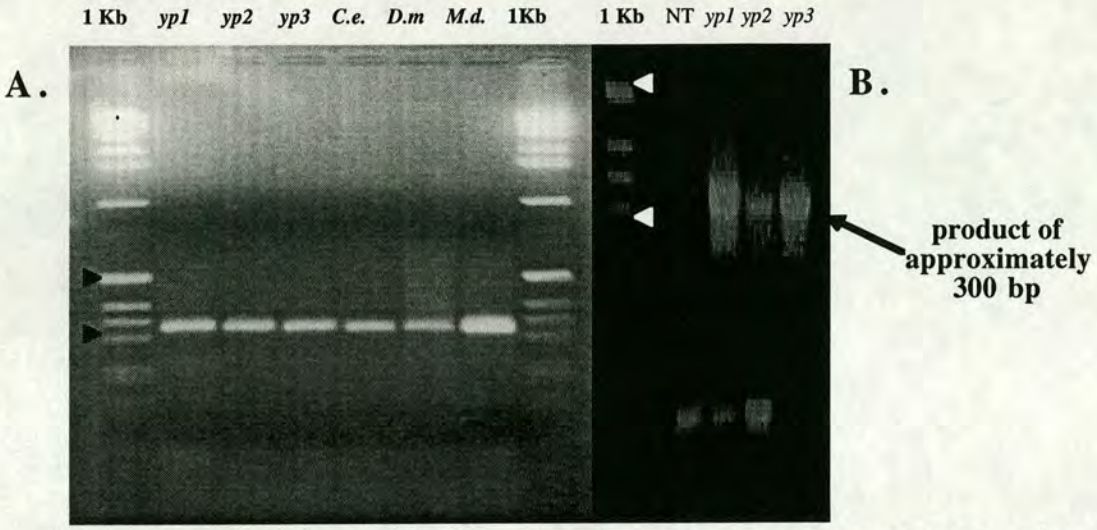


Figure 3.2.3. A. PCR amplification of *yolk protein* encoding sequences using redundant oligonucleotides F^{yp} and R^{yp}. *'yp1'*, *'yp2'* and *'yp3'* refer to *D. melanogaster yolk protein* subclones pGEMYP1, pGEMYP2 and pGEMYP3 respectively (i.e. are positive controls; 10ng template).*'C.e.'*, *'D.m.'* and *'M.d.'* refer to cDNA templates (constructed by reverse transcription of total RNA) derived from *C. erythrocephala, D. melanogaster* and *M. domestica* adult females respectively. Approximately 25ng first strand cDNA template was used in reverse transcriptase PCR reactions. **B. Control Reactions (prior to optimisation).** Product is clearly amplified in reactions containing positive control template DNA (i.e. the same controls as in **A.**), but is absent from reactions lacking a template (NT). NOTE : Equivalent Gibco BRL 1Kb ladder fragments in **A.** and **B.** are marked with arrowheads

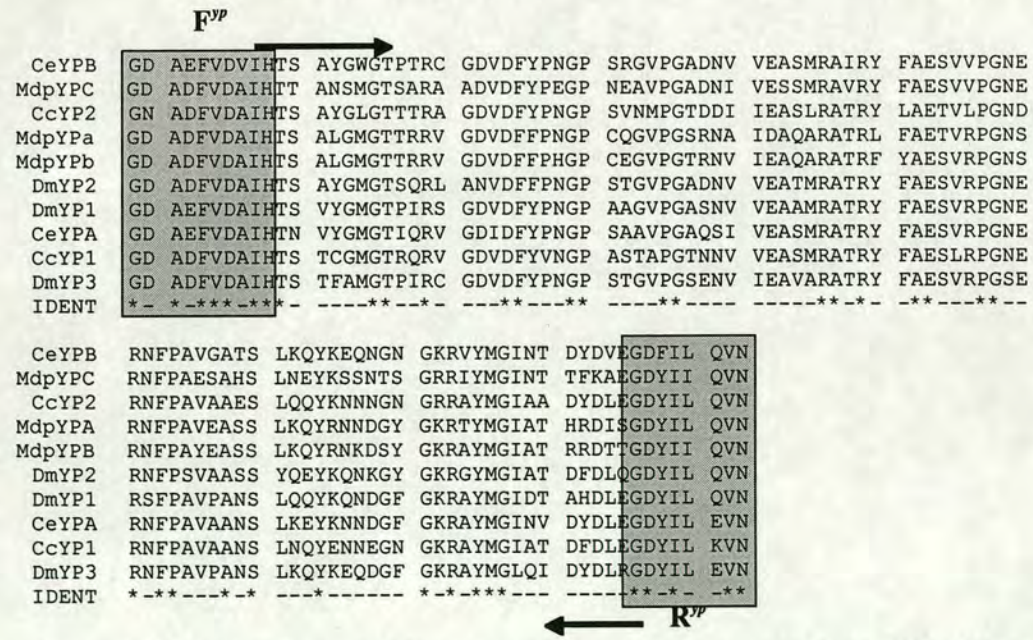


Figure 3.2.4. Alignment of the polypeptide sequences encoded by RT-PCR products from amplification of *M. domestica* 1st strand cDNA with other Dipteran yolk proteins over the amplified region. The oligonucleotide regions are boxed, and the directions of extension are indicated by arrows. Nomenclature of the YP's is as figure 3.2.1. The *M. domestica* PCR products are designated MdpYPA, MdpYPB, and MdpYPC.

Of the ten recombinant colonies obtained, only seven contained inserts of the expected size (300bp). Preliminary sequence analysis of these inserts revealed several recombinants contained identical inserts, and that three independent sequences were present. To characterise these sequences further, individual recombinants representative of each class of insert were analysed further. Both strands of the inserts contained within these recombinants (termed MdpYPA, MdpYPB, and MdpYPC) were sequenced. As illustrated in figure 3.2.4, extensive homology is evident between the polypeptides encoded by these PCR products and other Dipteran yolk protein polypeptides. Significantly however, none of the sequences are identical, supporting the hypothesis that recombinants MdpYPA, MdpYPB and MdpYPC contain sequences representative of three independent *M. domestica yolk protein* genes. Additionally, the fact that these sequences are derived from RNA, as opposed to genomic sequences, infers that these genes are functional.

3.2.2 Isolation of *C. erythrocephala* and *M. domestica* genomic *yp* sequences

The high polypeptide sequence homology contained within the RT-PCR amplified regions and other Dipteran yolk proteins suggests the *M. domestica* RT-PCR products could be used as *yolk protein* specific probes in cross-species hybridisations. Therefore, the 300bp *Eco*RI inserts contained within recombinants MdpYPA, MdpYPB, and MdpYPC were gel purified. These gel purified fragments were mixed to approximately equimolar concentrations, radioactively labelled, and used to screen approximately 1.5 x 10^5 pfu from both *C. erythrocephala* (self constructed, see materials and methods) and *M. domestica* (kindly donated by C. Tortiglione) genomic libraries (a total probe concentration of 10ng/ml hybridisation solution was used; this mixed probe is hereafter referred to as the cross-species *yp* probe number 1 [csYP1]). Since the homology between probe and target sequences was anticipated to be high (identical in the case of *M. domestica* genomic sequences), high stringency hybridisation conditions were employed to reduce the probability of false positive isolation. Therefore, hybridisations were carried out at 42°C in 50% (v/v) Formamide hybridisation solution, and high stringency post-hybridisation washing procedures were carried out.

Numerous positive recombinant phage derived from the *C. erythrocephala* genomic library were identified in the primary screen, and were subsequently characterised further by Claudia Tortiglione. Of seven positive recombinant phage identified in the primary screen of the *M. domestica* genomic library, secondary and tertiary screens demonstrated that only five were true positives. These phage, again, have been characterised further by Claudia Tortiglione. In all cases, the phage remaining after secondary and tertiary screens contain sequences showing strong homology to the original csYP1 probe. Sequence analysis of some of these cross-hybridising fragments confirms that they contain yolk protein encoding sequences (C. Tortiglione, pers.

comm.). The organisation of four of the five *M. domestica* phage isolated (termed MdgYPA to MdgYPE), as determined by Claudia Tortiglione, is shown in figure 3.2.5.

3.2.3 Isolation of *M. domestica* yolk protein encoding cDNA's

Since genes encoding the *M. domestica* yolk proteins had not been isolated previously their stage-, tissue-, and sex-specific expression patterns had only been inferred by the analysis of vitellogenesis in this species. The size of the presumed yolk proteins present in *M. domestica* appears to correlate well with those isolated in other species (i.e. they are between 40 - 50kDa), although conflicting reports have been presented (Adams and Filipi, 1983; DeBianchi *et al.*, 1985; Agui *et al.*, 1985; Martinez, 1991). The highest haemolymph titres of these proteins is apparent during vitellogenic stages of oogenesis, and the ovaries also synthesise vitellogenin (Note : a more comprehensive discussion of housefly vitellogenesis is presented in Results : Section III). Although not characterised individually, present data suggests that the *yolk protein* genes of *M. domestica* are expressed in a similar manner to *D. melanogaster*, and that a minimum of three genes are present.

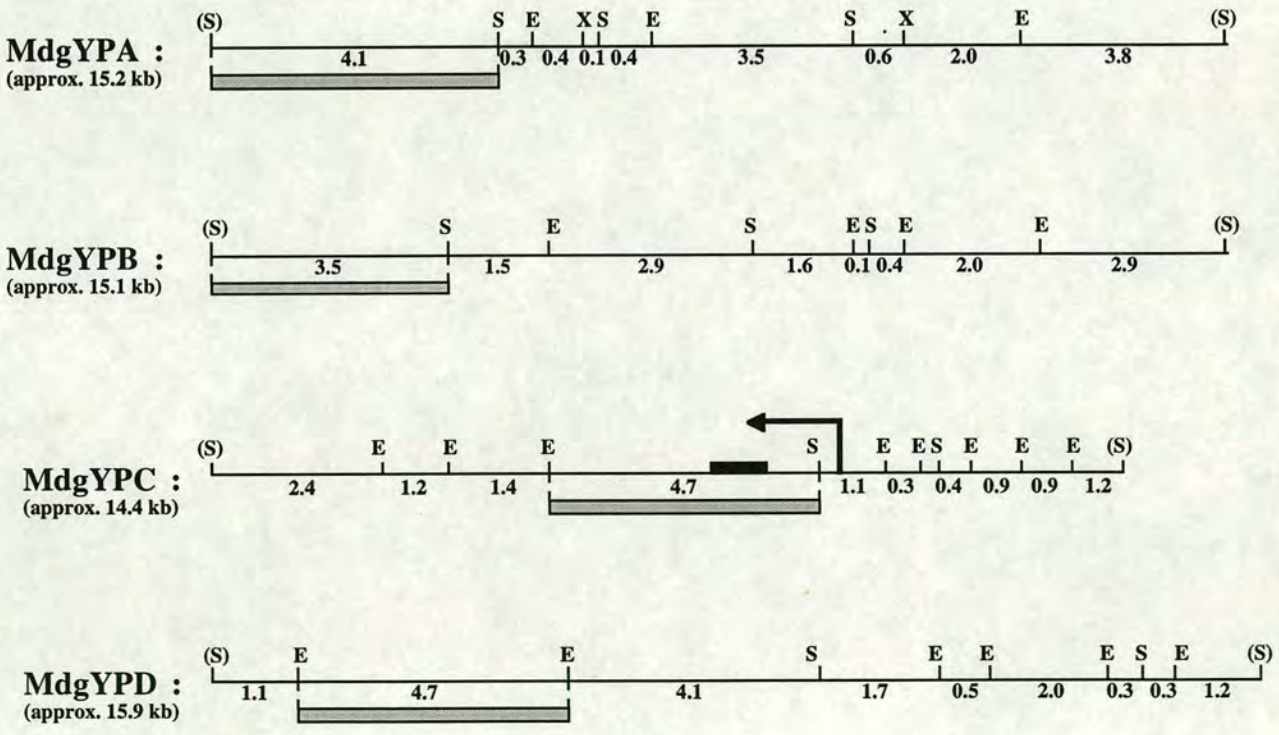


Figure 3.2.5. Organisation of the *Musca domestica yolk protein* genomic lambda bacteriophage clones isolated in library screens using probe csYP1. All data were kindly provided by C. Tortiglione. Lightly shaded boxed regions indicate restriction fragments cross-hybridising with probe csYP1 in subsequent Southern analyses. The black rectangle indicates the location of the PCR amplified region in MdgYPC (determined by subcloning and sequence analysis). The site of transcription initiation and the direction of transcription have been determined in this clone, as indicated by the arrow. All sizes refer to kilobases.

Conclusive determination of the expression profiles of the *M. domestica yolk protein* genes and the relative homology of the encoded proteins to other Dipteran yolk proteins would require the entire coding sequence. Therefore, probe csYP1 was used to screen an *M. domestica* ovarian cDNA library (kindly provided by D. Bopp, University of Zürich; approximately 1.5 x 10^5 pfu were screened). Of twenty putative positive recombinants selected from the numerous positive phage apparent in the primary screen, six failed to cross-hybridise in secondary or tertiary screens. The large number of positives apparent in the primary screen was anticipated since *yp* transcripts would be expected to be enriched in this library if, as suspected, expression occurs in the ovarian follicle cells. Since this cDNA library was constructed in the lambda ZAP II vector, phagemid rescue procedures were employed to isolate the insert cDNA sequences. This procedure circumvents the requirement for subcloning, since direct rescue of pBluescript SK^- recombinant plasmids derived from the recombinant phage is achieved.

Since a large number of recombinants (14) had to be characterised, and the entire pool of recombinants was expected to contain multiple copies of essentially identical inserts, a comparison of the products derived from *Sau*3A digestion of these recombinant plasmids was used to sort them into five preliminary classes (termed MdcYPA to MdcYPE respectively, see figure 3.2.6). Members of each class, although not always completely identical in restriction fragment banding pattern, were sufficiently similar to hypothesise that related sequences were present within the inserts. The apparent differences in banding patterns present within each class would most likely be attributable to variation introduced during first strand cDNA synthesis prior to library construction, such as in the length of extension.

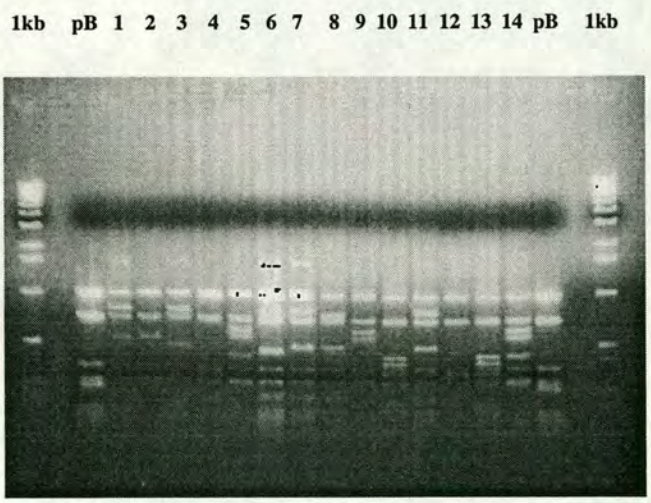


Figure 3.2.6. *Sau 3A* restriction endonuclease characterisation of putative *M. domestica yolk protein* cDNA's. '1kb' is Gibco BRL 1 kilobase ladder and 'pB' refers to *Sau 3A* digested pBluescript SK^- vector.

CLASSES	RECOMBINANTS	SEQUENCED CLONE
MdcYPA	1, 3	1
MdcYPB	2, 5, 9, 14	14
MdcYPC	4	4
MdcYPD	6, 7, 8, 11	7
MdcYPE	10, 12, 13	13

To characterise these insert classes further, one recombinant representative of each insert class (see figure 3.2.6), typically the clone containing the largest apparent insert, was partially sequenced (SK and T7 primers were used to extend into the termini of each clone). This revealed that recombinants MdcYPA and MdcYPD contained identical sequences at their termini, and were therefore likely to contain identical cDNA sequences. Thus, four apparently independent insert classes (MdcYPA, MdcYPB, MdcYPC, and MdcYPE) had been resolved. The entire sense strand sequence of the inserts present within each of these recombinants was determined by the construction of nested deletions, followed by automated sequencing (the nested deletions removed the SK priming site, and thus the T3 primer was used instead). Using the sense strand sequences determined as a template, primers were designed to sequence the anti-sense strand of subclones MdcYPA, MdcYPB and MdcYPE in their entirety.

Only one of the reverse primers designed to the sense strand of clone MdcYPA gave reliable sequence (and extended throughout the majority of the coding region), and so the sense strand was sequenced a second time to identify putative errors in those regions where the antisense strand sequence had not been determined. In this way, the complete sequence of these recombinant inserts was determined conclusively. Recombinants MdcYPA, MdcYPB, and MdcYPE contain three independent yolk protein encoding sequences which show extensive homology to other Dipteran yolk protein polypeptides, and therefore confirms that they do in fact represent transcripts derived from three independent *M. domestica yolk protein* genes (see figure 3.2.8). Recombinant MdcYPC, however, contains a truncated yolk protein encoding sequence fused to an apparently non-coding sequence with homology to mitochondrial DNA. Thus, this recombinant most likely derives from a co-ligation event during cDNA library construction (this view is supported by the fact that this was the only clone identified within the MdcYPC class of recombinant banding patterns, see figure 3.2.6). Regardless of this hypothesis, the *yolk protein* sequence contained within recombinant MdcYPC is essentially identical to the sequence present within MdcYPE, and is therefore redundant.

3.2.4 SEQUENCE ANALYSIS AND DATABASE SEARCHES

The cDNA sequences present within recombinants MdcypA, MdcypB, and MdcypE are presented in Appendix I (note each sequence is primarily named relative to its most likely homologue in *D. melanogaster*, as determined in homology searches). Analysis of these sequences reveals long open reading frames in all three recombinants, and that each recombinant does indeed contain novel sequences (i.e. three independent recombinants had been isolated). The predicted open reading frames (*Mdyp1* [MdcypA], nt 57-1287; *Mdyp2* [MdcypB], nt 96-1418; *Mdyp3* [MdcypE], nt 76-1281) encode putative translation products of 409aa (*Mdyp1* : 44.97kD), 439aa (*Mdyp2* : 48.73 kD), and 401aa (*Mdyp3* : 44.33kD) respectively. Consensus polyadenylation signals

(AAUAAA; reviewed by Proudfoot, 1991) are also apparent at nucleotides 1659-1664 (*Mdyp1)*, 1672-1677 (*Mdyp2*), and 1426-1431 (*Mdyp3*) respectively.

3.2.4.1 Evidence for marginal truncation of recombinant MdcYPB (*Mdyp2*)

The sequence derived from recombinant MdcypB, encompassing some 2kb, is clearly inconsistent with the transcript size (approx. 1.7kb) resolved in the Northern developmental analysis (Results Section III; figure 3.3.1). Close examination of the 5' sequence contained within recombinant MdcypB supports the hypothesis that a co-ligation event has occurred during library construction for several reasons. Firstly, the consensus AATAAA polyadenylation signal (Proudfoot, 1991) is apparent in this 5' leader (nt 295-300), as is a poly-A rich sequence (nt 332-350) downstream of this consensus polyadenylation signal. Secondly, the sequence immediately 3' of the poly-A rich sequence is the consensus sequence for restriction by *XhoI* (CTCGAG; nt 351-356), as would be expected with respect to the procedure employed in library construction (directional *EcoRI* [5']/ *XhoI* [3'] cloning was employed; D. Bopp, pers. comm.). This procedure involves *XhoI* restriction endonuclease digestion subsequent to the addition of *EcoRI* adapters to the 5' termini of newly synthesised cDNA. Thus it is conceivable that the *XhoI* restriction site present within the 5' leader of recombinant MdcypB represents a restriction site present in the endogenous gene sequence from which the MdcypB transcripts originally derived. During library construction, this site would be cleaved, yielding a cDNA which in itself is incapable of insertion into the *EcoRI/ XhoI* λ ZAP II vector, necessitating a coligation event. I therefore propose that this cDNA represents a marginally 5' truncated *Mdyp2* transcript (marginal since the approximate transcript size [1.7kb] evident in Northern hybridisations coincides well with the sequence contained within recombinant MdcypB from the point of truncation to the 3' terminus [1699nt]). The truncated MdcypB *(Mdyp2)* sequence (i.e. the sequence minus the co-ligated leader) is presented in Appendix I, figure A.3.

3.2.4.2 Sequence alignments and comparisons of Dipteran yolk proteins

Swissprot database searches using the entire predicted polypeptides encoded by recombinants MdcypA, MdcypB, and MdcypE were used to identify their most likely homologue in *D. melanogaster* (the GCG8 FASTA program was employed). This search suggested recombinant MdcypA contains sequence representative of the transcript derived from the *M. domestica* homologue of *Dmyp1*, MdcypB represents a homologue of *Dmyp2*, and MdcypE represents a homologue of *Dmyp3*. These recombinants, therefore, are hereafter referred to as *Mdyp1*, *Mdyp2,* and *Mdyp3* respectively in reference to the sequences from which they derive. However, this nomenclature can not be considered truly reliable since the relative identity and similarity values between these proteins is not significantly different (see Table 3.2.1).

An alignment of the polypeptides encoded by *Mdyp1*, *Mdyp2*, and *Mdyp3* with previously isolated Dipteran yolk proteins is presented in figure 3.2.7, and alignments of specific yolk protein families (i.e. restricted to just yolk protein-1, yolk protein-2, and yolk protein-3 homologues) are presented in figures 3.2.8, 3.2.9, and 3.2.10. Consistent with previous findings (Rina and Savakis, 1991; Martinez and Bownes, 1994), a high degree of sequence identity is apparent, particularly in C-terminal regions (the last 2/5ths of the proteins).

Figure 3.2.7 : Pileup alignment of Dipteran yolk proteins

```
          1                                                            50
MdYP1     MNPLGVVCFV AFVAVGALVS QSEDY..S.. .........P KPAYWVKPTE
MdYP3     MNPLVILGFV AMVAVGSLAS PTNQN..S.. .........M KPSQWLKPSE
CcYP2     MNPLTIFCLV AVLLSAATAH ....R..GSN AIRNNL.... QPSGXLSPRE
CeYPB     .......... .......... .......... .......... ..........
CcYP1     MNPLKIFCFL ALVIAVASAN KHGKN..KDN AGPNSL.... KPTDWLSVEE
CeYPA     MNPLRIVCVA ALLLAAGSA. ....N..GNL SGLNKL.... RPSQWLSSSQ
DmYP1     MNPMRVLSLL A.CLAVAALA KPNGR..MDN SVNQAL.... KPSQWLSGSQ
DmYP3     MMSLRICLLA TCLLVAAHAS K........D ASNDRL.... KPTKWLTATE
DmYP2     MNPLRTLCVM ACLLAVAMGN PQSGN..RSG RRSNSLDNVE QPSNWVNPRE
MdYP2     MNPLRTVCLM MGVLALASAY SAGPRPMSMN SNRNTIKNSM KPTSWMSIST

          MNPL------ ---------- ---------- ---------- -P--W-----

          51                                                           100
MdYP1     LGDTPSVNEL TCEELENMPL EKGDTLMCKL YHLSQIDYSV SPNFCPSPTN
MdYP3     LESTPSLDEL TFEELEKMPL EKGAKLMRKI YHLAQIENSV SPNFVPSPSN
CcYP2     LEDMPAINEI TFEKLQEMPA EEAADLVNKI YHLSQMSRNI EPSYAPSPNQ
CeYPB     .......... .......... .......... .HLSQAGRAI EPSFVPKASE
CcYP1     LQSMTAIDDI TLQQLENMSV EDAERKIEKI YHLSQINHAL EPSYVPSPSN
CeYPA     LDKLPRSMEI SLQKLESMSV EKGAELMQKL YHLSQINNDL KPSFVPSSSN
DmYP1     LEAIPALDDF TIERLENMNL ERGAELLQQV YHLSQIHHNV EPNYVPSG..
DmYP3     LENVPSLNDI TWERLENQPL EQGAKVIEKI YHVGQIKHDL TPSFVPSPSN
DmYP2     VEELPNLKEV TLKKLQEMSM EEGATLLDKL YHLSQFNHVF KPDYTPEPSQ
MdYP2     LQSLPSLKEI KLKQLEEMSA FEGADLINRL YHLAQATQAL EPTYAPRASE

          L---P---E- ----L--M-- E--------- YHL-Q----- -P---P----

          101                                                          150
MdYP1     VPVHSFNNKG EKETSNLNKH STLPEEKPKF DEQEVTVFIT GLPQSLEDVK
MdYP3     VPVYIFNGKG EKETCNLNNY VDIAKNKPKF GEQEVTVFIT GLPQSLDDVK
CcYP2     IPAYTYTPTG QRVNFNLNQL VATAQQQPNF GKQEVTVFIT GLPNKSSAML
CeYPB     IPAFLITPDN KKVNFKLSEL PKIAKEEKSF GDEEVTVYIT GLPQKTETVK
CcYP1     VPVMLMKPNG QSQQTNHNEL VEAAKQQPNF GDEEVTIFIT GMPQTSSAVL
CeYPA     VPCYIVKPNG KKVSTSLDKL ASACKQQPNF GEEEVTILIT GLPATTETVR
DmYP1     IQVYVPKPNG DKTVAPLNEM IQRLKQKQNF GEDEVTIIVT GLPQTSETVK
DmYP3     VPVWIIKSNG QKVECKLNNY VETAKAQPGF GEDEVTIVLT GLPKTSPAQQ
DmYP2     IRGYIVGERG QKIEFNLNTL VEKVKRQQKF GDDEVTIFIQ GLPETNTQVQ
MdYP2     IPAFLVTPDN QK.SIQLNEL PRVAREQSHC GKQEVTVFIT GLPSKLESVK

          -P-Y-----G -K----LN-- ---------F GE-EVT-FIT GLP-----V-

          151                                                          200
MdYP1     TPNTKLIQSY IQRYT..... .......... ........KK PEAPQGEDQS
MdYP3     KANTRLIQAY IQRYS..... .......... ........QK PTPPRDDDKS
CcYP2     TANQKLVQAY LQAYNG.... RVQVQGEQG. .....DDSNQ DTSSSEESSN
CeYPB     KATRKLVQAY MQRYNGQAPE RNSVRYEDD. .....SSEKR NPSSSEEDEW
CcYP1     KANKKLVQAY MQRYNGQQQP ...INGNKDY DYGSSQG.NQ GATSSEEDYS
CeYPA     KANRKLIDAY LQRYSTKRQQ ......PSKF DY..SGE.KM ARTSSEED.S
DmYP1     KATRKLVQAY MQRYNLQQQR QHGKNGNQDY QDQSNEQRKN QRTSSEEDYS
DmYP3     KAMRRLIQAY VQKYNLQQLQ KNAQEQQQQL .....KSSDY DYTSSEEA.A
DmYP2     KATRKLVQAY QQRYNLQPYE TTD....... ...YSNEEQS QRSSSEEQQT
MdYP2     EATRSLTQAY MQRYN.DESS YYQNSATSSS SSHYHLNKKQ QRSDSDEDNS

          -A---LVQAY -QRY------ ---------- ---------- ----SEE---

          201                                                          250
MdYP1     K.....WENE KPVGGHLVVI DLGHAITNVE RYATLNVKET GKMIGKTLAE
MdYP3     K.....WENE QPVGGHLVVI DLGHTITDME RYASLDVKET GKMIGKTFAE
CcYP2     RPNG....QQ PKPNGNLVVI DLGAVIRNFE DLVLLDINRV GAAIGNSLVQ
CeYPB     KSNS....N. .KPSGNLVVI ELGNVLTNMK NYVFLDVEKT GMEIGDILVQ
CcYP1     ESWK....NQ KSTKGNLVII NLGSTLTNMK RFALLDVEQT GNMIGKTLVQ
CeYPA     NEWQ....NQ QASSGNLVII DLGNELNSFK RFSLLDVDET GAMIRSAIVE
DmYP1     EEVK....NA KTQSGDIIVI DLGSKLNTYE RYAMLDIEKT GAKIGKWIVQ
DmYP3     DQWK....SA KAASGDLIII DLGSTLTNFK RYAMLDVLNT GAMIGQTLID
DmYP2     QRRKQNGEQD DTKTGDLIVI QLGNAIEDFE QYATLNIERL GEIIGNRLVE
MdYP2     N......... QKPSGCLVVV KFGDTISDFE EHATVDTEKV GKEVGKFWLQ

          ---------- ----G-LVVI -LG------- -Y--LDV--- G--IG--L--

          251                                                          300
MdYP1     LEKESNVDLE DLHVIGQGIG ANVAGAAGKA FKDVTTHKLG RITVLDPARQ
MdYP3     LMDECDVDVE DMHVVAQGIA TNVGGSAGKD FKDITTHKLD RITALDPARQ
CcYP2     LTAQADVPQE VINIVAQGIA AHVAGAAARQ YTRQTGNTLR RITAMDPSKI
CeYPB     CTDKADVPQE IVHVIGSNIA AHVAGAVGRQ FTRETGHQLR RITGLDPSKI
CcYP1     LTNEVDVPQE IIHIVAQCIG AQVAGAAGRQ YKRLTGHQLR RITALDPSKI
CeYPA     MTEKCDVADE TIHVVAQGIA AHVAGAAGNE FTRKTGRQLR RITALDPSKI
DmYP1     MVNELDMPFD TIHLIGQNVG AHVAGAAAQE FTRLTGHKLR RVTGLDPSKI
DmYP3     LTNK.GVPQE IIHLIGQGIS AHVAGAAGNK YTAQTGHKLR RITGLDPAKV
DmYP2     LTNTVNVPQE IIHLIGSGPA AHVAGVAGRQ FTRQTGHKLR RITALDPTKI
MdYP2     LLEKTNCNRD NVHLIGSNLG ANIAGAAGRQ YTKVTNHQLR RITGLDPVKC

          L-----V--E -IH-I---I- A-VAGAAG-- F---T-H-LR RIT-LDP-K-
```

```
          301                                                        350
MdYP1     VGKDPKVLTG LSRCSVKFVD VIHTSALGMG TTRRVGDVDF FSNGTCQGVP
MdYP3     VAKNPKVLSG LARGSANFVD AIHTSALGMG TTRRVGDVDF FPHGPCEGVP
CcYP2     YARKPNTLVG LARGNADFVD AIHTSAYGLG TTTRAGDVDF YPNGPSVNMP
CeYPB     YAQTQKSLRG LARGDAEFVD VIHTSAYGWG TPTRCGDVDF YPNGPSRGVP
CcYP1     FAKNRNALTG LARGDADFVD AIHTSTCGMG TRQRVGDVDF YVNGPASTAP
CeYPA     LAKNPHSLTG LSRGDAEFVD AIHTNVYGMG TIQRVGDIDF YPNGPSAAVP
DmYP1     VAKSKNTLTG LARGDAEFVD AIHTSVYGMG TPIRSGDVDF YPNGPAAGVP
DmYP3     LSKRPQILGG LSRGDADFVD AIHTSTFAMG TPIRCGDVDF YPNGPSTGVP
DmYP2     YGKPEERLTG LARGDADFVD AIHTSAYGMG TSQRLANVDF FPNGPSTGVP
MdYP2     FAKDPETLTG LARGDAEFVD AIHTTANSMG TSARAADVDF YPEGPNEAVP

          --K----L-G L-RG-A-FVD AIHTS-YGMG T--R-GDVDF YPNGP---VP

          351                                                        400
MdYP1     GSRNAIDAQA RATRLFGETV RPGNSRNFPA VEASSLIQYR NNDGYGKRTY
MdYP3     GTRNVIEAQA RATRFYAESV RPGNSRNFPA LEASSLKQYR NKDSYGKRAY
CcYP2     GTDDIIEASL RATRYLAETV LPGNDRNFPA VAAESLQQYK NNNGNGRRAY
CeYPB     GADNVVEASM RAIRYFAESV VPGNERNFPA VGATSLKQYK EQNGNGKRVY
CcYP1     GTNNVVEASM RATRYFAESL RPGNERNFPA VAANSLNQYE NNEGNGKRAY
CeYPA     GAQSIVEASM RATRYFAESV RPGNERNFPA VAANSLKEYK NNDGFGKRAY
DmYP1     GASNVVEAAM RATRYFAESV RPGNERSFPA VPANSLQQYK QNDGFGKRAY
DmYP3     GSENVIEAVA RATRYFAESV RPGSERNFPA VPANSLKQYK EQDGFGKRAY
DmYP2     GADNVVEATM RATRYFAESV RPGNERNFPS VAASSYQEYK QNKGYGKRGY
MdYP2     GADNIVESPM RAVRYFAESV VPGNERNFPA ESAHSLNEYK NSNTSGRRIY

          G--NVVEA-- RATRYFAESV -PGNERNFPA V-A-SL--Y- ---G-GKR-Y

          401                                                        450
MdYP1     MGIATHRDIS GDYMLEVNAE SPYGKRTPAR KQKSYHGFHQ TSYAKSNENY
MdYP3     MGIATRRDTT GDYILEVNEQ TPFGKR.SAP QQRSVQSFNS ENY.......
CcYP2     MGIAADYDLE GDYILQVNAK SPFGKSAPAQ KQNSYHGIHQ GAGRPN....
CeYPB     MGINTDYDVE GDFILQVNGK SPFGRSTPAQ KQQNYHNVHK PWKMSSRDF.
CcYP1     MGIATDFDLE GDYILKVNPK SPFGKSAPAQ KQRRYHGLHQ SWKSGKNQNQE
CeYPA     MGINVDYDLE GDYILEVNSK GPFGKRAPVQ KQNSYHGVQQ FLEKEGQ...
DmYP1     MGIDTAHDLE GDYILQVNPK SPFGRNAPAQ KQSSYHGVHQ AWNTNQDSKDYQ
DmYP3     MGLQIDYDLR GDYILEVNAK SPFGQRSPAH KQAAYHGMHH AQN.......
DmYP2     MGIATDFDLQ GDYILQVNSK SPFGRSTPAQ KQTGYHQVHQ PWRQSSSNQGSRRQ
MdYP2     MGIITTFKVE GDYMLQVNTK SPFGRSTPVQ KQQNVHGVHK SWKMSSSRDEE

          MGI----D-- GDYIL-VN-K SPFG---PA- KQ--YH--H- ----------
```

Figure 3.2.7. Pileup alignment Dipteran yolk protein polypeptides. Species are indicated by the first two name characters (i.e. 'Md' represents *M. domestica*, 'Cc' represents *C. capitata*, 'Ce' represents *C. erythrocephala*, and 'Dm' refers to *D. melanogaster*), whereas the final three characters define the class of yolk protein. A consensus sequence (requiring an identity or conservative substitution in eight out of the 10 aligned sequences) is indicated below the alignment. The conserved SSEE and GDADFVDA motifs referred to in the main text are boxed.

```
          1                                                            50
CcYP1     MNPLKIFCFL ALVIAVASAN KHGKNKDNAG PNSLKPTDWL SVEELQSMTA
CeYPA     MNPLRIVCVA ALLLAAGSA. ....NGNLSG LNKLRPSQWL SSSQLDKLPR
DmYP1     MNPMRVLSLL A.CLAVAALA KPNGRMDNSV NQALKPSQWL SGSQLEAIPA
MdYP1     MNPLGVVCFV AF.VAVGAL. .....VSQSE DYSPKPAYWV KPTELGDTPS

           MNPL------ A---A----- ---------- -----P--W- ----L-----

          51                                                          100
CcYP1     IDDITLQQLE NMSVEDAERK IEKIYHLSQI NHALEPSYVP SPSNVPVMLM
CeYPA     SMEISLQKLE SMSVEKGAEL MQKLYHLSQI NNDLKPSFVP SSSNVPCYIV
DmYP1     LDDFTIERLE NMNLERGAEL LQQVYHLSQI HHNVEPNYVP SG..IQVYVP
MdYP1     VNELTCEELE NMPLEKGDTL MCKLYHLSQI DYSVSPNFCP SPTNVPVHSF

           --------LE -M--E----- ----YHLSQI -----P---P S---V-----

          101                                                         150
CcYP1     KPNGQSQQTN HNELVEAAKQ QPNFGDEEVT IFITGMPQTS SAVLKANKKL
CeYPA     KPNGKKVSTS LDKLASACKQ QPNFGEEEVT ILITGLPATT ETVRKANRKL
DmYP1     KPNGDKTVAP LNEMIQRLKQ KQNFGEDEVT IIVTGLPQTS ETVKKATRKL
MdYP1     NNKGEKETSN LNKHSTLPEE KPKFDEQEVT VFITGLPQSL EDVKTPNTKL

           ---G------ ---------- ---F-E-EVT I-ITGLP--- --V-----KL

          151                                                         200
CcYP1     VQAYMQRYNG QQQP...ING NKDYDYGSSQ G.NQGATSSE EDYSESWKNQ
CeYPA     IDAYLQRYST KRQQ..... PSKFDY..SG E.KMARTSSE ED.SNEWQNQ
DmYP1     VQAYMQRYNL QQQRQHGKNG NQDYQDQSNE QRKNQRTSSE EDYSEEVKNA
MdYP1     IQSYIQRYT. ......... ......... ..KKPEAPQG EDQSK.WENE

           ---Y-QRY-- ---------- ---------- ---------- ED-S----N-

          201                                                         250
CcYP1     KSTKGNLVII NLGSTLTNMK RFALLDVEQT GNMIGKTLVQ LTNEVDVPQE
CeYPA     QASSGNLVII DLGNELNSFK RFSLLDVDET GAMIRSAIVE MTEKCDVADE
DmYP1     KTQSGDIIVI DLGSKLNTYE RYAMLDIEKT GAKIGKWIVQ MVNELDMPFD
MdYP1     KPVGGHLVVI DLGHAITNVE RYATLNVKET GKMIGKTLAE LEKESNVDLE

           ----G--V-I -LG------- R---L-V--T G--I------ ---------E

          251                                                         300
CcYP1     IIHIVAQCIG AQVAGAAGRQ YKRLTGHQLR RITALDPSKI FAKNRNALTG
CeYPA     TIHVVAQGIA AHVAGAAGNE FTRKTGRQLR RITALDPSKI LAKNPHSLTG
DmYP1     TIHLIGQNVG AHVAGAAAQE FTRLTGHKLR RVTGLDPSKI VAKSKNTLTG
MdYP1     DLHVIGQGIG ANVAGAAGKA FKDVTTHKLG RITVLDPARQ VGKDPKVLTG

           --H---Q-I- A-VAGAA--- F---T---L- RIT-LDP--- --K----LTG

          301                                                         350
CcYP1     LARGDADFVD AIHTSTCGMG TRQRVGDVDF YVNGPASTAP GTNNVVEASM
CeYPA     LSRGDAEFVD AIHTNVYGMG TIQRVGDIDF YPNGPSAAVP GAQSIVEASM
DmYP1     LARGDAEFVD AIHTSVYGMG TPIRSGDVDF YPNGPAAGVP GASNVVEAAM
MdYP1     LSRCSVKFVD VIHTSALGMG TTRRVGDVDF FSNGTCQGVP GSRNAIDAQA

           L-R----FVD -IHT---GMG T--R-GDVDF Y-NG-----P G----VEA--

          351                                                         400
CcYP1     RATRYFAESL RPGNERNFPA VAANSLNQYE NNEGNGKRAY MGIATDFDLE
CeYPA     RATRYFAESV RPGNERNFPA VAANSLKEYK NNDGFGKRAY MGINVDYDLE
DmYP1     RATRYFAESV RPGNERSFPA VPANSLQQYK QNDGFGKRAY MGIDTAHDLE
MdYP1     RATRLFGETV RPGNSRNFPA VEASSLIQYR NNDGYGKRTY MGIATHRDIS

           RATR-F-E-- RPGN-R-FPA V-A-SL--Y- -NDG-GKR-Y MGI----D--

          401                                                    442
CcYP1     GDYILKVNPK SPFGKSAPAQ KQRRYHGLHQ SWKSGKNQNQ E.
CeYPA     GDYILEVNSK GPFGKRAPVQ KQNSYHGVQQ FLEKEGQ... ..
DmYP1     GDYILQVNPK SPFGRNAPAQ KQSSYHGVHQ AWNTNQDSKD YQ
MdYP1     GDYMLEVNAE SPYGKRTPAR KQKSYHGFHQ TSYAKSNENY ..

           GDY-L-VN-- -PFG---P-- KQ--YHG--Q ---------- --
```

Figure 3.2.8. Pileup Alignment of Dipteran yolk protein-1 homologues. Please refer to figure 3.2.7 for a description of the nomenclature. Identity or conservative substitutions are indicated in the consensus sequence below the alignment.

```
          1                                                            50
CcYP2     MNPLTIFCLV AVLLSAATAH ........RG S..NAIRNNL QPSGXLSPRE
CeYPB     .......... .......... .......... .......... ..........
DmYP2     MNPLRTLCVM ACLLAVAMGN PQSGNRSGRR S..NSLDNVE QPSNWVNPRE
MdYP2     MNPLRTVCLM MGVLALASAY SAGPRPMSMN SNRNTIKNSM KPTSWMSIST

           ---------- ---------- ---------- ---------- ----------

          51                                                           100
CcYP2     LEDMPAINEI TFEKLQEMPA EEAADLVNKI YHLSQMSRNI EPSYAPSPNQ
CeYPB     .......... .......... .......... .HLSQAGRAI EPSFVPKASE
DmYP2     VEELPNLKEV TLKKLQEMSM EEGATLLDKL YHLSQFNHVF KPDYTPEPSQ
MdYP2     LQSLPSLKEI KLKQLEEMSA FEGADLINRL YHLAQATQAL EPTYAPRASE

           ---------- ---------- ---------- -HL-Q----- -P-Y-P----

          101                                                          150
CcYP2     IPAYTYTPTG QRVNFNLNQL VATAQQQPNF GKQEVTVFIT GLPNKSSAML
CeYPB     IPAFLITPDN KKVNFKLSEL PKIAKEEKSF GDEEVTVYIT GLPQKTETVK
DmYP2     IRGYIVGERG QKIEFNLNTL VEKVKRQQKF GDDEVTIFIQ GLPETNTQVQ
MdYP2     IPAFLVTPDN QK.SIQLNEL PRVAREQSHC GKQEVTVFIT GLPSKLESVK

           I--------- ------L--L ---------- G--EVTVFI- GLP-------

          151                                                          200
CcYP2     TANQKLVQAY LQAYNG.... .....RVQVQ GEQGDDSNQD TSSSEESS..
CeYPB     KATRKLVQAY MQRYNG.... .QAPERNSVR YEDDSSEKRN PSSSEEDE..
DmYP2     KATRKLVQAY QQRYNL.... .QPYETTDYS NEEQS....Q RSSSEEQQTQ
MdYP2     EATRSLTQAY MQRYNDESSY YQNSATSSSS SHYHLNKKQQ RSDSDED...

           -A---L-QAY -Q-YN----- ---------- ---------- -S-SEE----

          201                                                          250
CcYP2     .NRPNGQQPK PN.GNLVVID LGAVIRNFED LVLLDINRVG AAIGNSLVQL
CeYPB   . WKSNSN..K PS.GNLVVIE LGNVLTNMKN YVFLDVEKTG MEIGDILVQC
DmYP2     RRKQNGEQDD TKTGDLIVIQ LGNAIEDFEQ YATLNIERLG EIIGNRLVEL
MdYP2   .. ..NSNQKP S..GCLVVVK FGDTISDFEE HATVDTEKVG KEVGKFWLQL

           ----N----- ---G-LVVI- LG-------- ---------G --IG------

          251                                                          300
CcYP2     TAQADVPQEV INIVAQGIAA HVAGAAARQY TRQTGNTLRR ITAMDPSKIY
CeYPB     TDKADVPQEI VHVIGSNIAA HVAGAVGRQF TRETGHQLRR ITGLDPSKIY
DmYP2     TNTVNVPQEI IHLIGSGPAA HVAGVAGRQF TRQTGHKLRR ITALDPTKIY
MdYP2     LEKTNCNRDN VHLIGSNLGA NIAGAAGRQY TKVTNHQLRR ITGLDPVKCF

           --------E- ---I-----A -VAG---RQ- T--T---LRR IT-LDP-K-Y

          301                                                          350
CcYP2     ARKPNTLVGL ARGNADFVDA IHTSAYGLGT TTRAGDVDFY PNGPSVNMPG
CeYPB     AQTQKSLRGL ARGDAEFVDV IHTSAYGWGT PTRCGDVDFY PNGPSRGVPG
DmYP2     GKPEERLTGL ARGDADFVDA IHTSAYGMGT SQRLANVDFF PNGPSTGVPG
MdYP2     AKDPETLTGL ARGDAEFVDA IHTTANSMGT SARAADVDFY PEGPNEAVPG

           ------L-GL ARG-A-FVD- IHT-A---GT --R---VDFY P-GP----PG

          351                                                          400
CcYP2     TDDIIEASLR ATRYLAETVL PGNDRNFPAV AAESLQQYKN NNGNGRRAYM
CeYPB     ADNVVEASMR AIRYFAESVV PGNERNFPAV GATSLKQYKE QNGNGKRVYM
DmYP2     ADNVVEATMR ATRYFAESVR PGNERNFPSV AASSYQEYKQ NKGYGKRGYM
MdYP2     ADNIVESPMR AVRYFAESVV PGNERNFPAE SAHSLNEYKN SNTSGRRIYM

           -D--VE--MR A-RYFAE-V- PGNERNFP-- -A-S---YK- ----G-R-YM

          401                                                          450
CcYP2     GIAADYDLEG DYILQVNAKS PFGKSAPAQK QNSYHGIHQG AGRPN.....
CeYPB     GINTDYDVEG DFILQVNGKS PFGRSTPAQK QQNYHNVHKP WKMSSRDF..
DmYP2     GIATDFDLQG DYILQVNSKS PFGRSTPAQK QTGYHQVHQP WRQSSSNQGSRRQ
MdYP2     GIITTFKVEG DYMLQVNTKS PFGRSTPVQK QQNVHGVHKS WKMSSSRDEE

           GI-------G DY-LQVN-KS PFG-S-P-QK Q---H-VH-- ----------
```

Figure 3.2.9. Pileup Alignment of Dipteran yolk protein-2 homologues. Please refer to figure 3.2.7 for a description of the nomenclature. Identity or conservative substitutions are indicated in the consensus sequence below the alignment.

```
          1                                                            50
DmYP3     MMSLRICLLA TCLLVAAHAS KDASNDRLKP TKWLTATELE NVPSLNDITW
MdYP3     MNPLVILGFV AMVAVGSLAS PTNQNS.MKP SQWLKPSELE STPSLDELTF

           M--L-I---- ----V---AS ----N---KP --WL---ELE --PSL---T-

          51                                                           100
DmYP3     ERLENQPLEQ GAKVIEKIYH VGQIKHDLTP SFVPSPSNVP VWIIKSNGQK
MdYP3     EELEKMPLEK GAKLMRKIYH LAQIENSVSP NFVPSPSNVP VYIFNGKGEK

           E-LE--PLE- GAK---KIYH --QI-----P -FVPSPSNVP V-I----G-K

          101                                                          150
DmYP3     VECKLNNYVE TAKAQPGFGE DEVTIVLTGL PKTSPAQQKA MRRLIQAYVQ
MdYP3     ETCNLNNYVD IAKNKPKFGE QEVTVFITGL PQSLDDVKKA NTRLIQAYIQ

           --C-LNNYV- -AK--P-FGE -EVT---TGL P-------KA --RLIQAY-Q

          151                                                          200
DmYP3     KYNLQQLQKN AQEQQQQLKS SDYDYTSSEE AADQWKSAKA ASGDLIIIDL
MdYP3     RY........ .......... SQKPTPPRDD DKSKWENEQP VGGHLVVIDL

           -Y-------- ---------- S--------- ----W----- --G-L--IDL

          201                                                          250
DmYP3     GSTLTNFKRY AMLDVLNTGA MIGQTLIDLT NK.GVPQEII HLIGQGISAH
MdYP3     GHTITDMERY ASLDVKETGK MIGKTFAELM DECDVDVEDM HVVAQGIATN

           G-T-T---RY A-LDV--TG- MIG-T---L- ----V--E-- H---QGI---

          251                                                          300
DmYP3     VAGAAGNKYT AQTGHKLRRI TGLDPAKVLS KRPQILGGLS RGDADFVDAI
MdYP3     VGGSAGKDFK DITTHKLDRI TALDPARQVA KNPKVLSGLA RGSANFVDAI

           V-G-AG---- --T-HKL-RI T-LDPA---- K-P--L-GL- RG-A-FVDAI

          301                                                          350
DmYP3     HTSTFAMGTP IRCGDVDFYP NGPSTGVPGS ENVIEAVARA TRYFAESVRP
MdYP3     HTSALGMGTT RRVGDVDFFP HGPCEGVPGT RNVIEAQARA TRFYAESVRP

           HTS---MGT- -R-GDVDF-P -GP--GVPG- -NVIEA-ARA TR--AESVRP

          351                                                          400
DmYP3     GSERNFPAVP ANSLKQYKEQ DGFGKRAYMG LQIDYDLRGD YILEVNAKSP
MdYP3     GNSRNFPALE ASSLKQYRNK DSYGKRAYMG IATRRDTTGD YILEVNEQTP

           G--RNFPA-- A-SLKQY--- D--GKRAYMG -----D--GD YILEVN---P

          401                  421
DmYP3     FGQRSPAHKQ AAYHGMHHAQ N
MdYP3     FGKRS.APQQ RSVQSFNSEN Y

           FG-RS-A--Q ---------- -
```

Figure 3.2.10. Pileup Alignment of Dipteran yolk protein-3 homologues. Please refer to figure 3.2.7 for a description of the nomenclature. Identity or conservative substitutions are indicated in the consensus sequence below the alignment

	MdYP1		MdYP2		MYP3	
	SIM	IDE	SIM	IDE	SIM	IDE
MdYP1			58.77	40.99	82.54	69.08
MdYP2					63.50	42.25
MdYP3						
CcYP1	64.30	47.19	65.21	47.24	69.58	52.12
CcYP2	62.72	44.69	66.43	46.28	64.59	46.88
CeYPA	63.05	46.80	65.31	45.69	67.83	51.37
CeYPB	61.95	46.02	73.01	58.24	65.17	47.15
DmYP1	64.37	46.93	65.66	46.40	67.67	48.87
DmYP2	61.37	46.46	64.83	48.51	65.34	46.14
DmYP3	64.59	48.13	60.67	41.01	68.5	51.0

Table 3.2.1. Percentage amino acid similarity (SIM) and identity (IDE) between Dipteran yolk proteins (derived from gap alignments, GCG8)

In addition to pileup alignments, percentage amino acid similarity and identity between the various Dipteran yolk protein polypeptides was determined using gap alignment software (GCG8), and are presented in table 3.2.1. Clearly, in the majority of instances, only minor variation in these values is observed between species. Of interest, however, is the high degree of similarity (82.54%) and identity (69.08%) apparent between the polypeptides encoded by *Mdyp1* and *Mdyp3*, suggesting these two genes may have arisen relatively recently as a result of gene duplication, as is proposed to be the case for *Dmyp1* and *Dmyp2*. Indeed, the dendogram relationship predicted by alignments with pig and rat lipases over the conserved lipid binding domain suggests this may be the case (see figure 3.2.12). This would therefore suggest that despite the FASTA predictions, the gene we have termed Mdyp3 may in fact represent a homologue of Dmyp2 and *vice* versa. However, an analysis of the genomic organisation of each of the three *Musca domestica yolk protein* genes would be required to determine this conclusively.

In general, the *M. domestica* yolk protein sequences fit the pileup alignments well, although several features are of interest. There appears to be poor conservation of the SSEE motif (see figure 3.2.7), a region of unknown function (although highly conserved between previously isolated Dipteran yolk proteins), particularly in MdYP1 and MdYP3. This motif is more conserved in MdYP2, although their is still evidence of divergence (notably several Aspartic acid [D] residues located in place of the first Serine

and the third Glutamic acid residues [i.e. the sequence DSDE is observed]). In addition, there is evidence for sequence divergence in the highly conserved GDADFVDA motif (see figure 3.2.7) implicated in the association of ecdysteroid conjugates (Bownes, 1992), in MdYP1 and MdYP3 (the sequences CSVKFVDV and GSANFVDA are observed respectively). The relevance of these substitutions will be discussed more fully in the discussion section.

3.2.4.3 POST TRANSLATIONAL MODIFICATIONS

3.2.4.3.1 SIGNAL PEPTIDE SEQUENCES

The signal peptide localisation signals present within the N-terminal *M. domestica* yolk protein sequences were constrained according to consensus sequences derived by Von Heijne (1983) and Perlman and Halvorsen (1983). Essentially, the signal peptide comprises three domains; the 'N', 'H', and 'C' domains respectively. The 'N' region, typically ranging in size from 1-5 residues, should contain at least one positively charged residue (i.e. Lysine [K] or Arginine [R[). The 'H', or central hydrophobic region, comprises 7-20 residues and typically contains the amino acids Phenylalanine (F), Leucine (L), Isoleucine (I), Valine (V), Alanine (A), Methionine (M), and Tryptophan (W). Finally the 'C' region, typically more polar in nature than the 'H' domain, ranges in size from 4-7 residues. Those residues at positions -3 and -1 relative to the site of cleavage should be small neutral amino acids (i.e. Alanine [A], Serine [S], Glycine [G], Cystine [C], Threonine [T], or Glutamine [Q]), and the residue at position -2 should be a large aromatic residue although a hydrophobic/ charged residue is tolerated.

3.2.4.3.1.1 *Musca domestica* yolk protein-1 signal peptide

The 'N' region most likely comprises residues 1-3, although a charged residue is conspicuously absent from this domain. The 'H' region, most easily defined relative to the site of cleavage, probably comprises residues 4-18, since the S^{20} and S^{22} residues conform with the -3/ -1 consensus sequence. The Glutamine residue (Q^{21}) positioned at -2 relative to the site of cleavage would, however, not be considered optimal since it is neither aromatic nor hydrophobic. Thus, the 'C' domain is predicted to comprise residues 19-22, and cleavage is therefore predicted to occur between residues S^{22} and E^{23}.

3.2.4.3.1.2 *Musca domestica* yolk protein-2 signal peptide

The 'N' domain comprises residues 1-5, and contains a positively charged residue (R^5) typically found within this region. Residues A^{19} and S^{21} conform to the predicted -3/ -1 consensus in the 'C' region, as does the large aromatic Tyrosine residue (Y^{20}) located at the putative -2 position. Taking this into consideration, the 'H' domain most likely comprises residues 6-17, the 'C' region residues 18-21, and cleavage is predicted to occur between residues S^{21} and A^{22}.

3.2.4.3.1.3 *Musca domestica* yolk protein-3 signal peptide

Similar to MdYP1, the 'N' region of the MdYP3 signal peptide (most likely residues 1-3) lacks a positively charged residue typically found within this domain. Residues S^{20} and T^{22} conform well to the -3/ -1 consensus, and the aromatic Proline residue (P^{21}) located at the putative -2 position is consistent with cleavage occurring between residues T^{22} and N^{23}. Thus the 'H' domain is predicted to comprise residues 4-18, and the 'C' domain residues 19-22.

3.2.4.3.2 TYROSINE SULPHATION

Putative Tyrosine sulphation sites present within the *Musca domestica* yolk protein polypeptides were identified using the consensus constraints derived by Huttner and Baeuerle (1988). In all three polypeptides tyrosine sulphation sites with good homology to the consensus were identified. Tyrosine sulphation in both MdYP1 and MdYP3 is predicted to occur at Tyrosine^{372} (Y^{372}), since the residues flanking these Tyrosine residues in both instances show good homology to the consensus sequence. MdYP2 contains two putative Tyrosine sulphation sites (Y^{329} and Y^{401}), both of which show good similarity to the consensus, although a third Aspartic acid (D) or Glutamic acid (E) residue would normally be expected in those five residues either side of Tyrosine^{401} (i.e. -5 to +5 relative to the Tyrosine residue a minimum of 3 D/E residues are normally observed).

3.2.4.3.3 N-GLYCOSYLATION

Prosite database searches using the entire predicted polypeptides encoded by *Mdyp1, Mdyp2* and *Mdyp3* suggest only the polypeptides encoded by *Mdyp1* and *Mdyp2* contain putative N-glycosylation sites at residues N^{302} [MdYP1] and N^{381} [MdYP2] respectively. It should be noted, however, that it is not known if previously identified Dipteran yolk proteins are in fact glycosylated *in-vivo*, even though consensus sites for such modifications have been identified.

3.2.5 SECTION II : DISCUSSION

The high degree of homology evident between the yolk proteins encoded by the *Ceratitis capitata, Calliphora erythrocephala* and *Drosophila melanogaster yolk protein* genes facilitated the design of redundant oligonucleotides necessary for cross-species PCR amplification procedures. These oligonucleotides have been used to isolate three novel and independent yolk protein encoding sequences derived from the *M. domestica* genome.

Clearly in all cases these *M. domestica* yolk proteins show a high degree of homology with previously isolated Dipteran yolk proteins, particularly in C-terminal regions, suggesting a high degree of selective pressure to maintain the structure and function of these polypeptides. Early suggestions that these proteins function merely as a

nutritional source for the developing embryo now seem unlikely, and indeed although this is clearly an important function of these proteins, it must be of secondary importance. This is evident since if the primary function of these proteins is merely as a nutritional source, presumably only those domains required for the correct localisation of the proteins would be conserved. It now seems most likely that the primary role of the yolk proteins is to act as a storage molecule for inactivated ecdyesteroid conjugates, and as such the controlled decay of yolk spheres in the embryo would act as a master regulator of the hormonal control of embryogenesis. In this case, the two functions of the yolk proteins would be intrinsically linked. The controlled decay of yolk spheres would be a pre-requisite for correct hormone release, and in addition this constrains the decay of the yolk spheres to a defined time frame (i.e. yolk is degraded at a rate consistent with the embryos requirement for amino acids).

It is interesting to note, therefore, that domains implicated in the association of ecdysteroid conjugates (particularly the motif GDADFVDA) do not appear to be well conserved in two of the three *M. domestica* yolk proteins. Structural analysis suggests in the case of the lipases, this region forms one of four β-sheets over the conserved region, and that these β-sheets form a large hydrophobic pocket around the active site (Persson *et al.*, 1991). However, FASTA database comparisons over the YP region conserved with pig pancreatic triacylglycerol lipase (De Caro *et al.*, 1981) and rat hepatic lipase (Komaromy and Schotz, 1987), i.e. the lipid binding domain of the lipases (see Persson *et al.*, 1989), reveals that the *Musca domestica* yolk proteins are more closely related to the lipase sequences than their *Drosophila* counterparts, suggesting these amino acid substitutions may in fact have little significance with respect to ecdysteroid binding (see figure 3.2.11 and the dendogram presented in figure 3.2.12). Of the substitutions present in the *Musca* yolk protein sequences in highly conserved residues, there seems to be interchange between Alanine to Valine residues, and Aspartic acid (D) and Serine residues (S). Both Alanine (A) and Valine (V) are neutral hydrophobic residues which would not interfere with the formation of β-sheets structures, although the bulkier nature of the Valine side chain may introduce minor conformational changes. Also, substitution of Aspartic acid with Serine is unlikely to interrupt the β-sheet, and it therefore seems likely that these amino acid substitutions will have little effect on the ability of the *Musca* yolk proteins to bind ecdysteroid conjugates. This will however require confirmation by direct measurement of bound ecdysteroid concentrations.

```
          251                                                        300
MdYP1      LEKESNVDLE DLHVIGQGIG ANVAGAAGKA FKDVTTHKLG RITVLDPARQ
mdYP2      LLEKTNCNRD NVHLIGSNLG ANIAGAAGRQ YTKVTNHQLR RITGLDPVKC
mdYP3      LMDECDVDVE DMHVVAQGIA TNVGGSAGKD FKDITTHKLD RITALDPARQ
dmYP1      MVNELDMPFD TIHLIGQNVG AHVAGAAAQE FTRLTGHKLR RVTGLDPSKI
dmYP2      LTNTVNVPQE IIHLIGSGPA AHVAGVAGRQ FTRQTGHKLR RITALDPTKI
dmYP3      LTNK.GVPQE IIHLIGQGIS AHVAGAAGNK YTAQTGHKLR RITGLDPAKV
pig.lipp   LKSSLGYSPS NVHVIGHSLG SHAAGEAGRR TN..G..TIE RITGLDPAEP
rat.liph   LEESMKFSRS KVHLIGYSLG AHVSGFAGSS MG..GKRKIG RITGLDPAGP

CON        L---------- --H-IG---- --V-G-AG-- ---------- RIT-LDP---

          301                                                        350
mdYP1      VGKDPKVLTG LSRCSVKFVD VIHTSA.... ..LGMGTTRR VGDVDFFSNG
mdYP2      FAKDPETLTG LARGDAEFVD AIHTTA.... ..NSMGTSAR AADVDFYPEG
mdYP3      VAKNPKVLSG LARGSANFVD AIHTSA.... ..LGMGTTRR VGDVDFFPHG
dmYP1      VAKSKNTLTG LARGDAEFVD AIHTSV.... ..YGMGTPIR SGDVDFYPNG
dmYP2      YGKPEERLTG LARGDADFVD AIHTSA.... ..YGMGTSQR LANVDFFPNG
dmYP3      LSKRPQILGG LSRGDADFVD AIHTST.... ..FAMGTPIR CGDVDFYPNG
pig.lipp   CFQGTPELVR LDPSDAKFVD VIHTDAAPII PNLGFGMSQT VGHLDFFPNG
rat.liph   MFEGTSPNER LSPDDANFVD AIHT.FTREH MGLSVGIKQP IAHYDFYPNG

CON        -------L-- L----A-FVD -IHT------ --L--G---- ----DF-P-G


          351                                                        400
mdYP1      TCQGVPGSRN AID..... ..........  .AQARATRLF GETVRPGNSR
mdYP2      PNEAVPGADN IVE..... ..........  .SPMRAVRYF AESVVPGNER
mdYP3      PCEGVPGTRN VIE..... ..........  .AQARATRFY AESVRPGNSR
dmYP1      PAAGVPGASN VVE..... ..........  .AAMRATRYF AESVRPGNER
dmYP2      PSTGVPGADN VVE..... ..........  .ATMRATRYF AESVRPGNER
dmYP3      PSTGVPGSEN VIE..... ..........  .AVARATRYF AESVRPGSER
pig.lipp   GKQMPGCQKN ILSQIVDIDG IWEGTRDFVA CNHLRSYKYY ADSILNPD..
rat.liph   GSFQPGCHFL ELYKHIAEHG LNAITQT.IN CAHERSVHLF IDSLQHSNL.

CON        ---------N ---------- ---------- ----R---YF -ESV------

          401                                                        450
mdYP1      NFPAVEASSL IQYRNNDGY. ...GKRTYMG IATHR..... .DISGDYMLE
mdYP2      NFPAESAHSL NEYKNSNTS. ...GRRIYMG IITTF..... .KVEGDYMLQ
mdYP3      NFPALEASSL KQYRNKDSY. ...GKRAYMG IATRR..... .DTTGDYILE
dmYP1      SFPAVPANSL QQYKQNDGF. ...GKRAYMG IDTAH..... .DLEGDYILQ
dmYP2      NFPSVAASSY QEYKQNKGY. ...GKRGYMG IATDF..... .DLQGDYILQ
dmYP3      NFPAVPANSL KQYKEQDGF. ...GKRAYMG LQIDY..... .DLRGDYILE
pig.lipp   GFAGFPCDSY NVFTANKCFP CPSEGCPQMG HYADRFPGKT NGVSQVFYLN
rat.liph   QNTGFQCSNM DSFSQGLCLN CKKGRCNSLG YDIRRI...G HVKSKTLFLI

CON        -F------S- --Y------- ---G----MG ---------- ------Y-L-
```

Figure 3.2.11. Alignment of *Drosophila melanogaster* (dmYP) and *Musca domestica* (MdYP) yolk proteins with pig triacylglycerol lipase and rat hepatic lipase sequences over the region of homology in the lipid binding domain of the lipases. The region with apparent divergence in the amino acid sequence of two of the three *M. domestica* yolk proteins is boxed. The consensus sequence represents amino acid identity in seven of the eight aligned sequences.

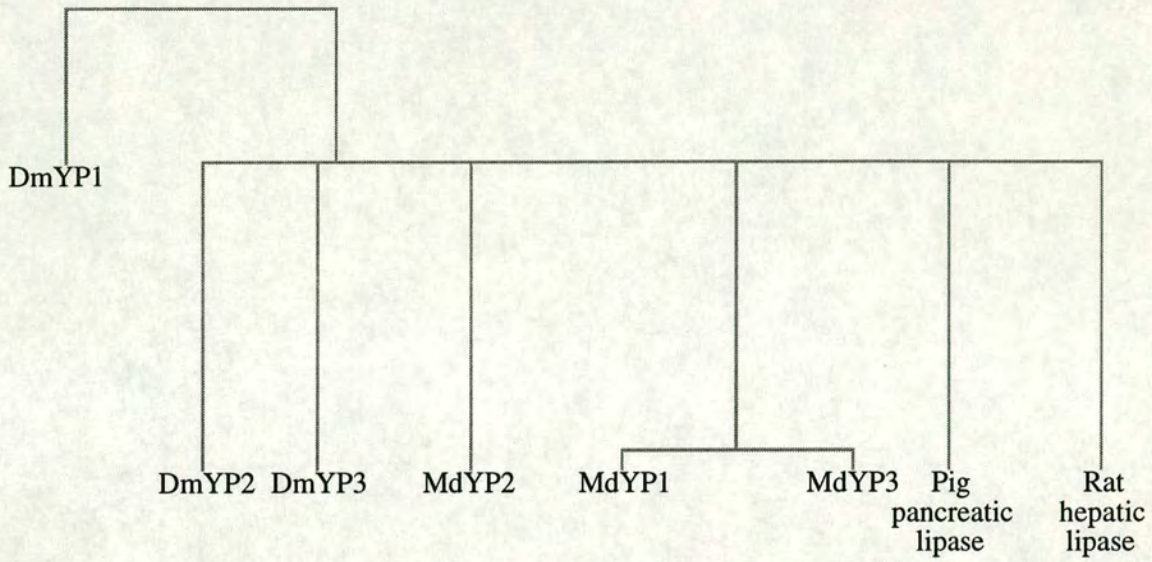


Figure 3.2.12. Dendogram (not to scale) illustrating the relationship between the *Drosophila* and *Musca* yolk proteins and pig pancreatic lipase and rat hepatic lipase over the conserved lipid binding domain.

3.3 SECTION III

The isolation of *M. domestica yolk protein* sequences allows an extensive characterisation of their expression profiles. In particular, it is important to determine whether the endogenous genes are expressed in a manner analogous to *Drosophila*, since this is a pre-requisite for the reporter gene analysis of *yolk protein* promoter elements discussed in results, section II. Although the expression of individual *yolk protein* genes has not been investigated in *M. domestica*, the process of vitellogenesis in this species has been relatively well characterised.

3.3.1 Vitellogenesis in *Musca domestica*

In *M. domestica*, as is observed in *C. erythrocephala* (Martinez and Bownes, 1994), oogenesis proceeds in a cyclic fashion termed interovariole synchrony (the polytrophic ovaries develop follicles in synchronous cycles; Morrison and Davies, 1964; Adams, 1981). Egg maturation is dependent on a proteinaceous diet in most *M. domestica* strains (i.e. they are anautogenic), although continued maintenance of strains on a sugar diet has generated several autogenic strains in which the first gonotrophic cycle is completed even in the absence of a protein meal (Agui *et al.*, 1985). This nutritional response is effected by the egg development neurosecretory hormone (EDNH) pathway by induction of ecdysone synthesis, and is mediated via the corpus-cardiacum (Adams and Nelson, 1990; Adams and Gerst, 1991, 1992, 1993). Interestingly, ecdysteroid titres correlate with haemolymph vitellogenin levels in a linear fashion in the first gonotrophic cycle, and both peak approximately 12-24 hours after emergence (Adams and Gerst, 1993). In contrast to *Drosophila* however where the ovaries are not thought to contribute significantly to free ecdysteroid titres (Bownes, 1989), the ovaries of *M. domestica* are a major site of ecdysteroid synthesis (Adams *et al.*, 1985, 1988) during the vitellogenic stages of oogenesis. In addition, juvenile hormone also stimulates vitellogenin synthesis and is involved in the progression of oogenesis (Adams, 1974; Adams and Filipi, 1988). In females in which the corpus-cardiacum-allatum complex and the ovaries have been surgically removed, no vitellogenin is detected in the haemolymph. However, application of ZR515 (a juvenile hormone analogue) and 20-hydroxyecdysone restores normal haemolymph vitellogenin levels in these females, even though 20-hydroxyecdysone can only induce trace synthesis levels of vitellogenin when applied on its own (for a review of Dipteran juvenile hormone function see Kelly *et al.*, 1987).

In addition to the apparent analogies between *Drosophila* and *M. domestica* hormonal control of vitellogenesis, both the fat body and the ovaries synthesise vitellogenins (Adams and Filipi, 1983; DeBianchi *et al.*, 1985). *In vitro* culture experiments in the presence of [^{35}S] Methionine clearly demonstrate that *M. domestica* ovaries synthesise vitellogenin during those oogenic stages equivalent to stages 8 - 10B of *D. melanogaster* oogenesis. Also, consistent with the cyclic nature of vitellogenesis in

M. domestica, maximal synthesis of vitellogenin in the fat body is co-ordinated with oogenesis, and occurs during vitellogenic stages (i.e. stages 8 - 10B).

Thus, although not previously characterised at a molecular level, the expression of the *M. domestica yolk protein* genes is expected to occur in a manner analogous to that observed in *Drosophila*. The isolation of *M. domestica yolk protein* sequences (results section II) allows an extensive characterisation of the expression of individual *yolk protein* genes, and is discussed below.

3.3.2 Expression of the *M. domestica yolk protein* genes

3.3.2.1 Northern analysis

In order to characterise the sex-, stage-, and tissue-specific transcription profiles of the *M. domestica yolk protein* genes, Northern analysis was carried out. Total RNA was extracted from four ovary pairs, and the corresponding carcasses from these ovary dissections, such that RNA was extracted from both ovaries and carcasses during stages 7 - 14 of oogenesis. In this way a developmental profile of *yp* gene expression in the fat body and ovaries during pre-vitellogenic to post-vitellogenic stages was obtained. Individual tracks of a 1.4% (w/v) formaldehyde denaturing agarose gel were loaded with total RNA equivalent to an extraction from an individual fly (i.e. one quarter of the original extraction), and three replica filters were produced (i.e. identical samples were loaded on each gel, and electrophoresis/ blotting procedures were identical). Each filter was probed with radioactively labelled individual *M. domestica yolk protein* cDNA sequences (i.e. MdcYPA [*Mdyp1*], MdcYPB [*Mdyp2*] and MdcYPE [*Mdyp3*] recombinant plasmids; high stringency hybridisations and post-hybridisation washing procedures were employed), and the results of subsequent autoradiography are depicted in figure 3.3.1. It should be noted, however, that the staging used in this analysis can only be regarded as an approximation since the requirement for RNA extraction precluded a detailed examination of egg chamber morphology. Direct measurements of the isotope emissions from the filters were also recorded using a Molecular Dynamics Phosphoimager and analysed using ImageQuant software. Histograms depicting these measurements are presented in figure 3.3.2.

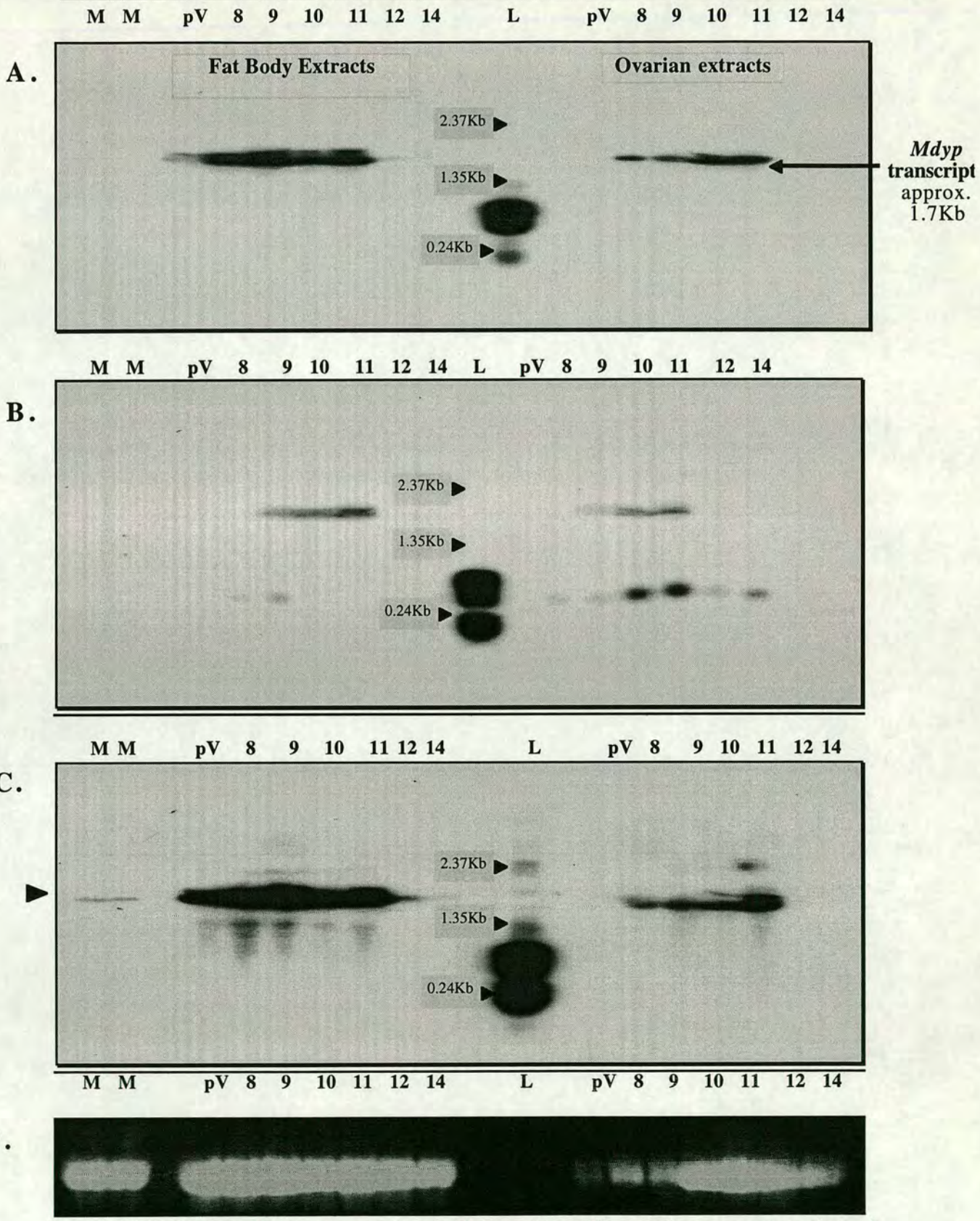
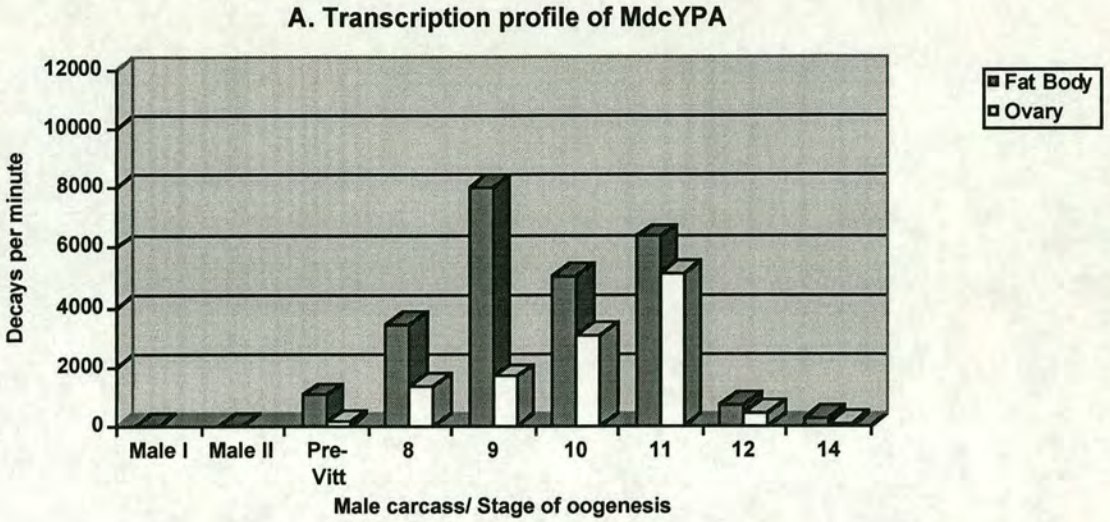
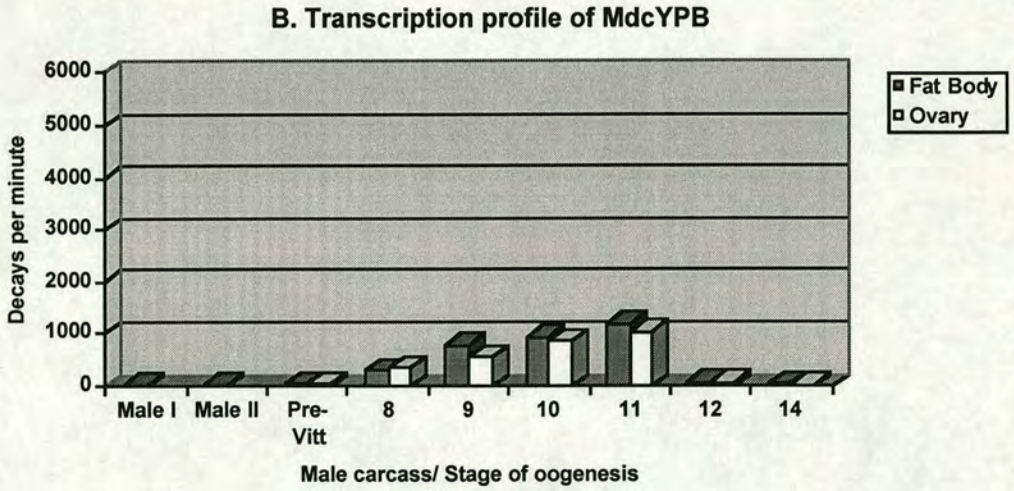


Figure 3.3.1. Replica Northern analysis of *M. domestica yolk protein* gene expression. A. Northern probed with MdcYPA. **B.** Northern probed with MdcYPB. **C.** Northern probed with MdcYPE (note a longer exposure [overnight] of this is presented to highlight the expression in males [arrowhead]). Exposure of the filters presented in **A.** and **B.** was for 4 hours. **D.** Photograph from one of the Ethidium Bromide stained denaturing gels from which the filters presented in A, B, and C derive as a means of estimating loading equivalence. Samples represent extracts from whole adult males (M), pre-vitellogenic (pV) egg chambers, and stages 8-14 of oogenesis. Carcass extracts are staged identically, since the ovarian extracts derived from identical individuals. 'L' represents RNA ladder marker (Gibco BRL)





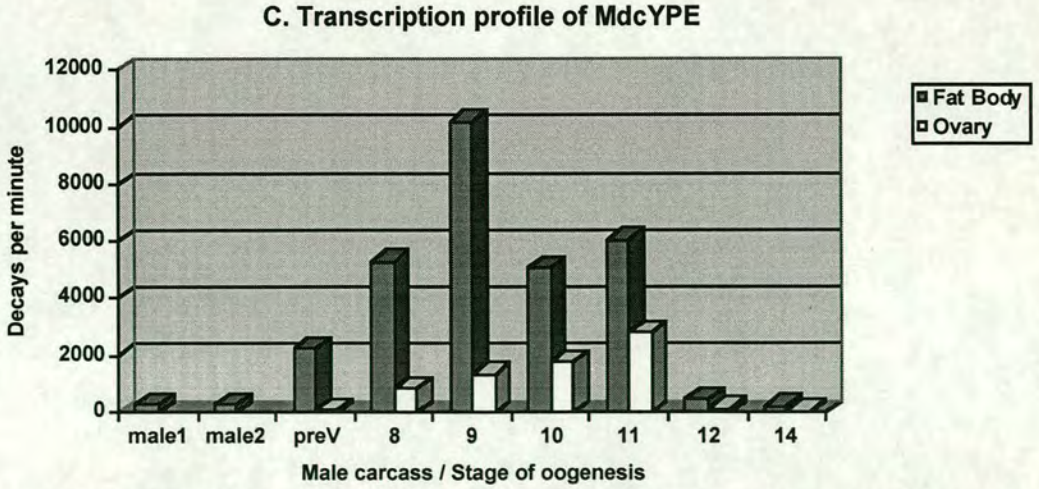


Figure 3.3.2. Histograms depicting direct measurements of isotope decay from the Northern filters presented in figure 3.3.1.

In order to test the equivalence of loading between the individual tracks on the gels, the filters were then stripped and re-probed with cDNA sequences which would be anticipated to be highly conserved between species. Initially the *D. melanogaster* ribosomal protein RP49 was used, and then subsequently a *D. melanogaster* α_1 - tubulin probe was tested (note, *D. melanogaster* probes were used since to our knowledge no suitable probes have been isolated from *M. domestica*). In neither case were any cross-hybridising sequences apparent. This lack of homology does not appear to be the result of loss of target sequences in the stripping procedures since these filters have subsequently been used in other analyses, and show strong cross-hybridisation. It therefore seems most likely that the stringency conditions used in the loading quantitation analysis (i.e. high) were prohibitive, and would require optimising. Time limitations, however, prevented further optimisation of the hybridisation procedure. Therefore, as an indication of the relative equivalence of loading between the various samples, figure 3.3.1(D) shows the Ethidium Bromide stained samples prior to transfer, and clearly shows approximately equal staining in the carcass extractions. The ovary extractions show more evident variation in staining, probably as a result of the increase in mass of the egg chamber during progression of oogenesis (i.e. egg chambers are much smaller during pre-vitellogenic stages than those at a later stage of development).

It is also interesting to note that in addition to the abundant 1.7kb transcript detected in this analysis, we also observe a slower migrating transcript (perhaps 100 - 200bp greater in length) in many of the developmental stages, particularly when probing with either *Mdyp1* or *Mdyp3*. We favour the view that this transcript probably represents unspliced hnRNA derived from the *Mdyp* locus in question since its abundance seems to be directly proportional to the abudance of the main 1.7kb transcript (see figure 3.3.1A, particularly in the fat body extracts), and the *Drosophila yolk* protein genes all contain one or two short introns (see review by Bownes *et al.*, 1993). We can not, however, exclude the possibility that this derives from non-specific cross hybridisation, or that these transcripts in fact represent splice variants. An analysis of the genomic organisation of the three *Musca domestica yolk protein* genes will therefore be necessary to determine this conclusively.

3.3.2.2 Fat body Transcription

Consistent with previously reported data (e.g. the *in vitro* analysis reported by DeBianchi *et al.*, 1985), the abundance of *yp* transcripts present in the fat body (as judged by the transcripts apparent in the carcass extracts) is correlated with the stage of oogenesis, and is cyclic (we presume the fat body is the only site of YP synthesis in the carcass). Interestingly, transcripts cross hybridising with probes MdcYPA [*Mdyp1*] and MdcYPE [*Mdyp3*] are more abundant than those cross hybridising with MdcYPB [*Mdyp2*]. Both *Mdyp1* and *Mdyp3* are extensively transcribed during the vitellogenic stages of oogenesis (stages 8 - 11 in this analysis), and transcript levels are much reduced after these stages although still weakly apparent at stage 14. Surprisingly, both *Mdyp1* and *Mdyp3*

transcripts are also abundant during pre-vitellogenic stages. DeBianchi *et al.* (1985) report that vitellogenin is present in the haemolymph of female flies containing pre-vitellogenic eggs, whilst Adams and Filipi (1983) detected no vitellogenin in the haemolymph prior to stage 7. Our results are consistent with those of DeBianchi *et al.* (1985) who demonstrated using *in vitro* culture experiments in the presence of [^{3}H] Leucine that the fat body synthesises vitellogenin intensely prior to and during the vitellogenic stages of oogenesis, and decreases subsequent to this. It should be noted however that the fly stocks used in our experiments have been maintained on sugar, and appeared to be autogenic (i.e. the first gonotrophic cycle proceeded even in the absence of a proteinaceous meal). Such differences in transcription may merely reflect variation in the dependence on a protein meal between the various fly strains. An analysis of transcription in both anautogenic and autogenic strains would therefore be of interest.

In contrast to these results, *Mdyp2* transcripts are not detected prior to stage 8, and accumulate at a generally later stage than transcripts derived from *Mdyp1* and *Mdyp3*. Maximal *Mdyp2* fat body transcript abundance is detected during stages 9 - 11 in this analysis, and no transcripts are detected post stage 11. It is interesting to note that a faster migrating transcript (of approximately 500-600bp) is also present, perhaps reflecting a degree of non-specific cross-hybridisation. However, the high degree of homology between the *Mdyp's* suggests this may not be the case, since if this is indeed non-specific hybridisation, it seems reasonable to assume it would be evident in hybridisations with *Mdyp1* or *Mdyp3*. It is possible, however, that this faster migrating transcript is a degradation product of the full length *Mdyp2* transcript, since the reduced abundance of *Mdyp2* transcripts relative to either *Mdyp1* or *Mdyp3* may be a consequence of reduced transcript stability, as is known to be the case with *Dmyp3* transcripts (Williams and Bownes, 1986).

One particularly intriguing result apparent from the Northern analysis is that *Mdyp3* transcripts are detected in male whole body total RNA extracts. This result is not due to experimental error since replica filters give the same result, as do filters containing male total RNA samples from independent isolations. Although these transcripts could derive from cross-hybridisation of the *Mdyp3* cDNA probe with closely related sequences, this seems unlikely for two reasons. Firstly, the transcripts size (1.7kb) detected in males is identical to that observed in females, and secondly, each cDNA probe reveals independent expression profiles (particularly between *Mdyp1* or *Mdyp3* and *Mdyp2*) suggesting each probe has high specificity for its target sequence (i.e. each cDNA probe does not seem to cross-react with transcripts derived from other *yp* genes). If, for example, a degree of cross-hybridisation is occurring between probe *Mdyp1* and transcripts derived from *Mdyp3* (and vice-versa), why is there no transcript apparent in male carcass total RNA when probed with *Mdyp1*? The generation of mutants deficient of *Mdyp3* transcription could resolve this issue.

It should be noted however that the *Mdyp3* cross-reacting transcript apparent in males is present at a greatly reduced level relative to those apparent in females during the vitellogenic stages of oogenesis. Adams and Filipi (1983) report that two polypeptides of a similar size to female vitellogenin subunits are detected in the male haemolymph, but that they do not cross-react with an *M. domestica* anti-vitellin polyclonal antibody. However, A. Dübendorfer (pers. comm.) has observed vitellogenins in male haemolymph extracts from a variety of *M. domestica* strains. Clearly, further analysis is required to establish why vitellogenins are present in the male haemolymph of some *M. domestica* strains.

3.3.2.3 Ovarian Transcription

The ovarian-specific expression profiles of *Mdyp1*, *Mdyp2* and *Mdyp3* are also illustrated in figure 3.3.1, and clearly demonstrate that transcription is restricted to vitellogenic stages of oogenesis. In all cases, low transcript abundance is detected in ovaries containing pre-vitellogenic (prior to stage 7) and post-vitellogenic (subsequent to stage 11) egg chambers. (note : the staging method employed in this analysis can only be considered an approximation since the requirement for rapid RNA extraction precluded an extensive analysis of egg chamber morphology). Thus, ovarian expression is restricted to stages 8 - 11 of oogenesis as judged by this crude assay. In general, transcripts are detected at a lower abundance in the ovaries relative to those detected in female carcass extracts. Also, consistent with the expression profiles apparent in the female carcass, transcripts derived from *Mdyp2* are less abundant in ovaries than transcripts derived from either *Mdyp1* or *Mdyp3*. Maximal transcript abundance in all cases is detected in ovaries containing stage 10 or stage 11 egg chambers. These results are entirely consistent with the *in vitro* culture results previously reported (Adams and Filipi, 1983; DeBianchi *et al.*, 1985).

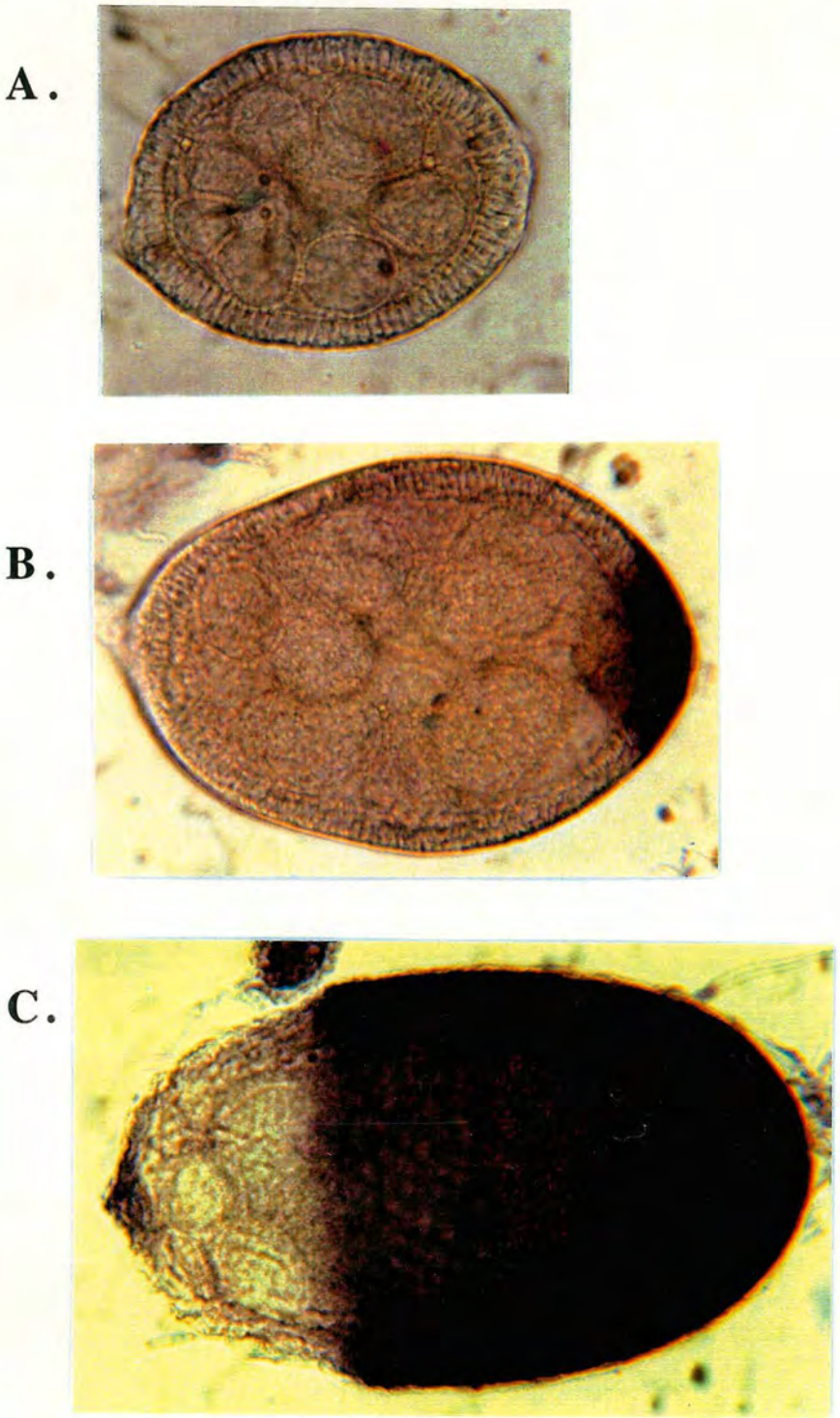


Figure 3.3.3. Whole mount *M. domestica* ovarian *in-situ* hybridisations to RNA. **A.** Pre-vitellogenic egg chamber (stage 5). **B.** Early stage 8 egg chamber. **C.** Stage 9 egg chamber.

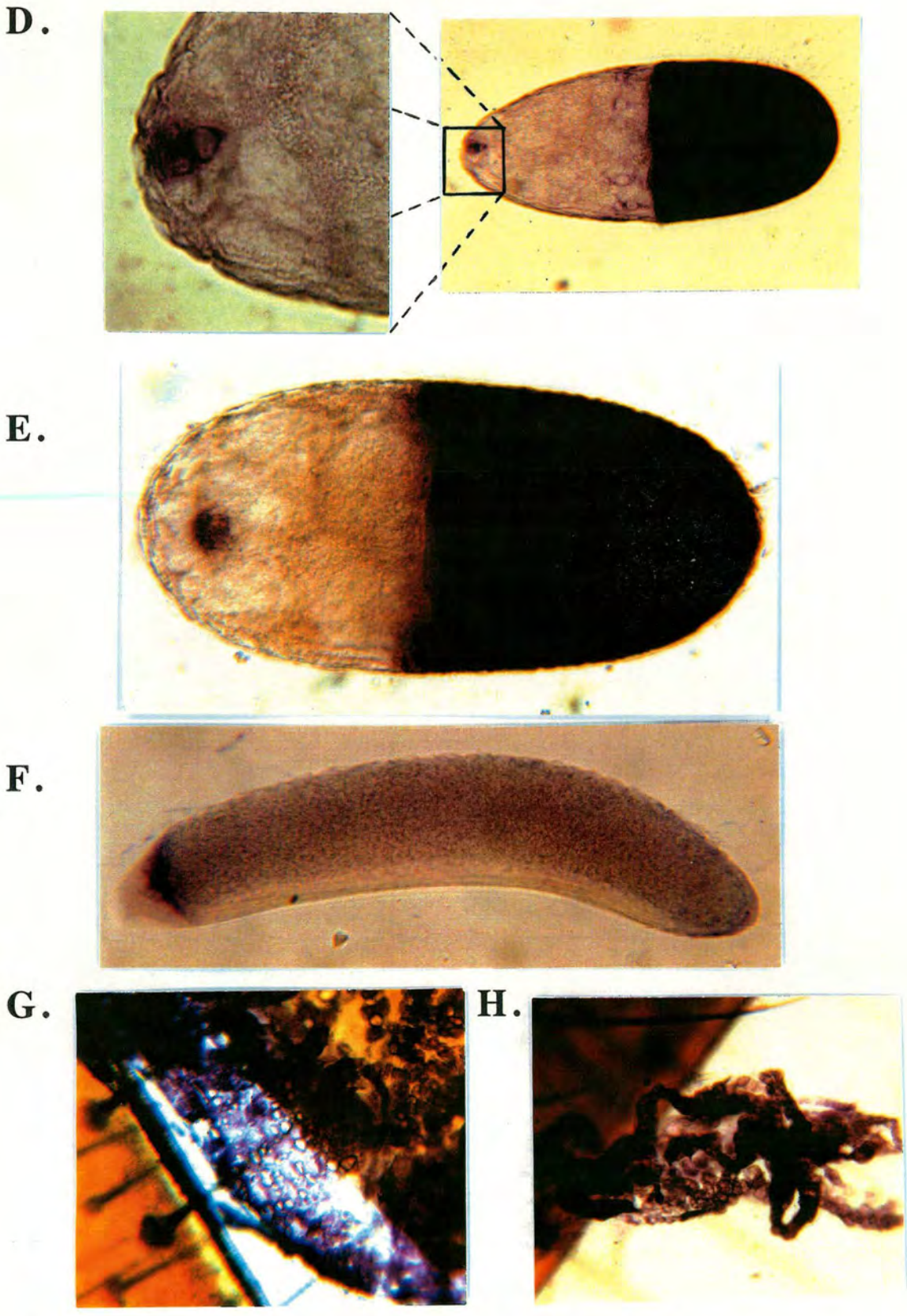


Figure 3.3.3.(continued) Whole mount *M. domestica* ovarian *in-situ* hybridisations to RNA. D. Staining in the presumptive borders cells prior to migration **E.** Stage 10 egg chamber also showing border cell staining during migration **F.** Post-vitellogenic egg chamber (stage 14). **G.** Female fat body tissue. **H.** Male fat body tissue

3.3.3 *IN SITU* HYBRIDISATION ANALYSIS

Since the Northern analysis described has poor resolution with respect to the temporal and spatial expression of the *M. domestica yolk protein* genes, whole mount tissue *in situ* hybridisations to RNA were carried out to determine these factors more precisely. In all cases, ovaries were dissected into PBS solution, and the corresponding carcass remains were further dissected to remove the gut rudiments prior to hybridisation (since fat body-specific transcription was anticipated, and the fat body was presumed to be the major site of carcass yolk protein synthesis). Similarly, male carcasses were prepared by removing the gut rudiments and gonads.

Each dissected tissue was probed with individual *M. domestica yolk protein* cDNA clones (the whole insert and vector was used [note : vector controls were performed, see subsequent discussion]). In all cases, the expression profiles revealed by each cDNA probe (i.e. MdcYPA, MdcYPB or MdcYPE) were identical. Therefore, the subtle quantitative differences in tissue-specific expression of individual *M. domestica yp* genes apparent in the Northern analysis could not be resolved in the *in situ* analysis. This is most likely due to differences in hybridisation stringency in the Northern and *in situ* procedures, and may reflect a degree of cross-reactivity between *yolk protein* transcripts and the cDNA probes in the *in situ* analysis.

3.3.3.1 Fat body Transcription

The Northern analysis described revealed sex-specific carcass expression of two of the *M. domestica yolk protein* genes (*Mdyp1* and *Mdyp2*), and greatly reduced expression of *Mdyp3* in the male carcass relative to the female carcass. Despite this, *in situ* hybridisations with any of the three *M. domestica yp* cDNA probes revealed identical expression profiles in male and female fat body tissue (see figure 3.3.3[g] and figure 3.3.3[h]), with strong hybridisation being apparent in all cases. However, an identical result was also obtained with a vector probe lacking an insert (i.e. a negative control), suggesting the expression profiles revealed in this preliminary analysis were a consequence of non-specific hybridisation and did not reflect the true expression profiles of the endogenous *M. domestica yp* genes. The most simplistic explanation of this result is that either the pre-absorption of the secondary antibody was insufficient, or that the stringency of the hybridisation was too low. Neither of these interpretations, however, seems to be correct. Extensive pre-absorption of the secondary antibody did not result in a commensurate decrease in non-specific signal, nor did an increase in the stringency of the hybridisation (data not shown; incremental temperatures from 45°C to 50°C were tested; temperatures greater than 50°C were not tested since degeneration of the tissue was apparent). An alternative hypothesis, based upon the glutinous nature of fat body tissue (which is known to interfere in other experimental procedures in *D. melanogaster*, M.

Bownes, pers. comm.), is that either the probe or the secondary antibody is adhering to the tissue non-specifically (in this case pre-absorption of the secondary antibody have no effect since the tissue itself would require blocking, and would not reflect the presence of cross-reacting antigens). An increase in the concentration of non-specific competitor salmon sperm DNA (concentrations of 100 µg/ml, 200 µg/ml, and 400 µg/ml were tested) did not abolish the non-specific signal, but did result in a proportional reduction in signal intensity in male fat body tissue relative to females (data not shown; i.e. the signal intensity appeared to be greater in female, as opposed to male, fat body tissue, and that this reduction in signal intensity appeared to be proportional to the increase in salmon sperm DNA concentration). Although time limitations prevented further optimisation of this procedure, it seems likely that the *in situ* analysis in fat body tissue will be resolved. Perhaps the addition of a non-specific protein competitor (such as Bovine Serum Albumin) would be more effective at blocking the tissue prior to and during hybridisation, although this was not tested.

Therefore, no conclusive data could be obtained from the *in situ* analysis of *yp* expression in fat body tissue. However, the Northern analysis previously described clearly demonstrates extensive transcription of the *M. domestica yp* genes in the females carcass, and based on analogy with other insect systems it seems likely that transcription is restricted to fat body tissue. Optimisation of the fat body *in situ* analysis would certainly demonstrate this more conclusively, and may also resolve whether expression of *Mdyp3* in the male carcass is indeed occurring as would appear to be the case from the Northern hybridisations.

3.3.3.2 Ovarian Transcription

The ovarian expression of the *M domestica yolk protein* genes, as revealed by *in situ* hybridisation are presented in figures 3.3.3[a-f] (note : only one developmental expression profile is presented since all were essentially identical). Consistent with the ovarian expression of the *D. melanogaster* and *C. erythrocephala yolk protein* genes (see review by Bownes *et al.*, 1993; Martinez and Bownes, 1994), the *M. domestica yp* genes are expressed in the ovarian follicle cells during the vitellogenic stages of oogenesis (comparable to stages 8 - 10B in *D. melanogaster*). Expression initiates during early stage 8 in those follicle cells adjacent to the presumptive oocyte (i.e. at the posterior of the egg chamber; see figure 3.3.3[b]). This expression domain then expands to encompass those follicle cells migrating posteriorly during late stage 8 and stage 9, such that during the early stages of follicle cell migration only a small anterior domain of the egg chamber remains unstained (see figure 3.3.3[c]). Interestingly, in contrast to the expression of the *D. melanogaster yp* genes (Logan and Wensink, 1990), *M. domestica yp* transcripts are detected in the presumptive border cells during late stage 9 (at the anterior tip) and during their subsequent migration through the nurse cell region (see figures 3.3.3[d] and 3.3.3[e]). Whether this border cell expression is maintained once these cells reach the

anterior tip of the developing oocyte could not be resolved due to the overlying expression in the columnar follicle cells. It seems likely, however, that expression in the border cells will be maintained until *yp* transcription is extinguished in the overlying follicle cell population.

Expression of the *M. domestica yp* genes is maintained in the columnar follicle cells overlying the developing oocyte until approximately stage 10 (figure 3.3.3[e]). Staining is also apparent in a small region adjacent to the anterior tip of the elongating oocyte during stage 13 and stage 14 (see figure 3.3.3[f]). However, this signal has been observed in many *in situ* analyses (M. Bownes, pers. comm.), and probably reflects non-specific hybridisation due to degeneration of the nurse cells (indeed in support of this hypothesis, signal was also apparent in negative control samples [data not shown]).

Therefore, consistent with the previously reported expression profiles of other Dipteran *yolk protein* genes, the *M. domestica yp* genes are expressed during stages 8 - 10 of oogenesis in the ovarian follicle cells.

3.3.4 SECTION III : DISCUSSION

The Northern hybridisation and *in situ* analysis of *M. domestica yolk protein* gene expression described lends further credence to the hypothesis that Dipteran *yolk protein* expression is conserved throughout Dipteran evolution (Martinez, 1991; Martinez and Bownes, 1992). Expression is restricted to the vitellogenic stages of oogenesis (i.e. stages 8 - 10), and the cyclic nature of *M. domestica* oogenesis is mirrored by cyclic transcription of *yolk protein* genes in the fat body. Ovarian transcription is restricted to the ovarian columnar follicle cells and the border cells, and is not detected in the squamous epithelial cells overlying the nurse cell region.

Expression in the border cells was somewhat unanticipated since in *D. melanogaster* a repressor of border cell expression has been identified in the ovarian enhancer-2 (OE-2) element located in the intergenic spacer between *yp1* and *yp2* (Logan and Wensink, 1990). It should also be noted that expression of the *M. domestica yolk protein* genes in the border cells has been confirmed independently by germ line transformation experiments (Tortiglione and Bownes, submitted). Reporter gene constructs under the control of *M. domestica yp* promoter/ enhancer elements show border cells expression in host *D. melanogaster* ovaries. This result, therefore, also substantiates the results of Logan and Wensink (1990), since expression of the endogenous *D. melanogaster yolk protein* genes in the border cells can clearly be resolved. Analogous results are also apparent in reporter gene constructs under the control of *C. erythrocephala yp* promoter/ enhancer elements (i.e. expression is observed in the host *D. melanogaster* ovaries; Tortiglione and Bownes, submitted). We favour the view that repression of border cell expression has been acquired in the more evolutionary advanced *D. melanogaster* species, rather than lost in the more primitive *M. domestica* and

C. erythrocephala species. In either case, the acquisition or loss of border cell *yp* repression must reflect a degree of selective pressure, and it is therefore intriguing why this selective pressure only exerts an influence over one of the three species analysed. The fact that border cell *yp* expression is apparent in *M. domestica* egg chambers would suggest that it is not detrimental to the process of oogenesis, nor to the correct activity of the nurse cells. With this in mind, it would be interesting to use germ line transformation experiments to induce native *yp* expression *D. melanogaster* border cells (using a *D. melanogaster yp* cDNA under the control of *M. domestica yp* regulatory elements, or under the control of an OE-2 construct lacking the domain required for border cell repression). Any detrimental effects associated with border cell *yp* expression would then become apparent.

4. FINAL DISCUSSION

Despite the use of a wide variety of approaches, the original intention of the work reported here (i.e. the isolation of sex determination genes from *Calliphora erythrocephala* and *Musca domestica*) proved impractical at this time. The results presented in Section I (encompassing both antibody and nucleotide cross-hybridisation analyses, as well as PCR based approaches) at best suggest an analysis of species more closely related to *D. melanogaster* is a pre-requisite to further analysis in *C. erythrocephala* and *M. domestica*, and at worst suggest a homologue of the *D. melanogaster dsx* sex-determination gene is unlikely to be present in these two species.

An analysis of sex determination gene homologues in species closely related to *D. melanogaster* is also of great interest since the relative rates of gene divergence could be estimated, and substantiate the results presented here (i.e. evidence of rapid gene divergence would suggest distantly related species may have acquired alternative sex determination mechanisms). In any case, comparisons of such sequences could identify regions within the encoded proteins under strong selective pressure, which in turn would facilitate the design of redundant oligonucleotides for cross-species PCR based applications. It is interesting to note, however, that the gene sequence of a newly isolated *M. domestica Sxl* homologue is highly conserved relative to the *D. melanogaster* gene, but appears to be non-functional with respect to sex determination (D. Bopp, pers. comm.).

Since the molecular analysis of sex determination genes proved unsuccessful, we have attempted to substantiate the results presented in section I indirectly using an analysis of *yolk protein* gene expression in *C. erythrocephala* and *M. domestica*. Martinez and Bownes (1994) have clearly demonstrated sex-specific expression of two *C. erythrocephala yolk protein* genes. Indeed, the expression profiles of these *C. erythrocephala* genes is essentially identical to the expression profile of the *D. melanogaster yolk protein* genes, suggesting regulation of their expression is effected by similar factors.

The isolation and characterisation of the *M. domestica yolk protein* genes reported here was therefore not only of evolutionary interest, but also a pre-requisite for the reporter gene analysis described. The results presented support previous findings of dipteran *yolk protein* conservation, revealing a high degree of conservation not only at a molecular level, but also with respect to their expression profiles. The observation that weak expression of the *M. domestica yolk protein*-3 (*Mdyp3*) gene is detected in whole male extracts (presumably reflecting expression in the fat body) clearly needs to be investigated. Although conclusive determination that this is indeed the case would require a mutant deficient of *Mdyp3* expression, perhaps a more immediate approach would be an analysis of reporter gene expression using a construct under the control of *Mdyp3* promoter elements. However, reporter gene constructs under the control of either *C. erythrocephala* or *M. domestica yp* promoter elements are not expressed sex-

specifically in transgenic *D. melanogaster* adults (Tortiglione and Bownes, submitted). Thus, the reporter gene analysis (of weak expression in males) may only be productive in transgenic *M. domestica* adults, where expression would be expected to mimic the endogenous *Mdyp3* gene. At present, however, *M. domestica* germline transformation procedures have not been described, although research in this area is proceeding (R. Nöthiger, pers. comm.).

Interestingly, since reporter gene constructs under the control of *M. domestica* and *C. erythrocephala yp* promoter elements do generate stage- and tissue-specific expression of the reporter gene in transgenic *D. melanogaster* adults, it can be inferred that those elements which define this pattern of expression are conserved between these species (Tortiglione and Bownes, submitted). Conservation of those elements defining the stage- and tissue-specific expression of the *yp* genes suggests strong selective pressure to maintain the expression profile of these genes, and it is therefore particularly intriguing that those elements defining the sex-specific expression of the *yp* genes appear, from these results, to have diverged. Clearly, sex-specific expression of the *yolk protein* genes is maintained in the species analysed (Martinez and Bownes, 1994; this thesis, results section III), albeit by what at present appear to constitute either unrelated or highly diverged regulatory proteins/ elements. Further analysis of the regulation of *yp* expression in these, and indeed other species, is therefore of great interest since within this region there appears to be high conservation of some, and high divergence of other, regulatory mechanisms whilst maintaining the overall pattern of expression. Indeed, it will be interesting to compare the sequence of promoter/ enhancer elements from a variety of dipteran species in an attempt to identify conserved regulatory domains (presumably the regulatory elements responsive to sex-specific induction will not be conserved between these species).

Collation of the data presented in this thesis, and those of Tortiglione and Bownes (submitted) suggests that the sex-specific expression of the *yolk protein* genes is not regulated in a similar manner in the dipteran species analyses. Conclusive determination that genes involved in regulating sexual differentiation in *Calliphora* and *Musca* are not homologues of *Drosophila* sex-determination genes, however, will require the molecular characterisation of sex determination genes from these two species (e.g. the 'M' and 'F' factors in *M. domestica*, see introduction). Comparisons of gene sequence, and mode of action, should then identify any functional similarities between various dipteran sex determination genes.

Consistent with previous data from a wide variety of dipteran species, the *yolk protein* genes are highly conserved throughout dipteran evolution, suggesting a more defined role in development than as a nutritional storage molecule. The observation that yolk spheres isolated from *D. melanogaster* embryos are associated with inactivated ecdysteroid conjugates, coupled with the high degree of evolutionary conservation,

suggests the primary role of the yolk proteins may be to control the release of ecdysone during the early stages of embryogenesis. Interestingly, this suggests an almost symbiotic relationship between the release of ecdysone and the provision of nutrients to the developing embryo. The release of nutrients from the yolk spheres must be tightly co-ordinated with embryogenesis to ensure correct development (i.e. stored nutrients are not exhausted before the embryo becomes self supportive), as must the release of ecdysone.

Of the two proposed functions of the yolk proteins, however, the controlled release of ecdysone seems a more logical primary role for these proteins, since any polypeptide with appropriate localisation signals could supply the oocyte with nutrients required for later development (and as such the regions outwith those residues defining the correct localisation of the protein would not be expected to be highly conserved). In addition, it seems somewhat unlikely that this differential expression merely reflects an inability on the part of the ovarian follicle cells to synthesise sufficient protein. Expression of the *yolk protein* genes in the fat body clearly leads to complications, since it necessitates a mechanism for sex-specific induction, and presumably introduces further localisation requirements. Perhaps synthesis of yolk proteins in the female fat body reflects a functional requirement of these proteins, such as the sequestration of ecdysteroid conjugates from the circulating haemolymph prior to deposition in the developing oocyte.

In conclusion, therefore, the dipteran *yolk protein* genes present us with many intriguing questions which should be resolved. The regulation of their expression in the various diptera is clearly of great interest, since there is evidence of divergence amongst some of the regulatory elements, whilst maintaining the overall expression profile. Does expression in the fat body reflect a requirement for passage through the haemolymph prior to deposition? What functions do the conserved domains apparent between the various dipteran yolk proteins serve (i.e. are they structural constraints, binding motifs, localisation signals)? Some of these question can be resolved using transgenic approaches (either using reporter gene analyses to dissect regulatory elements, or by introducing point mutations or deletions in conserved domains and analysing the effects *in vivo*), others may best be approached in larger organisms (such as *C. erythrocephala* or *M. domestica*) where more accurate assays of ecdysone concentrations in the circulating haemolymph can be determined.

5. APPENDIX I

NOTE : underlined sequences in the 3' termini of each sequence conform to the AAUAAA consensus polyadenylation sequence (reviewed by Proudfoot, 1991)

Figure A.1 : *Musca domestica yolk protein-1 (Mdyp1)* cDNA

(derived from cDNA clone MdYPA)

```
    GAATTCGGGACGAGCCCGTACGACTTTGGGAATTTGTGAAACTGACCGACAGAAGGATGA
1   ---------+---------+---------+---------+---------+---------+   60
    CTTAAGCCCTGCTCGGGCATGCTGAAACCCTTAAACACTTTGACTGGCTGTCTTCCTACT

                                                          M  N

    ATCCATTGGGAGTTGTGTGCTTTGTAGCCTTTGTGGCTGTCGGCGCCTTGGTCTCACAAT
61  ---------+---------+---------+---------+---------+---------+  120
    TAGGTAACCCTCAACACACGAAACATCGGAAACACCGACAGCCGCGGAACCAGAGTGTTA

      P  L  G  V  V  C  F  V  A  F  V  A  V  G  A  L  V  S  Q  S

    CGGAGGACTACAGTCCAAAGCCAGCGTATTGGGTGAAACCCACCGAATTGGGGGACACAC
121 ---------+---------+---------+---------+---------+---------+  180
    GCCTCCTGATGTCAGGTTTCGGTCGCATAACCCACTTTGGGTGGCTTAACCCCCTGTGTG

      E  D  Y  S  P  K  P  A  Y  W  V  K  P  T  E  L  G  D  T  P

    CATCTGTGAATGAGCTCACCTGTGAGGAGCTGGAGAATATGCCATTGGAAAAGGGAGACA
181 ---------+---------+---------+---------+---------+---------+  240
    GTAGACACTTACTCGAGTGGACACTCCTCGACCTCTTATACGGTAACCTTTTCCCTCTGT

      S  V  N  E  L  T  C  E  E  L  E  N  M  P  L  E  K  G  D  T

    CATTGATGTGTAAACTCTACCACCTGTCCCAAATCGACTACTCTGTGTCACCCAACTTTT
241 ---------+---------+---------+---------+---------+---------+  300
    GTAACTACACATTTGAGATGGTGGACAGGGTTTAGCTGATGAGACACAGTGGGTTGAAAA

      L  M  C  K  L  Y  H  L  S  Q  I  D  Y  S  V  S  P  N  F  C

    GCCCAAGCCCCACCAATGTCCCCGTCCACAGTTTCAACAACAAGGGTGAAAAGGAAACCT
301 ---------+---------+---------+---------+---------+---------+  360
    CGGGTTCGGGGTGGTTACAGGGGCAGGTGTCAAAGTTGTTGTTCCCACTTTTCCTTTGGA

      P  S  P  T  N  V  P  V  H  S  F  N  N  K  G  E  K  E  T  S

    CCAACTTGAACAAACATTCGACATTGCCCGAGGAAAAGCCGAAATTTGACGAACAAGAAG
361 ---------+---------+---------+---------+---------+---------+  420
    GGTTGAACTTGTTTGTAAGCTGTAACGGGCTCCTTTTCGGCTTTAAACTGCTTGTTCTTC

      N  L  N  K  H  S  T  L  P  E  E  K  P  K  F  D  E  Q  E  V

    TCACCGTTTTCATCACCGGTCTTCCCCAGTCCTTGGAGGATGTTAAGACGCCCAACACCA
421 ---------+---------+---------+---------+---------+---------+  480
    AGTGGCAAAAGTAGTGGCCAGAAGGGGTCAGGAACCTCCTACAATTCTGCGGGTTGTGGT

      T  V  F  I  T  G  L  P  Q  S  L  E  D  V  K  T  P  N  T  K
```

```
      AATTGATCCAGTCCTACATTCAACGTTACACCAAAAAACCCGAAGCACCCCAGGGTGAGG
481   ---------+---------+---------+---------+---------+---------+   540
      TTAACTAGGTCAGGATGTAAGTTGCAATGTGGTTTTTTGGGCTTCGTGGGGTCCCACTCC

        L  I  Q  S  Y  I  Q  R  Y  T  K  K  P  E  A  P  Q  G  E  D

      ATCAATCGAAATGGGAAAATGAAAAACCTGTGGGCGGTCATTTGGTTGTCATCGATTTGG
541   ---------+---------+---------+---------+---------+---------+   600
      TAGTTAGCTTTACCCTTTTACTTTTTGGACACCCGCCAGTAAACCAACAGTAGCTAAACC

        Q  S  K  W  E  N  E  K  P  V  G  G  H  L  V  V  I  D  L  G

      GCCATGCCATCACCAATGTTGAACGTTATGCCACTTTGAATGTCAAGGAGACCGGTAAAA
601   ---------+---------+---------+---------+---------+---------+   660
      CGGTACGGTAGTGGTTACAACTTGCAATACGGTGAAACTTACAGTTCCTCTGGCCATTTT

        H  A  I  T  N  V  E  R  Y  A  T  L  N  V  K  E  T  G  K  M

      TGATTGGCAAGACCTTGGCTGAGTTGGAGAAGGAGAGCAATGTCGATTTGGAAGATCTCC
661   ---------+---------+---------+---------+---------+---------+   720
      ACTAACCGTTCTGGAACCGACTCAACCTCTTCCTCTCGTTACAGCTAAACCTTCTAGAGG

        I  G  K  T  L  A  E  L  E  K  E  S  N  V  D  L  E  D  L  H

      ATGTCATTGGTCAGGGTATTGGTGCCAATGTTGCTGGTGCTGCTGGTAAGGCTTTCAAGG
721   ---------+---------+---------+---------+---------+---------+   780
      TACAGTAACCAGTCCCATAACCACGGTTACAACGACCACGACGACCATTCCGAAAGTTCC

        V  I  G  Q  G  I  G  A  N  V  A  G  A  A  G  K  A  F  K  D

      ACGTTACCACACACAAATTGGGTCGCATCACTGTCTTGGACCCTGCCAGACAGGTGGGCA
781   ---------+---------+---------+---------+---------+---------+   840
      TGCAATGGTGTGTGTTTAACCCAGCGTAGTGACAGAACCTGGGACGGTCTGTCCACCCGT

        V  T  T  H  K  L  G  R  I  T  V  L  D  P  A  R  Q  V  G  K

      AGGATCCCAAAGTTTTGACCGGTTTGTCTCGTTGCAGTGTCAAATTCGTTGATGTCATCC
841   ---------+---------+---------+---------+---------+---------+   900
      TCCTAGGGTTTCAAAACTGGCCAAACAGAGCAACGTCACAGTTTAAGCAACTACAGTAGG

        D  P  K  V  L  T  G  L  S  R  C  S  V  K  F  V  D  V  I  H

      ACACCTCCGCCTTGGGTATGGGTACCACCCGTCGTGTCGGTGATGTTGATTTCTTCTCCA
901   ---------+---------+---------+---------+---------+---------+   960
      TGTGGAGGCGGAACCCATACCCATGGTGGGCAGCACAGCCACTACAACTAAAGAAGAGGT

        T  S  A  L  G  M  G  T  T  R  R  V  G  D  V  D  F  F  S  N

      ATGGTACATGTCAAGGTGTTCCCGGTAGTCGCAATGCCATCGATGCCCAAGCTCGTGCCA
961   ---------+---------+---------+---------+---------+---------+   1020
      TACCATGTACAGTTCCACAAGGGCCATCAGCGTTACGGTAGCTACGGGTTCGAGCACGGT

        G  T  C  Q  G  V  P  G  S  R  N  A  I  D  A  Q  A  R  A  T

      CACGTCTCTTTGGCGAAACCGTTCGTCCCGGAAACAGCCGCAACTTCCCTGCCGTTGAGG
1021  ---------+---------+---------+---------+---------+---------+   1080
      GTGCAGAGAAACCGCTTTGGCAAGCAGGGCCTTTGTCGGCGTTGAAGGGACGGCAACTCC

        R  L  F  G  E  T  V  R  P  G  N  S  R  N  F  P  A  V  E  A
```

```
      CCAGTTCTTTGATACAGTACCGCAACAATGATGGCTATGGCAAACGTACCTATATGGGTA
1081  ---------+---------+---------+---------+---------+---------+  1140
      GGTCAAGAAACTATGTCATGGCGTTGTTACTACCGATACCGTTTGCATGGATATACCCAT

        S   S   L   I   Q   Y   R   N   N   D   G   Y   G   K   R   T   Y   M   G   I

      TTGCCACCCACCGTGATATCTCCGGTGACTACATGTTGGAGGTTAACGCCGAGAGTCCCT
1141  ---------+---------+---------+---------+---------+---------+  1200
      AACGGTGGGTGGCACTATAGAGGCCACTGATGTACAACCTCCAATTGCGGCTCTCAGGGA

        A   T   H   R   D   I   S   G   D   Y   M   L   E   V   N   A   E   S   P   Y

      ATGGCAAGAGAACCCCCGCCCGCAAACAAAAATCATACCACGGTTTTCATCAAACTTCCT
1201  ---------+---------+---------+---------+---------+---------+  1260
      TACCGTTCTCTTGGGGGCGGGCGTTTGTTTTTAGTATGGTGCCAAAAGTAGTTTGAAGGA

        G   K   R   T   P   A   R   K   Q   K   S   Y   H   G   F   H   Q   T   S   Y

      ATGCCAAAAGCAATGAAAACTATTAGAAAGTTGATTGTCTTGGCAAAAGAAAAAGATTCT
1261  ---------+---------+---------+---------+---------+---------+  1320
      TACGGTTTTCGTTACTTTTGATAATCTTTCAACTAACAGAACCGTTTTCTTTTTCTAAGA

        A   K   S   N   E   N   Y   *

      CTGGAAATGACTCCCAACAACTTAAGCTGTGAATATTTGTCGAACAAAAACAATATTTAT
1321  ---------+---------+---------+---------+---------+---------+  1380
      GACCTTTACTGAGGGTTGTTGAATTCGACACTTATAAACAGCTTGTTTTTGTTATAAATA

      TTTCACCCAAGCCGAGAGTCCCTATGGCAAGAGAACCCCCGCCCGCCTTCCAAAATCATA
1381  ---------+---------+---------+---------+---------+---------+  1440
      AAAGTGGGTTCGGCTCTCAGGGATACCGTTCTCTTGGGGGCGGGCGGAAGGTTTTAGTAT

      CCACGGTTTTCATCAAACTTCCTATTCCCCAAGCAATGGAAACTATTAGGAAGTTGGATG
1441  ---------+---------+---------+---------+---------+---------+  1500
      GGTGCCAAAAGTAGTTTGAAGGATAAGGGGTTCGTTACCTTTGATAATCCTTCAACCTAC

      TTTGCCCCAAGAAAAAGATTCTTTGGAAATGACTCCGAAAAAATTAAGCTGTGAATATTT
1501  ---------+---------+---------+---------+---------+---------+  1560
      AAACGGGGTTCTTTTTCTAAGAAACCTTTACTGAGGCTTTTTTAATTCGACACTTATAAA

      GTCGGCCCCAAACAATAATTAAAATCCCAATTTGTGTAATTTTTTAGAGATGAAAAAAAA
1561  ---------+---------+---------+---------+---------+---------+  1620
      CAGCCGGGGTTTGTTATTAATTTTAGGGTTAAACACATTAAAAAATCTCTACTTTTTTTT

      ACATCGGAGATAAATTATTGATTATTTAATTAAAAAAAAATAAAACTCTGAGCATTTCCA
1621  ---------+---------+---------+---------+---------+---------+  1680
      TGTAGCCTCTATTTAATAACTAATAAATTAATTTTTTTTTATTTTGAGACTCGTAAAGGT

      CCGAAAAAAAAAAAAAAAAAAAAAAAAAAACTCGAG
1681  ---------+---------+---------+-------  1717
      GGCTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGCTC
```

Figure A.2 : *Musca domestica yolk protein-2 (Mdyp2)* cDNA and 5' coligated region (boxed)

(derived from cDNA clone MdYPB; *Xho*I restriction site implicated in co-ligation is located at nt 351 - 356)

```
    GAATTCGGCACGAGTCGTCTTCGAGTATAGAAGCATTATGAGAGCTAAGTGGCGTAAGAA
1   ---------+---------+---------+---------+---------+---------+   60
    CTTAAGCCGTGCTCAGCAGAAGCTCATATCTTCGTAATACTCTCGATTCACCGCATTCTT

    GCGTATGCGTAGGTCGAAGCGTCAGCGCAGAAAGATGCGTGCAAGGCCCAAATAAATTGA
61  ---------+---------+---------+---------+---------+---------+  120
    CGCATACGCATCCAGCTTCGCAGTCGCGTCTTTCTACGCACGTTCCGGGTTTATTTAACT

    TACTGGAAATCCCGTTATTACAATAATTCCGTACACCTTCTAAACAAATGGGAAGCTTCA
121 ---------+---------+---------+---------+---------+---------+  180
    ATGACCTTTAGGGCAATAATGTTATTAAGGCATGTGGAAGATTTGTTTACCCTTCGAAGT

    AGTTATTACTTTCTTGTGCGCACCCCATTATCACTGTGTGATGATGGTCACGTCAAGTCT
181 ---------+---------+---------+---------+---------+---------+  240
    TCAATAATGAAAGAACACGCGTGGGGTAATAGTGACACACTACTACCAGTGCAGTTCAGA

    CGAAAGCAGGCTGATATTACAACAGCTTCCGGTGCCATGGATGTTTTTTGAAGAAATAAA
241 ---------+---------+---------+---------+---------+---------+  300
    GCTTTCGTCCGACTATAATGTTGTCGAAGGCCACGGTACCTACAAAAAACTTCTTTATTT

    GAGTATCTTCTCCGGATTTCTACTCGAAGTGAAAATATAAAAAAAAAAAACTCGAGACTA
301 ---------+---------+---------+---------+---------+---------+  360
    CTCATAGAAGAGGCCTAAAGATGAGCTTCACTTTTATATTTTTTTTTTTTGAGCTCTGAT

    GTTCTCTCTCTCTCTCTCCGCGTGTTCTCTGGTTTCTGGTATTACAGAAGAAGATCCGGT
361 ---------+---------+---------+---------+---------+---------+  420
    CAAGAGAGAGAGAGAGGCGCACAAGAGACCAAAGACCATAATGTCTTCTTCTAGGCCA

    TCGTGTTTTTGTTAATCTCTAGACAATGAATCCTTTGCGTACAGTTTGCCTGATGATGGG
421 ---------+---------+---------+---------+---------+---------+  480
    AGCACAAAAACAATTAGAGATCTGTTACTTAGGAAACGCATGTCAAACGGACTACTACCC

                                M  N  P  L  R  T  V  C  L  M  M  G

    CGTCCTGGCCTTGGCTAGTGCCTACAGTGCTGGTCCCCGTCCCATGAGCATGAATTCGAA
481 ---------+---------+---------+---------+---------+---------+  540
    GCAGGACCGGAACCGATCACGGATGTCACGACCAGGGGCAGGGTACTCGTACTTAAGCTT

     V  L  A  L  A  S  A  Y  S  A  G  P  R  P  M  S  M  N  S  N

    TCGTAATACGATTAAGAATAGCATGAAACCCACCTCATGGATGTCGATTTCGACATTGCA
541 ---------+---------+---------+---------+---------+---------+  600
    AGCATTATGCTAATTCTTATCGTACTTTGGGTGGAGTACCTACAGCTAAAGCTGTAACGT

     R  N  T  I  K  N  S  M  K  P  T  S  W  M  S  I  S  T  L  Q
```

```
      ATCGCTGCCCTCGTTGAAGGAAATCAAATTGAAGCAACTGGAAGAGATGTCGGCCTTTGA
601   ---------+---------+---------+---------+---------+---------+  660
      TAGCGACGGGAGCAACTTCCTTTAGTTTAACTTCGTTGACCTTCTCTACAGCCGGAAACT

       S  L  P  S  L  K  E  I  K  L  K  Q  L  E  E  M  S  A  F  E

      GGGAGCCGATTTGATCAATCGCTTGTATCACCTGGCCCAAGCCACCCAGGCCTTGGAGCC
661   ---------+---------+---------+---------+---------+---------+  720
      CCCTCGGCTAAACTAGTTAGCGAACATAGTGGACCGGGTTCGGTGGGTCCGGAACCTCGG

       G  A  D  L  I  N  R  L  Y  H  L  A  Q  A  T  Q  A  L  E  P

      CACCTATGCCCCCAGGGCGAGTGAGATCCCAGCCTTCCTTGTAACACCCGATAACCAGAA
721   ---------+---------+---------+---------+---------+---------+  780
      GTGGATACGGGGGTCCCGCTCACTCTAGGGTCGGAAGGAACATTGTGGGCTATTGGTCTT

       T  Y  A  P  R  A  S  E  I  P  A  F  L  V  T  P  D  N  Q  K

      ATCGATTCAATTGAACGAGTTGCCACGTGTGGCCCGCGAACAATCACACTGCGGCAAACA
781   ---------+---------+---------+---------+---------+---------+  840
      TAGCTAAGTTAACTTGCTCAACGGTGCACACCGGGCGCTTGTTAGTGTGACGCCGTTTGT

       S  I  Q  L  N  E  L  P  R  V  A  R  E  Q  S  H  C  G  K  Q

      GGAGGTTACCGTTTTCATTACCGGCCTCCCCTCGAAATTGGAAAGTGTCAAGGAAGCCAC
841   ---------+---------+---------+---------+---------+---------+  900
      CCTCCAATGGCAAAAGTAATGGCCGGAGGGGAGCTTTAACCTTTCACAGTTCCTTCGGTG

       E  V  T  V  F  I  T  G  L  P  S  K  L  E  S  V  K  E  A  T

      CCGCAGCTTGACCCAGGCCTATATGCAACGTTACAATGATGAGTCCAGCTATTACCAAAA
901   ---------+---------+---------+---------+---------+---------+  960
      GGCGTCGAACTGGGTCCGGATATACGTTGCAATGTTACTACTCAGGTCGATAATGGTTTT

       R  S  L  T  Q  A  Y  M  Q  R  Y  N  D  E  S  S  Y  Y  Q  N

      CTCTGCCACCAGCAGCAGCAGCAGCCACTACCACCTCAACAAGAAGCAACAACGTTCGGA
961   ---------+---------+---------+---------+---------+---------+  1020
      GAGACGGTGGTCGTCGTCGTCGTCGGTGATGGTGGAGTTGTTCTTCGTTGTTGCAAGCCT

       S  A  T  S  S  S  S  S  H  Y  H  L  N  K  K  Q  Q  R  S  D

      CAGTGATGAAGATAATTCCAACCAGAAACCTTCCGGCTGTTTGGTTGTTGTGAAATTCGG
1021  ---------+---------+---------+---------+---------+---------+  1080
      GTCACTACTTCTATTAAGGTTGGTCTTTGGAAGGCCGACAAACCAACAACACTTTAAGCC

       S  D  E  D  N  S  N  Q  K  P  S  G  C  L  V  V  V  K  F  G

      TGATACCATTAGCGATTTTGAGGAACACGCCACCGTGGACACCGAAAAGGTTGGTAAGGA
1081  ---------+---------+---------+---------+---------+---------+  1140
      ACTATGGTAATCGCTAAAACTCCTTGTGCGGTGGCACCTGTGGCTTTTCCAACCATTCCT

       D  T  I  S  D  F  E  E  H  A  T  V  D  T  E  K  V  G  K  E

      AGTCGGTAAGTTTTGGTTACAACTTTTGGAAAAGACCAACTGCAATCGCGACAATGTCCA
1141  ---------+---------+---------+---------+---------+---------+  1200
      TCAGCCATTCAAAACCAATGTTGAAAACCTTTTCTGGTTGACGTTAGCGCTGTTACAGGT

       V  G  K  F  W  L  Q  L  L  E  K  T  N  C  N  R  D  N  V  H
```

```
      TTTGATTGGCTCCAATTTGGGCGCCAACATTGCCGGAGCTGCTGGTCGTCAATACACCAA
1201  ---------+---------+---------+---------+---------+---------+  1260
      AAACTAACCGAGGTTAAACCCGCGGTTGTAACGGCCTCGACGACCAGCAGTTATGTGGTT

       L  I  G  S  N  L  G  A  N  I  A  G  A  A  G  R  Q  Y  T  K

      GGTTACCAATCATCAATTGCGCCGCATCACCGGTTTGGACCCCGTCAAATGCTTTGCCAA
1261  ---------+---------+---------+---------+---------+---------+  1320
      CCAATGGTTAGTAGTTAACGCGGCGTAGTGGCCAAACCTGGGGCAGTTTACGAAACGGTT

       V  T  N  H  Q  L  R  R  I  T  G  L  D  P  V  K  C  F  A  K

      GGATCCCGAGACATTGACCGGCTTGGCTCGTGGTGATGCTGAATTCGTCGATGCCATCCA
1321  ---------+---------+---------+---------+---------+---------+  1380
      CCTAGGGCTCTGTAACTGGCCGAACCGAGCACCACTACGACTTAAGCAGCTACGGTAGGT

       D  P  E  T  L  T  G  L  A  R  G  D  A  E  F  V  D  A  I  H

      CACCACTGCCAACAGCATGGGTACCTCTGCCCGTGCCGCTGATGTTGACTTCTACCCCGA
1381  ---------+---------+---------+---------+---------+---------+  1440
      GTGGTGACGGTTGTCGTACCCATGGAGACGGGCACGGCGACTACAACTGAAGATGGGGCT

       T  T  A  N  S  M  G  T  S  A  R  A  A  D  V  D  F  Y  P  E

      GGGACCCAATGAGGCTGTGCCCGGTGCCGACAATATTGTCGAATCTCCGATGCGTGCTGT
1441  ---------+---------+---------+---------+---------+---------+  1500
      CCCTGGGTTACTCCGACACGGGCCACGGCTGTTATAACAGCTTAGAGGCTACGCACGACA

       G  P  N  E  A  V  P  G  A  D  N  I  V  E  S  P  M  R  A  V

      CCGCTACTTTGCCGAATCCGTTGTGCCCGGCAATGAGCGTAACTTCCCAGCTGAGAGTGC
1501  ---------+---------+---------+---------+---------+---------+  1560
      GGCGATGAAACGGCTTAGGCAACACGGGCCGTTACTCGCATTGAAGGGTCGACTCTCACG

       R  Y  F  A  E  S  V  V  P  G  N  E  R  N  F  P  A  E  S  A

      CCACTCTTTGAACGAATACAAGAACAGCAACACCTCGGGACGTCGCATCTACATGGGCAT
1561  ---------+---------+---------+---------+---------+---------+  1620
      GGTGAGAAACTTGCTTATGTTCTTGTCGTTGTGGAGCCCTGCAGCGTAGATGTACCCGTA

       H  S  L  N  E  Y  K  N  S  N  T  S  G  R  R  I  Y  M  G  I

      CATCACCACCTTCAAGGTTGAGGGTGACTACATGTTGCAGGTGAACACCAAGAGTCCATT
1621  ---------+---------+---------+---------+---------+---------+  1680
      GTAGTGGTGGAAGTTCCAACTCCCACTGATGTACAACGTCCACTTGTGGTTCTCAGGTAA

       I  T  T  F  K  V  E  G  D  Y  M  L  Q  V  N  T  K  S  P  F

      CGGTCGTAGCACCCCTGTCCAAAAACAACAAAACGTCCATGGTGTCCACAAGTCGTGGAA
1681  ---------+---------+---------+---------+---------+---------+  1740
      GCCAGCATCGTGGGGACAGGTTTTTGTTGTTTTGCAGGTACCACAGGTGTTCAGCACCTT

       G  R  S  T  P  V  Q  K  Q  Q  N  V  H  G  V  H  K  S  W  K

      AATGTCCTCCTCTCGCGATGAGGAATAAATGGATGGAGAAGAAATCATTCAACTGACCCC
1741  ---------+---------+---------+---------+---------+---------+  1800
      TTACAGGAGGAGAGCGCTACTCCTTATTTACCTACCTCTTCTTTAGTAAGTTGACTGGGG

       M  S  S  S  R  D  E  E  *
```

```
      GGCAATGTCTAATGGAAACCTGAGAGATAGATGTTTGGAACTGGTTCAACAGTTCGAATC
1801  ---------+---------+---------+---------+---------+---------+  1860
      CCGTTACAGATTACCTTTGGACTCTCTATCTACAAACCTTGACCAAGTTGTCAAGCTTAG

      GAACAATTTTTTTTTTGTAATTTTGTAGTTTTTAAGTTCAATTAATATTGAAAAAAGAATT
1861  ---------+---------+---------+---------+---------+---------+  1920
      CTTGTTAAAAAAAAACATTAAAACATCAAAAATTCAAGTTAATTATAACTTTTTTCTTAA

      GAAATATCAATTATGTGCTGTTTAATTTATATAAGATAGATCCGCTTTTATTAAACCAAA
1921  ---------+---------+---------+---------+---------+---------+  1980
      CTTTATAGTTAATACACGACAAATTAAATATATTCTATCTAGGCGAAAATAATTTGGTTT

      AAATTAGACAAATAATAAGAGAATTGCATATAAAAAAGCCAAATAAAAAAAAAAAAAAA
1981  ---------+---------+---------+---------+---------+---------+  2040
      TTTAATCTGTTTATTATTCTCTTAACGTATATTTTTTCGGTTTATTTTTTTTTTTTTTTT

      AAACTCGAG
2041  ---------  2049
      TTTGAGCTC
```

Figure A.3 :*Musca domestica yolk protein-2 (Mdyp2)* cDNA

(derived from cDNA clone MdYPB; 5' co-ligated sequence removed)

```
    CTCGAGACTAGTTCTCTCTCTCTCTCTCCGCGTGTTCTCTGGTTTCTGGTATTACAGAAG
1   ---------+---------+---------+---------+---------+---------+   60
    GAGCTCTGATCAAGAGAGAGAGAGAGGCGCACAAGAGACCAAAGACCATAATGTCTTC

    AAGATCCGGTTCGTGTTTTTGTTAATCTCTAGACAATGAATCCTTTGCGTACAGTTTGCC
61  ---------+---------+---------+---------+---------+---------+  120
    TTCTAGGCCAAGCACAAAAACAATTAGAGATCTGTTACTTAGGAAACGCATGTCAAACGG

                                        M  N  P  L  R  T  V  C  L

    TGATGATGGGCGTCCTGGCCTTGGCTAGTGCCTACAGTGCTGGTCCCCGTCCCATGAGCA
121 ---------+---------+---------+---------+---------+---------+  180
    ACTACTACCCGCAGGACCGGAACCGATCACGGATGTCACGACCAGGGGCAGGGTACTCGT

      M  M  G  V  L  A  L  A  S  A  Y  S  A  G  P  R  P  M  S  M

    TGAATTCGAATCGTAATACGATTAAGAATAGCATGAAACCCACCTCATGGATGTCGATTT
181 ---------+---------+---------+---------+---------+---------+  240
    ACTTAAGCTTAGCATTATGCTAATTCTTATCGTACTTTGGGTGGAGTACCTACAGCTAAA

      N  S  N  R  N  T  I  K  N  S  M  K  P  T  S  W  M  S  I  S

    CGACATTGCAATCGCTGCCCTCGTTGAAGGAAATCAAATTGAAGCAACTGGAAGAGATGT
241 ---------+---------+---------+---------+---------+---------+  300
    GCTGTAACGTTAGCGACGGGAGCAACTTCCTTTAGTTTAACTTCGTTGACCTTCTCTACA

      T  L  Q  S  L  P  S  L  K  E  I  K  L  K  Q  L  E  E  M  S

    CGGCCTTTGAGGGAGCCGATTTGATCAATCGCTTGTATCACCTGGCCCAAGCCACCCAGG
301 ---------+---------+---------+---------+---------+---------+  360
    GCCGGAAACTCCCTCGGCTAAACTAGTTAGCGAACATAGTGGACCGGGTTCGGTGGGTCC

      A  F  E  G  A  D  L  I  N  R  L  Y  H  L  A  Q  A  T  Q  A

    CCTTGGAGCCCACCTATGCCCCCAGGGCGAGTGAGATCCCAGCCTTCCTTGTAACACCCG
361 ---------+---------+---------+---------+---------+---------+  420
    GGAACCTCGGGTGGATACGGGGGTCCCGCTCACTCTAGGGTCGGAAGGAACATTGTGGGC

      L  E  P  T  Y  A  P  R  A  S  E  I  P  A  F  L  V  T  P  D

    ATAACCAGAAATCGATTCAATTGAACGAGTTGCCACGTGTGGCCCGCGAACAATCACACT
421 ---------+---------+---------+---------+---------+---------+  480
    TATTGGTCTTTAGCTAAGTTAACTTGCTCAACGGTGCACACCGGGCGCTTGTTAGTGTGA

      N  Q  K  S  I  Q  L  N  E  L  P  R  V  A  R  E  Q  S  H  C

    GCGGCAAACAGGAGGTTACCGTTTTCATTACCGGCCTCCCCTCGAAATTGGAAAGTGTCA
481 ---------+---------+---------+---------+---------+---------+  540
    CGCCGTTTGTCCTCCAATGGCAAAAGTAATGGCCGGAGGGGAGCTTTAACCTTTCACAGT

      G  K  Q  E  V  T  V  F  I  T  G  L  P  S  K  L  E  S  V  K
```

```
      AGGAAGCCACCCGCAGCTTGACCCAGGCCTATATGCAACGTTACAATGATGAGTCCAGCT
541   ---------+---------+---------+---------+---------+---------+   600
      TCCTTCGGTGGGCGTCGAACTGGGTCCGGATATACGTTGCAATGTTACTACTCAGGTCGA

        E   A   T   R   S   L   T   Q   A   Y   M   Q   R   Y   N   D   E   S   S   Y

      ATTACCAAAACTCTGCCACCAGCAGCAGCAGCAGCCACTACCACCTCAACAAGAAGCAAC
601   ---------+---------+---------+---------+---------+---------+   660
      TAATGGTTTTGAGACGGTGGTCGTCGTCGTCGTCGGTGATGGTGGAGTTGTTCTTCGTTG

        Y   Q   N   S   A   T   S   S   S   S   S   H   Y   H   L   N   K   K   Q   Q

      AACGTTCGGACAGTGATGAAGATAATTCCAACCAGAAACCTTCCGGCTGTTTGGTTGTTG
661   ---------+---------+---------+---------+---------+---------+   720
      TTGCAAGCCTGTCACTACTTCTATTAAGGTTGGTCTTTGGAAGGCCGACAAACCAACAAC

        R   S   D   S   D   E   D   N   S   N   Q   K   P   S   G   C   L   V   V   V

      TGAAATTCGGTGATACCATTAGCGATTTTGAGGAACACGCCACCGTGGACACCGAAAAGG
721   ---------+---------+---------+---------+---------+---------+   780
      ACTTTAAGCCACTATGGTAATCGCTAAAACTCCTTGTGCGGTGGCACCTGTGGCTTTTCC

        K   F   G   D   T   I   S   D   F   E   E   H   A   T   V   D   T   E   K   V

      TTGGTAAGGAAGTCGGTAAGTTTTGGTTACAACTTTTGGAAAAGACCAACTGCAATCGCG
781   ---------+---------+---------+---------+---------+---------+   840
      AACCATTCCTTCAGCCATTCAAAACCAATGTTGAAAACCTTTTCTGGTTGACGTTAGCGC

        G   K   E   V   G   K   F   W   L   Q   L   L   E   K   T   N   C   N   R   D

      ACAATGTCCATTTGATTGGCTCCAATTTGGGCGCCAACATTGCCGGAGCTGCTGGTCGTC
841   ---------+---------+---------+---------+---------+---------+   900
      TGTTACAGGTAAACTAACCGAGGTTAAACCCGCGGTTGTAACGGCCTCGACGACCAGCAG

        N   V   H   L   I   G   S   N   L   G   A   N   I   A   G   A   A   G   R   Q

      AATACACCAAGGTTACCAATCATCAATTGCGCCGCATCACCGGTTTGGACCCCGTCAAAT
901   ---------+---------+---------+---------+---------+---------+   960
      TTATGTGGTTCCAATGGTTAGTAGTTAACGCGGCGTAGTGGCCAAACCTGGGGCAGTTTA

        Y   T   K   V   T   N   H   Q   L   R   R   I   T   G   L   D   P   V   K   C

      GCTTTGCCAAGGATCCCGAGACATTGACCGGCTTGGCTCGTGGTGATGCTGAATTCGTCG
961   ---------+---------+---------+---------+---------+---------+  1020
      CGAAACGGTTCCTAGGGCTCTGTAACTGGCCGAACCGAGCACCACTACGACTTAAGCAGC

        F   A   K   D   P   E   T   L   T   G   L   A   R   G   D   A   E   F   V   D

      ATGCCATCCACACCACTGCCAACAGCATGGGTACCTCTGCCCGTGCCGCTGATGTTGACT
1021  ---------+---------+---------+---------+---------+---------+  1080
      TACGGTAGGTGTGGTGACGGTTGTCGTACCCATGGAGACGGGCACGGCGACTACAACTGA

        A   I   H   T   T   A   N   S   M   G   T   S   A   R   A   A   D   V   D   F

      TCTACCCCGAGGGACCCAATGAGGCTGTGCCCGGTGCCGACAATATTGTCGAATCTCCGA
1081  ---------+---------+---------+---------+---------+---------+  1140
      AGATGGGGCTCCCTGGGTTACTCCGACACGGGCCACGGCTGTTATAACAGCTTAGAGGCT

        Y   P   E   G   P   N   E   A   V   P   G   A   D   N   I   V   E   S   P   M
```

```
      TGCGTGCTGTCCGCTACTTTGCCGAATCCGTTGTGCCCGGCAATGAGCGTAACTTCCCAG
1141  ---------+---------+---------+---------+---------+---------+  1200
      ACGCACGACAGGCGATGAAACGGCTTAGGCAACACGGGCCGTTACTCGCATTGAAGGGTC

        R  A  V  R  Y  F  A  E  S  V  V  P  G  N  E  R  N  F  P  A

      CTGAGAGTGCCCACTCTTTGAACGAATACAAGAACAGCAACACCTCGGGACGTCGCATCT
1201  ---------+---------+---------+---------+---------+---------+  1260
      GACTCTCACGGGTGAGAAACTTGCTTATGTTCTTGTCGTTGTGGAGCCCTGCAGCGTAGA

        E  S  A  H  S  L  N  E  Y  K  N  S  N  T  S  G  R  R  I  Y

      ACATGGGCATCATCACCACCTTCAAGGTTGAGGGTGACTACATGTTGCAGGTGAACACCA
1261  ---------+---------+---------+---------+---------+---------+  1320
      TGTACCCGTAGTAGTGGTGGAAGTTCCAACTCCCACTGATGTACAACGTCCACTTGTGGT

        M  G  I  I  T  T  F  K  V  E  G  D  Y  M  L  Q  V  N  T  K

      AGAGTCCATTCGGTCGTAGCACCCCTGTCCAAAAACAACAAAACGTCCATGGTGTCCACA
1321  ---------+---------+---------+---------+---------+---------+  1380
      TCTCAGGTAAGCCAGCATCGTGGGGACAGGTTTTTGTTGTTTTGCAGGTACCACAGGTGT

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

      AGTCGTGGAAAATGTCCTCCTCTCGCGATGAGGAATAAATGGATGGAGAAGAAATCATTC
1381  ---------+---------+---------+---------+---------+---------+  1440
      TCAGCACCTTTTACAGGAGGAGAGCGCTACTCCTTATTTACCTACCTCTTCTTTAGTAAG

        S  W  K  M  S  S  S  R  D  E  E  *

      AACTGACCCCGGCAATGTCTAATGGAAACCTGAGAGATAGATGTTTGGAACTGGTTCAAC
1441  ---------+---------+---------+---------+---------+---------+  1500
      TTGACTGGGGCCGTTACAGATTACCTTTGGACTCTCTATCTACAAACCTTGACCAAGTTG


      AGTTCGAATCGAACAATTTTTTTTTTGTAATTTTGTAGTTTTTAAGTTCAATTAATATTGA
1501  ---------+---------+---------+---------+---------+---------+  1560
      TCAAGCTTAGCTTGTTAAAAAAAAACATTAAAACATCAAAAATTCAAGTTAATTATAACT


      AAAAAGAATTGAAATATCAATTATGTGCTGTTTAATTTATATAAGATAGATCCGCTTTTA
1561  ---------+---------+---------+---------+---------+---------+  1620
      TTTTTCTTAACTTTATAGTTAATACACGACAAATTAAATATATTCTATCTAGGCGAAAAT


      TTAAACCAAAAAATTAGACAAATAATAAGAGAATTGCATATAAAAAAGCCAAATAAAAAA
1621  ---------+---------+---------+---------+---------+---------+  1680
      AATTTGGTTTTTTAATCTGTTTATTATTCTCTTAACGTATATTTTTTCGGTTTATTTTTT


      AAAAAAAAAAAAACTCGAG
1681  ---------+---------  1699
      TTTTTTTTTTTTTGAGCTC
```

Figure A.4 : *Musca domestica yolk protein-3 (Mdyp3)* cDNA

(derived from cDNA clone MdYPE)

```
    GAATTCGGCACGAGCAACAGTTGAAGTTGTCCGGCAAATATTAGCGGAACTTGGAAACAA
1   ---------+---------+---------+---------+---------+---------+   60
    CTTAAGCCGTGCTCGTTGTCAACTTCAACAGGCCGTTTATAATCGCCTTGAACCTTTGTT

    CAAATCGAAAACAAAATGAATCCTTTAGTAATTTTGGGCTTTGTGGCCATGGTGGCTGTA
61  ---------+---------+---------+---------+---------+---------+  120
    GTTTAGCTTTTGTTTTACTTAGGAAATCATTAAAACCCGAAACACCGGTACCACCGACAT

                    M  N  P  L  V  I  L  G  F  V  A  M  V  A  V

    GGATCGTTGGCATCGCCAACAAACCAGAACAGCATGAAGCCATCGCAATGGTTGAAACCC
121 ---------+---------+---------+---------+---------+---------+  180
    CCTAGCAACCGTAGCGGTTGTTTGGTCTTGTCGTACTTCGGTAGCGTTACCAACTTTGGG

    G  S  L  A  S  P  T  N  Q  N  S  M  K  P  S  Q  W  L  K  P

    AGTGAGCTGGAGAGTACACCATCCCTGGATGAGTTGACCTTCGAGGAATTGGAGAAAATG
181 ---------+---------+---------+---------+---------+---------+  240
    TCACTCGACCTCTCATGTGGTAGGGACCTACTCAACTGGAAGCTCCTTAACCTCTTTTAC

    S  E  L  E  S  T  P  S  L  D  E  L  T  F  E  E  L  E  K  M

    CCATTGGAAAAAGGAGCCAAATTGATGCGCAAAATATATCACTTGGCCCAAATCGAAAAC
241 ---------+---------+---------+---------+---------+---------+  300
    GGTAACCTTTTTCCTCGGTTTAACTACGCGTTTTATATAGTGAACCGGGTTTAGCTTTTG

    P  L  E  K  G  A  K  L  M  R  K  I  Y  H  L  A  Q  I  E  N

    TCTGTCTCGCCCAATTTCGTGCCCAGCCCCAGCAATGTGCCCGTCTACATTTTCAACGGC
301 ---------+---------+---------+---------+---------+---------+  360
    AGACAGAGCGGGTTAAAGCACGGGTCGGGGTCGTTACACGGGCAGATGTAAAAGTTGCCG

    S  V  S  P  N  F  V  P  S  P  S  N  V  P  V  Y  I  F  N  G

    AAGGGTGAAAAGGAGACTTGCAACTTGAACAACTACGTCGACATTGCCAAGAACAAGCCC
361 ---------+---------+---------+---------+---------+---------+  420
    TTCCCACTTTTCCTCTGAACGTTGAACTTGTTGATGCAGCTGTAACGGTTCTTGTTCGGG

    K  G  E  K  E  T  C  N  L  N  N  Y  V  D  I  A  K  N  K  P

    AAATTTGGCGAACAAGAAGTCACCGTTTTCATTACTGGCCTGCCCCAGTCCTTGGATGAT
421 ---------+---------+---------+---------+---------+---------+  480
    TTTAAACCGCTTGTTCTTCAGTGGCAAAAGTAATGACCGGACGGGGTCAGGAACCTACTA

    K  F  G  E  Q  E  V  T  V  F  I  T  G  L  P  Q  S  L  D  D

    GTCAAGAAGGCCAACACCCGATTGATCCAAGCCTACATTCAACGCTACAGCCAAAAACCC
481 ---------+---------+---------+---------+---------+---------+  540
    CAGTTCTTCCGGTTGTGGGCTAACTAGGTTCGGATGTAAGTTGCGATGTCGGTTTTTGGG

    V  K  K  A  N  T  R  L  I  Q  A  Y  I  Q  R  Y  S  Q  K  P
```

```
     ACTCCACCCAGGGATGATGACAAATCGAAATGGGAAAATGAACAACCCGTTGGCGGCCAT
541  ---------+---------+---------+---------+---------+---------+  600
     TGAGGTGGGTCCCTACTACTGTTTAGCTTTACCCTTTTACTTGTTGGGCAACCGCCGGTA

     T  P  P  R  D  D  D  K  S  K  W  E  N  E  Q  P  V  G  G  H

     TTGGTTGTCATCGATTTGGGTCACACCATCACCGACATGGAACGTTATGCCAGTTTGGAT
601  ---------+---------+---------+---------+---------+---------+  660
     AACCAACAGTAGCTAAACCCAGTGTGGTAGTGGCTGTACCTTGCAATACGGTCAAACCTA

     L  V  V  I  D  L  G  H  T  I  T  D  M  E  R  Y  A  S  L  D

     GTCAAGGAGACCGGTAAAATGATTGGCAAGACCTTTGCCGAGCTGATGGATGAGTGCGAT
661  ---------+---------+---------+---------+---------+---------+  720
     CAGTTCCTCTGGCCATTTTACTAACCGTTCTGGAAACGGCTCGACTACCTACTCACGCTA

     V  K  E  T  G  K  M  I  G  K  T  F  A  E  L  M  D  E  C  D

     GTCGATGTTGAGGATATGCATGTTGTTGCCCAGGGCATTGCTACCAATGTTGGCGGTTCG
721  ---------+---------+---------+---------+---------+---------+  780
     CAGCTACAACTCCTATACGTACAACAACGGGTCCCGTAACGATGGTTACAACCGCCAAGC

     V  D  V  E  D  M  H  V  V  A  Q  G  I  A  T  N  V  G  G  S

     GCTGGCAAGGACTTCAAGGACATTACCACCCACAAATTGGATCGCATCACCGCTTTGGAC
781  ---------+---------+---------+---------+---------+---------+  840
     CGACCGTTCCTGAAGTTCCTGTAATGGTGGGTGTTTAACCTAGCGTAGTGGCGAAACCTG

     A  G  K  D  F  K  D  I  T  T  H  K  L  D  R  I  T  A  L  D

     CCTGCCCGTCAAGTTGCCAAGAATCCCAAGGTGTTGTCCGGTTTAGCTCGTGGCTCCGCC
841  ---------+---------+---------+---------+---------+---------+  900
     GGACGGGCAGTTCAACGGTTCTTAGGGTTCCACAACAGGCCAAATCGAGCACCGAGGCGG

     P  A  R  Q  V  A  K  N  P  K  V  L  S  G  L  A  R  G  S  A

     AACTTTGTTGATGCCATTCACACTTCCGCTTTGGGTATGGGCACCACCCGTCGTGTTGGT
901  ---------+---------+---------+---------+---------+---------+  960
     TTGAAACAACTACGGTAAGTGTGAAGGCGAAACCCATACCCGTGGTGGGCAGCACAACCA

     N  F  V  D  A  I  H  T  S  A  L  G  M  G  T  T  R  R  V  G

     GATGTTGATTTCTTCCCTCATGGCCCCTGTGAAGGTGTTCCCGGTACCCGCAACGTAATT
961  ---------+---------+---------+---------+---------+---------+  1020
     CTACAACTAAAGAAGGGAGTACCGGGGACACTTCCACAAGGGCCATGGGCGTTGCATTAA

     D  V  D  F  F  P  H  G  P  C  E  G  V  P  G  T  R  N  V  I

     GAAGCCCAGGCTCGTGCCACCCGTTTCTATGCCGAATCTGTACGTCCCGGTAACAGCCGT
1021 ---------+---------+---------+---------+---------+---------+  1080
     CTTCGGGTCCGAGCACGGTGGGCAAAGATACGGCTTAGACATGCAGGGCCATTGTCGGCA

     E  A  Q  A  R  A  T  R  F  Y  A  E  S  V  R  P  G  N  S  R

     AATTTCCCAGCCCTTGAAGCCAGCTCCCTGAAGCAGTACCGCAACAAGGATAGCTATGGC
1081 ---------+---------+---------+---------+---------+---------+  1140
     TTAAAGGGTCGGGAACTTCGGTCGAGGGACTTCGTCATGGCGTTGTTCCTATCGATACCG

     N  F  P  A  L  E  A  S  S  L  K  Q  Y  R  N  K  D  S  Y  G
```

```
      AAACGTGCCTACATGGGTATTGCCACACGTCGTGACACCACCGGTGACTACATTTTGGAA
1141  ---------+---------+---------+---------+---------+---------+  1200
      TTTGCACGGATGTACCCATAACGGTGTGCAGCACTGTGGTGGCCACTGATGTAAAACCTT

      K  R  A  Y  M  G  I  A  T  R  R  D  T  T  G  D  Y  I  L  E

      GTCAATGAGCAGACTCCCTTTGGCAAGCGCTCAGCTCCCCAGCAAAGATCTGTCCAATCT
1201  ---------+---------+---------+---------+---------+---------+  1260
      CAGTTACTCGTCTGAGGGAAACCGTTCGCGAGTCGAGGGGTCGTTTCTAGACAGGTTAGA

      V  N  E  Q  T  P  F  G  K  R  S  A  P  Q  Q  R  S  V  Q  S

      TTCAACAGTGAAAACTATTGAGGAATGATAGATGGTTTTACTTCCATTTCTACCCCTGGC
1261  ---------+---------+---------+---------+---------+---------+  1320
      AAGTTGTCACTTTTGATAACTCCTTACTATCTACCAAAATGAAGGTAAAGATGGGGACCG

      F  N  S  E  N  Y  *

      GTTTAATTTTGAACTGTGAACTGTTTATAGTTTATGATGGAACTTTTTTATATTTGGACC
1321  ---------+---------+---------+---------+---------+---------+  1380
      CAAATTAAAACTTGACACTTGACAAATATCAAATACTACCTTGAAAAAATATAAACCTGG


      ACACATCCCCACATTTTTTAGATTTTAGTTGAAACATAAAGAAAAAATAAAAAAGAATAT
1381  ---------+---------+---------+---------+---------+---------+  1440
      TGTGTAGGGGTGTAAAAAATCTAAAATCAACTTTGTATTTCTTTTTTATTTTTTCTTATA


      TTGATATTAAAAAAAAAAAAAAAAAACTCGAG
1441  ---------+---------+---------+--  1472
      AACTATAATTTTTTTTTTTTTTTTTGAGCTC
```

6. BIBLIOGRAPHY

ABRAHAMSEN, N., A. MARTINEZ, T.KJAER, L. SØNDERGAARD, and M. BOWNES (1993) " *Cis*-regulatory sequences leading to the fat body, sex specific, expression of the *yolk protein* genes 1 and 2 of *Drosophila melanogaster* " MOL. GEN. GENET. 237 : 41 - 48

ADAMS, T.S. (1974) "The role of juvenile hormone in housefly ovarian follicle morphogenesis" J. INSECT PHYSIOL. 20 : 263 - 276

ADAMS, T.S. (1981) "Activation of successive ovarian gonotrophic cycles by the corpus allatum in the housefly, *Musca domestica* (Diptera : Muscidae)" INT. J. INVERT. REPROD. 3 : 41 - 48

ADAMS, T.S., and P.A. FILIPI (1983) "Vitellin and vitellogenin concentrations during oogenesis in the first gonotrophic cycle of the housefly, *Musca domestica*" J. INSECT PHYSIOL. 29(9) : 723 - 733

ADAMS, T.S., H.H. HAGEDORN, and G.D. WHEELOCK (1985) "Haemolymph ecdysteroid in the housefly, *Musca domestica*, during oogenesis and its relationship with vitellogenin levels" J. INSECT PHYSIOL. 31 : 91 - 97

ADAMS, T.S., T.J. KELLY, and C.W. WOODS (1988) "Relationship between the corpus cardiacum-allatum complex and the ovaries with the haemolymph ecdysteroid profile in the housefly, *Musca domestica*" J. INSECT. PHYSIOL. 34 : 1105 - 1109

ADAMS, T.S., and P.A. FILIPI (1988) "Interaction between juvenile hormone, 20-hydroxyecdysone, the corpus cardiacum-allatum complex, and the ovaries in regulating vitellogenin levels in the housefly, *Musca domestica*" J. INSECT PHYSIOL. 34 : 11 - 19

ADAMS, T.S., and D.R. NELSON (1990) "The influence of diet on ovarian maturation, mating, and pheromone production in the housefly, *Musca domestica*" INVERT. REPROD. DEV. 17 : 193 - 201

ADAMS, T.S., and J.W. GERST (1991) "The effect of pulse feeding a protein diet on ovarian maturation, vitellogenin levels, and ecdysteroid titre in houseflies, *Musca domestica*, maintained on sucrose" INVERT. REPROD. DEV. 20 : 49 - 57

ADAMS, T.S., and J.W. GERST (1992) "Interaction between diet and hormones on vitellogenin levels in the housefly, *Musca domestica*" INVERT. REPROD. DEV. 21 : 91 - 98

ADAMS, T.S., and J.W. GERST (1993) "Effect of diet on vitellogenin, vitellin and ecdysteroid levels during the second cycle of oogenesis in the housefly, *Musca domestica*" J. INSECT PHYSIOL. 39(10) : 835 - 843

AGUI, N., M. TAKAHASHI, Y. WADA, S. IZUMI, and S. TOMINO (1985) "The relationship between nutrition, vitellogenin, vitellin and ovarian development in the housefly, *Musca domestica* L." J. INSECT PHYSIOL. 31(9) : 715 - 722

ALBRECHT, E.B. and H.K. SALZ (1993) " The *Drosophila* sex determination gene *snf* is utilised for the establishment of the female-specific splicing pattern of *Sex-lethal* " GENETICS 134 : 801 - 807

ALTING-MEES, M.A., and J.M. SHORT (1989) "pBluescript II : gene mapping vectors" NUCLEIC ACIDS RESEARCH 17(22) : 9494

AMREIN, H., M. GORMAN, and R. NÖTHIGER (1988) " The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein " CELL 55 : 1025 - 1035

AMREIN, H., T. MANIATIS, and R. NÖTHIGER (1990) " Alternatively spliced transcripts of the sex-determining gene *tra-2* of *Drosophila* encode functional proteins of different size " EMBO J. 9(11) : 3619 - 3629

ANDREWS, A.T. (1991) "Electrophoresis of nucleic acids" In 'Essential Molecular Biology : A practical Approach (Volume I)' IRL PRESS (Ed. T. A. Brown)

AN, W., and P.C. WENSINK (1995a) "Three protein binding sites from an enhancer that directs the sex and fat body specific transcription of *Drosophila yolk protein* genes " EMBO J. 14(6) : 1221 - 1230

AN, W., and P.C. WENSINK (1995b) " Integrating sex- and tissue-specific regulation within a single *Drosophila* enhancer " GENES & DEVEL. 9 : 256 - 266

BACHILER, D., and L. SANCHÉZ (1986) " Mutations affecting dosage compensation in *Drosophila melanogaster* : Effects in the germline " DEVEL. BIOL. 118 : 379 - 384

BAEUERLE, P.A., and W.B. HUTTNER (1985) "Tyrosine sulfation of yolk proteins-1, 2 and 3 in *Drosophila melanogaster*" J. BIOL. CHEM. 260 : 6434 - 6439

BAEUERLE, P.A., F. LOTTSPEICH, and W.B. HUTTNER (1988) " Purification of yolk protein 2 of *Drosophila melanogaster* and identification of its site of tyrosine sulfation " J. OF BIOL. CHEM. 263(29) : 14925 - 14929

BAKER, B.S., and K. RIDGE (1980) " Sex and the single cell : On the action of major loci affecting sex determination in *Drosophila melanogaster* " GENETICS 94 : 383 - 423

BAKER, B.S. and J.M. BELOTE (1983) "Sex determination and dosage compensation in *Drosophila melanogaster* " ANN. REV. GENET. 17 : 345 - 393

BAKER, B.S., and M.F. WOLFNER (1988) " A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila melanogaster* " GENES & DEVEL. 2 : 477 - 489

BANDZIULIS, R.J., M.S. SWANSON, and G. GREYFUSS (1989) " RNA-binding proteins as developmental regulators " GENES AND DEVEL. 3 : 431 - 437

BARNETT, T., C. PACHL, J.P. GERGEN, and P.C. WENSINK (1980) " The isolation and characterisation of *Drosophila* yolk protein genes " CELL 21 : 729 - 738

BELL, A.E. (1954) " A gene in *Drosophila melanogaster* that produces all male progeny " GENETICS 39 : 958 - 959

BELL, L., J.I. HORABIN, P. SCHEDL, and T.W. CLINE (1991) " Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in *Drosophila* " CELL 65 : 229 - 239

BELL, L.R., E.M. MAINE, P. SCHEDL, and T.W.CLINE (1988) "*Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins " CELL 55 : 1037 - 1046

BELOTE, J.M., and J.C. LUCCHESI (1980) " Male-specific lethal mutations of *Drosophila melanogaster* " GENETICS 96 : 165 - 186

BELOTE, J.M., and B.S. BAKER (1980) " Sex determination in *Drosophila melanogaster* : Analysis of *transformer-2*, a sex transforming locus " PROC. NAT. ACAD. SCI. USA 79 : 1568 - 1572

BELOTE, J.M., and B.S. BAKER (1982) " Sex-determination in *Drosophila melanogaster* : Analysis of *transformer-2*, a sex-transforming locus" PROC. NATL. ACAD. SCI. USA 79 : 1568 - 1572

BELOTE, J.M. and B.S. BAKER (1983) " The dual functions of a sex determination gene in *Drosophila melanogaster* " DEVEL. BIOL. 95 : 512 - 517

BELOTE, J.M., A.M. HANDLER, M.F. WOLFNER, K.J. LIVAK, and B.S. BAKER (1985) " Sex-specific regulation of yolk protein gene expression in *Drosophila*" CELL 40 : 339 - 348

BELOTE, J.M. (1992) " Sex determination in *Drosophila melanogaster* : from the X:A ratio to *doublesex* " SEM. IN DEVEL. BIOL. 3 : 319 - 330

BERNSTEIN, M., and T.W. CLINE (1994) " Differential effects of *Sex-lethal* mutations on dosage compensation early in *Drosophila* development " GENETICS 136 : 1051 - 1061

BERNSTEIN, M. R.A. LERSCH, L. SUBRAHMANYAN, and T.W. CLINE (1995) "Transposon insertions causing constitutive *Sex-lethal* activity in *Drosophila melanogaster* affect *Sxl* sex-specific transcript splicing " GENETICS 139 : 631 - 648

BEVERLEY, M.S., and A.C. WILSON (1982) "Molecular evolution in *Drosophila* and higher Diptera I. Micro-complement fixation studies of larval haemolymph protein " J. MOL. EVOL. 18 : 251 - 264

BEVERLEY, M.S., and A.C. WILSON (1984) "Molecular evolution in *Drosophila* and higher Diptera II. A time scale for fly evolution " J. MOL. EVOL. 21 : 1 - 13

BIER, E., H. VAESSIN, S. SHEPHERD, K. LEE, M. McCALL, S. BARBEL, L. ACKERMAN, R. CARRETTO, T. UEMURA, E. GRELL, L.Y. JAN, and Y.N. JAN (1989) " Searching for pattern and mutation in the *Drosophila* genome with a P-*LacZ* vector " GENES & DEVEL. 3 : 1273 - 1287

BINGHAM, P.M., T. CHOU, I. MIMS, and Z. ZACHER (1988) " On/off regulation of gene expression at the level of splicing " TIG 4(5) : 135 - 138

BOGGS, R.T., P. GREGOR, S. IDRISS, J.M. BELOTE, and M. McKEOWN (1987) " Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene " CELL 50 : 739 - 747

BONE, J.R., J. LAVENDER, R. RICHMAN, M.J. PALMER, B.M. TURNER, and M. KURODA (1994) " Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila* " GENES & DEVEL. 8 : 96 - 104

BOPP, D., J.I. HORABIN, R.A. LERSCH, T.W. CLINE, and P. SCHEDL (1993) "Expression of the *Sex-lethal* gene is controlled at multiple levels during *Drosophila* oogenesis " DEVELOPMENT 118 : 797 - 812

BOPP, D., L.R. BELL, T.W. CLINE, and P. SCHEDL (1991) " Developmental distribution of female-specific *Sex-lethal* proteins in *Drosophila* development "GENES & DEVEL. 5 : 403 - 415

BOWNES, M., and L. DALE (1968) "A handbook of *Drosophila* development", Elsevier Biomedical Press (ed. R. Ranson) : 31 - 48

BOWNES, M., and B.D. HAMES (1977) " Accumulation and degradation of three yolk proteins in *Drosophila melanogaster* " J. EXP. ZOOL. 200 : 149 - 156

BOWNES, M. and B.D. HAMES (1978) " Analysis of yolk proteins in *Drosophila melanogaster*, translation in a cell-free system and peptide analysis " FEBS LETTERS 96 : 327 - 330

BOWNES, M., and R. NÖTHIGER (1981) " Sex-determining genes and vitellogenin synthesis in *Drosophila melanogaster* " DEVEL. BIOL. 89 : 225 - 236

BOWNES, M., M. BLAIR, R. KOZMA, and M. DEMPSTER (1983) " 20-hydroxyecdysone stimulates tissue-specific *yolk-protein* gene transcription in both male and female *Drosophila*" J. EMBRYOL. EXP. MORPHOL. 78 : 249 - 268

BOWNES, M. (1986) " Expression of the genes coding for vitellogenin (yolk protein) " ANN REV. ENTOMOL. 31 : 507 - 531

BOWNES, M., and M. BLAIR (1986) " The effects of a sugar diet and hormones on the expression of the *Drosophila yolk-protein* genes " J. INSECT PHYSIOL. 32(5) : 493 - 501

BOWNES, M., A. SCOTT, and M. BLAIR (1987) " The use of an inhibitor of protein synthesis to investigate the roles of ecdysteroids and sex-determination genes on the expression of the genes encoding the *Drosophila* yolk proteins " DEVELOPMENT 101 : 931 - 941

BOWNES, M., A. SHIRRAS, M. BLAIR, J. COLLINS, and A. COULSON (1988) " Evidence that insect embryogenesis is regulated by ecdysteroids released from yolk proteins " PROC. NAT. ACAD. SCI. USA 85 : 1554 - 1557

BOWNES, M. (1989) " The role of juvenile hormone, ecdysone and the ovary in the control of *Drosophila* vitellogenesis " J. INSECT PHYSIOL. 35(5) : 409 - 413

BOWNES, M., and G. REID (1990) " The role of the ovary and nutritional signals in the regulation of fat body *yolk protein* gene expression in *Drosophila melanogaster* " J. INSECT PHYSIOL. 36 : 471 - 479

BOWNES, M. (1992) " Why is there sequence similarity between insect yolk proteins and vertebrate lipases ? " J. OF LIPID RES. 33 : 777 - 780

BOWNES, M., E. RONALDSON, D. MAUCHLINE, and A MARTINEZ (1993) " Regulation of vitellogenesis in *Drosophila* " INT. J. INSECT MORPHOL. & EMBRYOL. 22(2-4) : 349 - 367

BOWNES, M., (1994) " The regulation of *yolk protein* genes, a family of sex differentiation genes in *Drosophila melanogaster* " BIOESSAYS 16(10) : 745 - 752

BOWNES, M., E. RONALDSON, and D. MAUCHLINE (1996) "20-hydroxyecdysone, but not juvenile hormone, regulation of *yolk protein* gene expression can be mapped to *cis*-acting DNA sequences" DEVEL. BIOL. 173(2) : 475 - 489

BRENNAN, M.D., T.G. WARREN, and A.P. MAHOWALD (1980) " Signal peptides and signal peptidase in *Drosophila melanogaster* " J. CELL BIOL. 87 : 516 - 520

BRENNAN, M.D., A.J. WEINER, T.J. GORALSKI, and A.P. MAHOWALD (1982) " The follicle cells are a major site of vitellogenin synthesis in *Drosophila melanogaster* " DEVEL. BIOL. 89 : 225 - 236

BRIDGES, C.B. (1916) " Non-disjunction as proof of the chromosome theory of heredity " GENETICS 1 : 1 - 52

BRIDGES, C.B. (1921) " Triploid intersexes in *Drosophila melanogaster* " SCIENCE 54 : 252 -254

BROWN, E.H., and R.C. KING (1961) " Studies on the expression of the *transformer* gene in *Drosophila melanogaster* " GENETICS 46 : 143 - 156

BULLOCK, W.D., J.M. FERNANDEZ, and J.M. SHORT (1987) "XL1 BLUE : A high efficiency plasmid transforming *recA Eschericia coli* strain with β-galactosidase selection" BIOTECHNIQUES 5(4) : 376 - 379

BURTIS, K.C., and B.S. BAKER (1989a) " A molecular analysis of *doublesex*, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila melanogaster* " GENES & DEVEL. 2 : 477 - 489

BURTIS, K.C., and B.S. BAKER (1989b) "*Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNA's encoding related sex-specific polypeptides" CELL 56 : 997 - 1010

BURTIS, K.C., K.T. COSCHIGANO, B.S. BAKER, and P.C. WENSINK (1991) " The *doublesex* proteins of *Drosophila melanogaster* bind directly to a sex-specific *yolk protein* gene enhancer " EMBO J. 10(9) :2577 - 2582

BURTIS, K.C., and M.F. WOLFNER (1992) " The view from the bottom : sex-specific traits and their control in *Drosophila* " SEM. IN DEVEL. BIOL. 3 : 331 - 340

BURTIS, K.C. (1993) " The regulation of sex determination and sexually dimorphic differentiation in *Drosophila* " CURR. OPIN. CELL BIOL. 5 : 1006 - 1014

BUTLER, B., V. PIRROTTA, I. IRMINGER-FINGER, and R. NÖTHIGER (1986) " The sex-determining gene *tra* of *Drosophila* : molecular cloning and transformation studies " EMBO J. 5 : 3607 - 3613

BUTTERWORTH, F.M., V.S. BURDE, and M. BOWNES (1992) " Mutant protein folding leads to female sterility in *Drosophila* " DEVEL. BIOL. 154 : 182 - 194

CAMPUZANO, S., L. CARRAMOLINO, C.V. CABRERA, M. RUIZ-GÓMEZ, R. VILLARES, A. BORONAT, and J. MODOLELL (1985) " Molecular genetics of the *achaete-scute* gene complex of *D. melanogaster* " CELL 40 : 327 - 338

CAUDY, M., E.H. GRELL, C. DAMBLY-CHAUDIERÉ, A. GHYSEN, L.Y. JAN, and Y.N. JAN (1988a) " The maternal sex-determination gene *daughterless* has zygotic activity necessary for the formation of peripheral nerves in *Drosophila* " GENES & DEVEL. 2 : 843 - 852

CAUDY, M., H. VÄSSIN, M. BRAND, R. TUMA, L.Y. JAN, and Y.N. JAN (1988b) " *daughterless*, a gene essential for both neurogenesis and sex determination, has sequence similarities to *myc* and the *achaete-scute* complex " CELL 55 : 1061 - 1067

CHASE, B.A., and B.S. BAKER (1995) " A genetic analysis of *intersex*, a gene regulating sexual differentiation in *Drosophila melanogaster* females " GENETICS 139 : 1649 - 1661

CHOMCZYNSKI, P., and N. SACCHI (1987) "" ANAL. BIOCHEM. 162 : 156

CHUNG, C.T., S.L. NIEMELA, and R.H. MILLER (1989) "One-step preparation of competent *Eschericia coli* : Transformation and storage of bacterial cells in the same solution" PROC. NATL. ACAD. SCI. USA 86 : 2172 - 2175

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. (1978) " Two closely linked mutations in *Drosophila melanogaster* that are lethal to opposite sexes and interact with *daughterless* " GENETICS 90 : 683 - 698

CLINE, T.W. (1979a) " A male-specific lethal mutation in *Drosophila melanogaster* that transforms sex " DEVEL. BIOL. 72 : 266 -275

CLINE, T.W. (1979b) " A product of the maternally-influenced *Sex-lethal* gene determines sex in *Drosophila melanogaster* females " GENETICS 139 : 1649 - 1661

CLINE, T.W. (1980) " Maternal and zygotic sex-specific gene interactions in *Drosophila melanogaster* " GENETICS 96 : 903 - 926

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* " DEVEL. BIOL. 95 : 260 - 274

CLINE, T.W. (1984) " Autoregulatory functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state " GENETICS 107 : 231 - 277

CLINE, T.W. (1986) " A female-specific lesion in an X-linked positive regulator of the *Drosophila* sex determination gene, *Sex-lethal* " GENETICS 113 : 641 - 663

CLINE, T.W. (1988) " Evidence that *sisterless-a* and *sisterless-b* are two of several discrete 'numerator' elements of the X/A sex determination signal in *Drosophila* that switch *Sxl* between two alternative stable expression states " GENETICS 119 : 829 - 862

CLINE, T.W. (1993) " The *Drosophila* sex determination signal : how do flies count to two ? " T.I.G. 9(11) : 385 - 390

COSCHIGANO, K.T., and P.C. WENSINK (1993) " Sex-specific transcriptional regulation by the male and female *doublesex* proteins of *Drosophila* " GENES & DEVEL. 7 : 42 - 54

COLLIER, G.E., and R.J. MacINTYRE (1977) "Microcomplement fixation studies on the evolution of α-glycerophosphate dehydrogenase within the genus *Drosophila* " PROC. NATL. ACAD. SCI. U.S.A. 74 : 684 - 688

CRONMILLER, C., and T.W. CLINE (1987) " The *Drosophila* sex determination gene *daughterless* has different functions in the germ line versus the soma " CELL 48 : 479 - 487

CRONMILLER, C., P. SCHEDL, and T.W. CLINE (1988) " Molecular characterisation of *daughterless*, a *Drosophila* sex determination gene with multiple roles in development" GENES & DEVEL. 2 : 1666 - 1676

CUMMINGS, C.A., and C. CRONMILLER (1994) " The *daughterless* gene functions together with *Notch* and *delta* in the control of ovarian follicle development in *Drosophila* " DEVELOPMENT 120 : 381 - 394

CURRIE, D.A., and M. BATE (1995) " Innervation is essential for the development and differentiation of a sex-specific adult muscle in *Drosophila melanogaster* " DEVELOPMENT 121 : 2549 - 2557

DAMBLY-CHAUDIÈRE, C., A. GHYSEN, L.Y. JAN, and Y.N. JAN (1988) " The determination of sense organs in *Drosophila* : interactions of *scute* with *daughterless* " ROUX'S ARCH. DEVEL. BIOL. 197 : 419 - 423

DeBIANCHI, A.G., M. COUTINHO, S.D. PERIERA, O. MARINOTTI, and H.J. TARGA (1985) "Vitellogenin and vitellin of *Musca domestica*. Quantification and synthesis by fat bodies and ovaries" INSECT BIOCHEM. 15(1) : 77 - 84

De CARO, J., M. BOUDOUARD, J. BONICEL, A. GUIDONI, P. DESNUELLE, and M. ROVERY (1981) "Porcine pancreatic lipase. Completion of primary structure" BIOCHEM. BIOPHYS. ACTA. 671 : 129 - 138

DENHOLM, I., M.G. FRANCO, P.G. RUBINI, and M VECCHI (1985) " Geographical variation in house-fly (*Musca domestica* L.) sex determinants within the British Isles " GENET. RES. CAMB. 47 : 19 -27

DELIDAKIS, C., A. PREISS, D.A. HARTLEY, and S. ARTAVANIS-TSAKONAS (1991) " Two genetically and molecularly distinct functions in early neurogenesis reside within the *Enhancer of split* locus of *Drosophila melanogaster* " GENETICS 129 : 803 - 823

DESHPADNE, G., J. STUKEY, and P. SCHEDL (1995) " *scute (sis-b)* function in *Drosophila* sex determination " MOL. & CELL BIOL. 15(8) : 4430 - 4440

DiMARIO, P.J., T.G. WARREN, and A.P. MAHOWALD (1987) " The purification and in vitro phosphorylation of vitellogenin from *Drosophila melanogaster* " INSECT BIOCHEM. 17 : 1187 - 1197

DOBZHANSKY, T., and J. SCHULTZ (1931) " Evidence for multiple sex factors in the X-chromosome of *Drosophila melanogaster* " PROC. NAT. ACAD. SCI. USA 17 : 513 - 518

DÜBENDORFER, A., D. HILFIKER-KLEINER, and R. NÖTHIGER (1992) " Sex determination mechanisms in Dipteran insects : the case of *Musca domestica* " SEM. IN DEVEL. BIOL. 3 : 349 - 356

DUFFY, J.B., and J.P. GERGEN (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 & DEVEL. 5 : 2176 - 2187

DYSON, N.J. (1991) "Immobilisation of nucleic acids and hybridisation analysis" In 'Essential Molecular Biology : A practical approach (Volume II)' IRL PRESS (Ed. T. A. Brown)

ERDMAN, S.E., and K.C. BURTIS (1993) " The *Drosophila doublesex* proteins share a novel zinc finger related DNA binding domain " EMBO J. 12(2) : 527 - 535

ERICKSON, J.W., and T.W. CLINE (1991) " Molecular nature of the *Drosophila* sex determination signal and its link to neurogenesis " SCIENCE 251 : 1071 - 1074

ERICKSON, J.W., and T.W. CLINE (1993) " A bZIP protein, *sisterless-a*, collaborates with bHLH transcription factors early in *Drosophila* development to determine sex " GENES & DEVEL. 7 : 1688 - 1072

FALB, D., and T. MANIATIS (1992) " A conserved regulatory unit implicated in tissue-specific gene expression in *Drosophila* and man " GENES & DEVEL. 6 : 454 - 465

FENG, Y., N.M. SCHIFF, and D.R. CAVENER (1991) " Organ-specific patterns of gene expression in the reproductive tract of *Drosophila* are regulated by the sex-determination genes " DEVEL. BIOL. 146 : 451 - 460

FLICKINGER, T.W., and H.K. SALZ (1994) " The *Drosophila* sex determination gene *snf* encodes a nuclear protein with sequence and functional similarity to the mammalian U1A snRNP protein " GENES & DEVEL. 8 : 914 - 925

FRANCO, M.G., P.G. RUBINI, and M. VECCHI (1982) " Sex-determinants and their distribution in various populations of *Musca domestica* L. of Western Europe " GENETICAL RESEARCH 40 : 279 - 293

FRIEDRICH, E., P.A. BAEUERLE, H. GAROFF, B. HOVEMANN, and W.B. HUTTNER (1988) " Expression, tyrosine sulfation, and secretion of yolk protein 2 of *Drosophila melanogaster* in mouse fibroblasts " J. OF BIOL. CHEM. 263 (29) : 14930 - 14938

FRISCHAUF, A.M., H. LEHRACH, A. POUSTKA, and N. MURRAY (1983) "*Lambda* replacement vectors carrying polylinker sequences" J. MOL. BIOL. 170 : 827 - 842

FUJIHARA, T., M. KAWABE, and K. OISHI (1978) " A sex-transformation gene in *Drosophila melanogaster* " J. HEREDITY 69 : 229 - 236

FUKUNAGA, A., A. TANAKA, and K. OISHI (1975) " *Maleless*, a recessive, autosomal mutant of *Drosophila melanogaster* that specifically kills male zygotes " GENETICS 81 : 135 - 141

GANS, M., C. AUDIT, and M. MASSON (1975) " Isolation and characterisation of sex-linked female sterile mutants in *Drosophila melanogaster* " GENETICS 81 : 683 - 704

GARABEDIAN, M.J., M. HUNG, and P.C. WENSINK (1985) " Independent control elements that determine *yolk protein* gene expression in alternative *Drosophila* tissues " PROC. NAT. ACAD. SCI. USA 82 : 1396 - 1400

GARABEDIAN, M.J., B.M. SHEPHERD, and P.C. WENSINK (1986) " A tissue-specific transcription enhancer from the *Drosophila yolk protein 1* gene " CELL 45 : 859 - 867

GARABEDIAN, M.J., A.D. SHIRRAS, M. BOWNES, and P.C. WENSINK (1987) " The nucleotide sequence of the gene coding for *Drosophila melanogaster* yolk protein 3" GENE 55 : 1 - 8

GARCIA-BELLIDO, A. (1979) " Genetic analysis of the *achaete-scute* system of *Drosophila melanogaster* " GENETICS 91 : 491 - 520

GARCIA-BELLIDO, A. (1981) " From the gene to the pattern : chaeta differentiation " In 'Cellular controls in differentiation' C.W. LLOYD and D.E. REES, EDS (New York : Academic Press) pp 281 - 304

GELTI-DOUKA, H., GINGERAS, T.R., and M.P. KAMBYSELLIS (1974) " Yolk proteins in *Drosophila* : Identification and site of synthesis " J. EXP. ZOOL. 187 : 167 - 172

GERGEN, J.P (1987) " Dosage compensation in *Drosophila* : Evidence that *daughterless* and *Sex-lethal* control X chromosome activity at the blastoderm stage of embryogenesis " GENETICS 117 : 477 - 485

GERGEN, J.P., and B.A. BUTLER (1988) " Isolation of the segmentation gene *runt* and analysis of its expression during embryogenesis " GENES & DEVEL. 2 : 1179 - 1193

GIORGI, F., P. LUCCHESI, A. MORELLI, and M. BOWNES (1993) " Ultrastructural analysis of *Drosophila* ovarian follicles differing in yolk polypeptide (yps) composition " DEVELOPMENT 117 : 319 - 328

GORALSKI, T.J., J. EDSTROM, and B.S. BAKER (1989) " The sex-determination locus *transformer-2* of *Drosophila* encodes a polypeptide with similarity to RNA binding proteins " CELL 56 : 1011 - 1018

GORMAN, M., M.I. KURODA, and B.S. BAKER (1993) " Regulation of the sex-specific binding of the *maleless* dosage compensation protein to the male X chromosome in *Drosophila* " CELL 72 : 39 - 49

GORMAN, M., A. FRANKE, and B.S. BAKER (1995) " Molecular characterisation of the *male-specific lethal-3* gene and investigations of the regulation of dosage compensation in *Drosophila* " DEVELOPMENT 121 : 463 - 475

GRANADINO, B., S. CAMPUZANO, and L. SÁNCHEZ (1990) " The *Drosophila melanogaster fl(2)d* gene is needed for the female-specific splicing of *Sex-lethal* RNA " EMBO J. 9(8) : 2597 - 2602

GRANADINO, B., A.S. JUÁN, P. SANTAMARIA, and L. SÁNCHEZ (1992) " Evidence of a dual function in *fl(2)d*, a gene needed for *Sex-lethal* expression in *Drosophila melanogaster* " GENETICS 130 : 597 - 612

GRANADINO, B., P. SANTAMARIA, and L. SÁNCHEZ (1993) " Sex determination in the germ line of *Drosophila melanogaster* : activation of the gene *Sex-lethal* " DEVELOPMENT 118 : 813 - 816

HAGEDORN, H.H., and J.G. KUNKEL (1979) " Vitellogenin and vitellin in insects " ANN. REV. ENTOMOL. 24 : 475 - 505

HÄGELE, K. (1985) " Identification of a polytene chromosome band containing a male sex-determiner of *Chironomus thummi thummi* " CHROMOSOMA 91 : 167 - 171

HARTLEY, D.A., A. PREISS, and S. ARTAVARIS-TSAKONAS (1988) "A deduced gene product from the *Drosophila* neurogenic locus, *Enhancer of split*, shows homology to mammalian G-protein b subunit " CELL 55 : 785 - 795

HEDLEY, M.L., and T. MANIATIS (1991) " Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to *tra-2* protein in vitro " CELL 65 : 579 - 586

HEINRICHS, V., and B.S. BAKER (1995) " The *Drosophila* SR protein RBP1 contributes to the regulation of *doublesex* alternative splicing by recognising RBP1 RNA target sequences " EMBO J. 14(16) : 3987 - 4000

HILDRETH, P.E. (1965) "*Doublesex*, a recessive gene that transforms both males and females of *Drosophila* into intersexes " GENETICS 51 : 659 - 678

HILFIKER, A., and R. NÖTHIGER (1991) " The temperature sensitive mutation *vir*^{ts} *(virilizer)* identifies a new gene involved in sex determination of *Drosophila* " ROUX'S ARCH. DEVEL. BIOL. 200 : 240 - 248

HILFIKER, A., Y. YANG, D.H. HAYES, C.A. BEARD, J.E. MANNING, and J.C. LUCCHESI (1994) " Dosage compensation in *Drosophila* : the X-chromosomal binding of MSL-1 and MLE is dependent on *Sxl* activity " EMBO J. 13(15) : 3542 - 3550

HILFIKER-KLEINER, D., A. DÜBENDORFER, A. HILFIKER, and R. NÖTHIGER (1993) " Developmental analysis of two sex-determining genes, M and F, in the housefly, *Musca domestica* " GENETICS 134 : 1187 - 1194

HILFIKER-KLEINER, D., A. DÜBENDORFER, A. HILFIKER, and R. NÖTHIGER (1994) " Genetic control of sex determination in the germ line and soma of the housefly, *Musca domestica* " DEVELOPMENT 120 : 2531 - 2538

HORABIN, J.I., and P. SCHEDL (1993a) " Regulated splicing of the *Drosophila Sex-lethal* male exon involves a blockage mechanism " MOL. & CELL. BIOL. 13(3) : 1408 - 1414

HORABIN, J.L., and P. SCHEDL (1993b) " *Sex-lethal* autoregulation requires multiple *cis*-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site " MOL. & CELL. BIOL. 13(12) : 7734 - 7746

HORSFALL, W.R., and J.F. ANDERSON (1963) " Thermally induced genital appendages in mosquitoes " SCIENCE 141 : 1183

HOSHIJIMA, K., K. INOUE, I. HIGUCHI, H. SAKAMOTO, and Y. SHIMURA (1991) " Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila* " SCIENCE 252 : 833 - 836

HUNG, M., T. BARNETT, C. WOOLFORD, and P.C. WENSINK (1982) " Transcript maps of the *Drosophila yolk protein* genes " J. MOL. BIOL. 154 : 581 - 602

HUNG, M., and P.C. WENSINK (1983) " Sequence and structure conservation in yolk proteins and their genes " J. MOL. BIOL. 164 : 481 - 489

HUTTNER, W.B., and P.A. BAEUERLE (1988) "Protein sulfation on tyrosine" MODERN CELL. BIOL. 6 : 97 - 140

INOUE, K., K. HOSHIJIMA, H. SAKAMOTO, and Y. SHIMURA (1990) " Binding of the *Drosophila Sex-lethal* gene product to the alternative splice site of *transformer* primary transcript " NATURE 344 : 461 - 463

INOUE, H., and T. HIROYOSHI (1986) " A maternal-effect sex-transformation mutant of the housefly, *Musca domestica* L. " GENETICS 112 : 469 - 482

ISAAC, P.G., and M. BOWNES (1982) " Ovarian and fat-body vitellogenin synthesis in *Drosophila melanogaster* " EUR. J. BIOCHEM. 123 : 527 - 534

JOWETT, T., and J.H. POSTLETHWAIT (1980) " The regulation of yolk polypeptide synthesis in *Drosophila* ovaries and fat body by 20-hydroxyecdysone and a juvenile hormone analogue " DEVEL. BIOL. 80 : 225 - 234

JURSNICH, V.A., and K.C. BURTIS (1993) " A positive role in differentiation for the male *doublesex* protein of *Drosophila* " DEVEL. BIOL. 155 : 235 - 249

KANIA, M.A., A.S. BONNER, J.B. DUFFY, and J.P. GERGEN (1990) " The *Drosophila* segmentation gene *runt* encodes a novel nuclear regulatory protein that is also expressed in the developing nervous system " GENES & DEVEL. 4 : 1701 - 1703

KELLEY, R.L., I. SOLOVYEVA, L.M. LYMAN, R. RICHMAN, V. SOLOVYEV, and M.I. KURODA (1995) " Expression of *msl-2* causes assembly of dosage compensation regulators on the X-chromosome and female lethality in *Drosophila* " CELL 81 : 867 - 877

KELLY, T.J., T.S. ADAMS, M.B. SCHWARTZ, M.J. BIRNBAUM, E.C. RUBENSTEIN, and R.B. IMBERSKI (1987) "Juvenile hormone and ovarian maturation in the Diptera : A review of recent results" INSECT BIOCHEM. 17(7) : 1089 - 1093

KEYES, L.N., T.W. CLINE, and P. SCHEDL (1992) " The primary sex determination signal of *Drosophila* acts at the level of transcription " CELL 68 : 933 - 943

KING, R.C. (1959) In *Drosophila Inf. Serv.* 33 : 142 - 143

KOMAROMY, M.C., and M.C. SCHOTZ (1987) "Cloning of rat hepatic lipase cDNA : evidence for a lipase gene family" PROC. NATL. ACAD. SCI. USA. 84: 1626-1650

KOZMA, R., and M. BOWNES (1986) "Yolk protein induction in males of several *Drosophila* species " INSECT BIOCHEM. 16 : 263 - 271

KURODA, M.I., M.J. KERNAN, R. KREBER, B. GANETZKY, and B.S. BAKER (1991) " The *maleless* protein associates with the X chromosome to regulate dosage compensation in *Drosophila* " CELL 66 : 935 - 947

LIDDELL, S., and M. BOWNES (1991a) " Characterisation, molecular cloning and sequencing of $yp3^{sl}$, a fertile yolk protein 3 mutant in *Drosophila* " MOL. GEN. GENET. 228 : 81 - 88

LIDDELL, S., and M. BOWNES (1991b) " Investigation of *cis*-acting sequences regulating expression of the gene encoding yolk protein 3 in *Drosophila melanogaster* " MOL. GEN. GENET. 230 : 219 - 224

LINDSLEY, D., and E. GRELL (1968) "Genetic variations of *Drosophila melanogaster*" Carnegie Institute of Washington, Publication no. 627

LIU, Y., and J.M. BELOTE (1995) " Protein-protein interactions among components of the *Drosophila* primary sex determination signal " MOL. GEN. GENET. 248 : 182 - 189

LOGAN, S.K., M.J. GARABEDIAN, and P.C. WENSINK (1989) " DNA regions that regulate the ovarian transcriptional specificity of *Drosophila yolk protein* genes " GENES & DEVEL. 3 : 1453 - 1461

LOGAN, S.K., and P.C. WENSINK (1990) " Ovarian follicle cell enhancers from the *Drosophila* yolk protein genes : different segments of one enhancer have different cell-type specificitys that interact to give normal expression " GENES & DEVEL. 4 : 613 - 623

LUCCHESI, J.C., and T. SKRIPSKY (1981) " The link between dosage compensation and sex differentiation in *Drosophila melanogaster* " BIOCHEM. GENET. 9 : 41 - 51

MAINE, E.M., H.K. SALZ, P. SCHEDL, and T.W. CLINE (1985a) " *Sex-lethal*, a link between sex determination and sexual differentiation in *Drosophila melanogaster* " COLD SPRING HARB. SYMP. QUANT. BIOL. 50 : 595 - 604

MAINE, E.M., H.K. SALZ, T.W. CLINE, and P. SCHEDL (1985b) " The *Sex-lethal* gene of *Drosophila* : DNA alterations associated with sex-specific lethal mutations " CELL 53 : 521 - 529

MARSH, J.L., and E. WIESCHAUS (1978) " Is sex determination in germ line and soma controlled by separate genetic mechanisms ? " NATURE 272 : 249 - 251

MARSHALL, T., and J.R.S. WHITTLE (1978) " Genetic analysis of the mutation *female-lethal* in *Drosophila melanogaster* " GENET. RES. 32 : 103 - 111

MARTINEZ, A. (1991) "The yolk proteins of *Drosophila* are conserved through Dipteran evolution" PhD Thesis, The University of Edinburgh

MARTINEZ, A., and M. BOWNES (1992) "The specificity of yolk protein uptake in the Cyclorrhaphan Diptera is conserved through evolution" J. MOL. EVOL. 35 : 444 - 453

MARTINEZ, A., and M. BOWNES (1994) "The sequence and expression pattern of the *Calliphora erythrocephala yolk protein A* and *B* genes" J. MOL. EVOL. 38 : 336 - 351

MASON, J.M. (1973) " A relationship between *daughterless* and the Y chromosome " In *Drosophila Inf. Serv.* 50 : 93

MARTIN, J., and B.T.O. LEE (1984) " A phylogenetic study of sex determiner location in a group of Australasian *Chironomus* species (Diptera, Chironimidae) " CHROMOSOMA 90 : 190 - 197

MATTOX, W., M.J. PALMER, and B.S. BAKER (1990) " Alternative splicing of the sex determination gene *transformer-2* is sex-specific in the germ line but not in the soma " GENES & DEVEL. 4 : 789 - 805

MATTOX, W., and B.S. BAKER (1991) " Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila* " GENES & DEVEL. 5 : 786 - 796

McDONALD, I.C., P. EVENSON, C.A. NICKEL, and O.A. JOHNSON (1978) " House fly genetics : Isolation of a female determining factor on chromosome 4 " ANN. REV. ENTOMOL. SOC. AMER. 71 : 692 - 694

McKEARIN, D.M., and A.C. SPRADLING (1990) " *bag of marbles* : a *Drosophila* gene required to initiate both male and female gametogenesis " GENES & DEVEL. 4 : 2242 - 2251

McKEOWN, M., J.M. BELOTE, and B.S. BAKER (1987) " A molecular analysis of *transformer*, a gene in *Drosophila melanogaster* that controls female sexual differentiation " CELL 48 : 489 - 499

McKEOWN, M., J.M. BELOTE, and R.T. BOGGS (1988) " Ectopic expression of the female *transformer* gene product leads to female differentiation of chromosomally male *Drosophila* " CELL 53 : 887 - 895

McROBERT, S.P., and L. TOMPKINS (1985) " The effect of *transformer, doublesex* and *intersex* mutations on the sexual behaviour of *Drosophila melanogaster* " GENETICS 111 : 89 - 96

MEVEL-NINIO, M., R. TERRACOL, and F.C. KAFATOS (1991) " The *ovo* gene of *Drosophila* encodes a zinc finger protein required for female germ line development " EMBO J. 10(8) : 2259 - 2266

MINOO, P., and J. POSTLETHWAIT (1985) " Biosynthesis of *Drosophila* yolk polypeptides " ARCH. INSECT BIOCHEM. PHYSIOL. 2 : 7 - 27

MORRISON, P.E., and M.D. DAVIES (1964) "Feeding of dry, chemically defined diets, and egg production in the adult house-fly" NATURE 201 : 104 - 105

MULLER, H.J (1955) " On the relationship between chromosome changes and gene mutations " BROOKHAVEN SYMP. 8 : 126 - 147

MULLER, H.J., and S. ZIMMERING (1960) " A sex linked lethal without evident effect in *Drosophila* males but partially dominant in females " GENETICS 45 : 1001 - 1002

MULLIGAN, P.K., J.D. MOHLER, and L.J. KALFAYAN (1988) " Molecular localisation and developmental expression of the *otu* locus of *Drosophila melanogaster* " MOL. & CELL. BIOL. 8 : 1481 - 1488

MURRE, C., P.S. McCAW, and D. BALTIMORE (1989a) " A new DNA binding and dimerisation motif in immunoglobin enhancer binding, *daughterless, MyoD,* and *myc* proteins " CELL 56 : 777 - 783

MURRE, C., P.S. McCAW, H. VAESSIN, M. CAUDY, L.Y. JAN, Y.N. JAN, C.V. CABRERA, J.N. BUSKIN, S.D. HAUSCHKA, A.B. LASSAR, H. WEINTRAUB, and D. BALTIMORE (1989b) " Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence " CELL 58 : 537 - 544

NAGOSHI, R.N., M. McKEOWN, K.C. BURTIS, J.M. BELOTE, and B.S. BAKER (1988) " The control of alternative splicing at genes regulating sexual differentiation in *Drosophila melanogaster* " CELL 53 : 229 - 236

NAGOSHI, R.N., and B.S. BAKER (1990) " Regulation of sex-specific RNA splicing at the *Drosophila doublesex* gene; *cis*-acting mutations in exon sequences alter sex-specific RNA splicing patterns " GENES & DEVEL. 4 : 89 - 97

NAGOSHI, R.N., J.S. PATTON, E. BAE, and P.K. GEYER (1995) " The somatic sex determines the requirement for *ovarian tumor* gene activity in the proliferation of the *Drosophila* germline " DEVELOPMENT 121 : 579 - 587

NÖTHIGER, R., and M. STEINMANN-ZWICKY (1985) " A single principle for sex determination in insects " COLD SPRING HARB. SYMP. 50 : 615 - 621

NÖTHIGER, R., M. LEUTHOLD, N. ANDERSON, P. GERSCHWILER, A. GRÜTER, W. KELLER, C. LEIST, M. ROOST, and H. SCHMID (1987) " Genetic and developmental analysis of the sex-determining gene *'doublesex' (dsx)* of *Drosophila melanogaster* " GENET. RES. 50 : 113 - 123

NÖTHIGER, R., M. JONGLEZ, M. LEUTHOLD, P. MEIER-GERSCHWILER, and T. WEBER (1989) " Sex determination in the germ line of *Drosophila* depends on genetic signals and inductive somatic factors " DEVELOPMENT 107 : 505 - 518

OLIVER, B., N. PERRIMON, and A.P. MAHOWALD (1987) " The *ovo* locus is required for sex-specific germ line maintenance in *Drosophila* " GENES & DEVEL. 1 : 913 - 923

OLIVER, B., N. PERRIMON, and A.P. MAHOWALD (1988) " Genetic evidence that the *sans-fille* locus is involved in *Drosophila* sex determination " GENETICS 120 : 159 - 171

OLIVER, B., D. PAULI, and A.P. MAHOWALD (1990) " Genetic evidence that the *ovo* locus is involved in *Drosophila* germ line sex determination " GENETICS 125 : 535 - 550

OLIVER, B., Y. KIM, and B.S. BAKER (1993) " *Sex-lethal*, master and slave : a hierarchy of germ-line sex determination in *Drosophila* " DEVELOPMENT 119 : 897 - 908

OLIVER, B., J. SINGER, V. LAGET, G. PENNETTA, and D. PAULI (1994) " Function of *Drosophila ovo*$^+$ in germ-line sex determination depends on X-chromosome number " DEVELOPMENT 120 : 3185 - 3195

PALMER, M.J., V.A. MERGNER, R. RICHMAN, J.E. MANNING, M.I. KURODA, and J.C. LUCCHESI (1993) " The *male-specific lethal-one (msl-1)* gene of *Drosophila melanogaster* encodes a novel protein that associates with the X chromosome in males " GENETICS 134 : 545 - 557

PALMER, M.J., R. RICHMAN, L. RICHTER, and M.I. KURODA (1994) " Sex-specific regulation of the *male-specific lethal-1* dosage compensation gene in *Drosophila* " GENES & DEVEL. 8 : 698 - 706

PARKHURST, S.M., H.D. LIPSHITZ, and D. ISH-HOROWICZ (1993) " *achaete-scute* feminising activities and *Drosophila* sex determination " DEVELOPMENT 117 : 737 - 749

PAROUSH, Z., R.L. FINLEY Jr., T. KIDD, S.M. WAINWRIGHT, P.W. INGHAM, R. BRENT, and D. ISH-HOROWICZ (1994) " *Groucho* is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts with hairy-related bHLH proteins " CELL 79 : 805 - 815

PAULI, D., and A.P. MAHOWALD (1990) " Germ-line sex determination in *Drosophila melanogaster* " T.I.G. 6(8) : 259 - 264

PAULI, D., B. OLIVER, and A.P. MAHOWALD (1993) " The role of the *ovarian tumor* locus in *Drosophila melanogaster* germ line sex determination " DEVELOPMENT 119 : 123 - 134

PAULI, D., B. OLIVER, and A.P. MAHOWALD (1995) " Identification of regions interacting with *ovo*^{D} mutations : Potential new genes involved in germline sex determination or differentiation in *Drosophila melanogaster* " GENETICS 139 : 713 - 732

PERLMAN, D., and H.O. HALVORSEN (1983) "A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides" J. MOL. BIOL. 167 : 391 - 401

PERSSON, B., G. BENGTSSON-OLIVECRONA, S. ENERBACK, T. OLIVECRONA, and H. JORNVALL (1989) " Structural features of lipoprotein lipase. Lipase family relationships, binding interactions, non-equivalence of lipase cofactors, vitellogenin similarities and functional subdivision of lipoprotein lipase " EUROP. J. BIOCHEM. 179 : 39 - 45

PERSSON, B., J. HORNVALL, T. OLIVECRONA, and G. BENGTSSON-OLIVECRONA (1991) " Lipoprotein lipases and vitellogenins in relation to the known three dimensional structure of pancreatic lipase " FEBS LETT. 288 : 33 - 36

POIRÉ, M., E. NIEDERER, and M. STEINMANN-ZWICKY (1995) " A sex-specific number of germ cells in embryonic gonads of *Drosophila* " DEVELOPMENT 121 : 1867 - 1873

POSTLETHWAIT, J.H., and R. KASCHNITZ (1978) " The synthesis of *Drosophila melanogaster* vitellogenins *in-vivo*, in culture, and in a cell-free translation system " FEBS LETTERS 95(2) : 247 - 251

POSTLETHWAIT, J.H., M. BOWNES, and T. JOWETT (1980) " Sexual phenotype and vitellogenin synthesis in *Drosophila melanogaster* " DEVEL. BIOL. 79 : 379 - 387

PROUDFOOT, N. (1991) " Poly(A) Signals" CELL 64 : 671 - 674

PULTZ, M.A., G.S. CARSON, and B.S. BAKER (1994) " A genetic analysis of *hermaphrodite*, a pleiotropic sex determination gene in *Drosophila melanogaster* " GENETICS 70 : 261 - 274

PULTZ, M.A., and B.S. BAKER (1995) " The dual role of *hermaphrodite* in the *Drosophila* sex determination regulatory hierarchy " DEVELOPMENT 121 : 99 - 111

REDFIELD, H. (1926) " The maternal inheritance of a sex-limited lethal effect in *Drosophila melanogaster* " GENETICS 11 : 482 - 502

RINA, M., and C. SAVAKIS (1991) "A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata* : sequence and structural conservation in the Dipteran yolk proteins and their genes" GENETICS 127 : 369 - 380

RONALDSON, E., and M. BOWNES (1995) " Two independent *cis*-acting elements regulate the sex- and tissue-specific expression of *yp3* in *Drosophila melanogaster* " GENET. RES. CAMB. 66 : 9 - 17

RYNER, L.C., and B.S. BAKER (1991) " Regulation of *doublesex* pre-mRNA processing occurs by 3'-splice site activation " GENES & DEVEL. 5 : 2071 - 2085

SÁNCHEZ, L., and R. NÖTHIGER (1982) " Clonal analysis of *Sxl*, a gene needed for female sexual development in *Drosophila melanogaster* " ROUX'S ARCH. DEVEL. BIOL 191 : 211 - 214

SAKAMOTO, H., K. INOUE, I. HIGUCHI, Y. ONO, and Y. SHIMURA (1992) " Control of *Drosophila Sex-lethal* pre-mRNA splicing by its own female-specific product " NUCLEIC ACIDS RES. 20(21) : 5563 - 5540

SALZ, H.K., E.M. MAINE, L.N. KEYES, M.E. SAMUELS, T.W. CLINE, and P. SCHEDL (1989) " The *Drosophila* female-specific sex-determination gene, *Sex-lethal*, has stage-, tissue- and sex-specific RNA's suggesting multiple modes of regulation " GENES & DEVEL. 3 : 708 - 719

SALZ, H.K. (1992) " The genetic analysis of *snf* : A *Drosophila* sex determination gene required for activation of *Sex-lethal* in both the germ line and the soma " GENETICS 130 : 547 - 554

SAMUELS, M.E., D. BOPP, R.A. COLVIN, R.F. ROSCIGNO, M.A. GARCIA-BELLIDO, and P. SCHEDL (1994) " RNA binding by *Sxl* proteins *in-vitro* and *in-vivo* " MOL. & CELL. BIOL. 14(7) : 4975 - 4990

SANDLER, L. (1972) " On the genetic control of genes located in the sex chromosome heterochromatin of *Drosophila melanogaster* " GENETICS 70 : 261 - 274

SASS, G.L., A.R. COMER, and L.L. SEARLES (1995) " The ovarian tumor protein isoforms of *Drosophila melanogaster* exhibit differences in function, expression, and localisation " DEVEL. BIOL. 167 : 201 - 212

SCHRONS, H., E. KNUST, and J.A. CAMPOS-ORTEGA (1992) " The *Enhancer of split* complex and adjacent genes in the 96F region of *Drosophila melanogaster* are required for segregation of neural and epidermal progenitor cells " GENETICS 132 : 481 - 503

SCHÜPBACH, T. (1982) " Autosomal mutations that interfere with sex determination in somatic cells of *Drosophila* have no direct effect on the germline " DEVEL. BIOL. 89 : 117 - 127

SCHÜPBACH, T. (1985) " Normal female germ cell differentiation requires the female X chromosome to autosome ratio and expression of *Sex-lethal* in *Drosophila melanogaster* " GENETICS 109 : 529 - 548

SCHÜPBACH, T., and E. WIESCHAUS (1989) " Female sterile mutations on the second chromosome of *Drosophila melanogaster*. I. Maternal effect mutations " GENETICS 121 : 101 - 117

SHIRRAS, A.D., and M. BOWNES (1987) " Separate DNA sequences are required for normal female and ecdysone-induced male expression of *Drosophila melanogaster* yolk protein 1" MOL. GEN. GENET. 210 : 153 - 155

SHORT, J.M., J.M. FERNANDEZ, J.A. SORGE, and W.D. HUSE (1988) "λ ZAP : a bacteriophage λ expression vector with *in vivo* excision properties" NUCLEIC ACIDS RESEARCH 16(15) : 7583 - 7600

SLEE, R., and M. BOWNES (1990) " Sex determination in *Drosophila melanogaster* " QUART. REV. BIOL. 65(2) : 175 - 204

SMITH, C.W.J., J.G. PATTON, and B. NADAL-GINARD (1989) " Alternative splicing in the control of gene expression " ANN. REV. GENET. 23 : 527 - 577

SØNDERGAARD, L., D. MAUCHLINE, P. EGETOFT, N. WHITE, P. WULF, and M. BOWNES (1995) " Nutritional response in a *Drosophila yolk protein* gene promoter" MOL. GEN. GENET. 248 : 25 - 32

SOSNOWSKI, B.A., J.M. BELOTE, and M. McKEOWN (1989) " Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage " CELL 58 : 449 - 459

SOSNOWSKI, B.A., D.D. DAVIS, R.T. BOGGS, S.J. MADIGAN, and M. McKEOWN (1994) " Multiple portions of a small region of the *Drosophila transformer* gene are required for efficient *in-vivo* sex-specific regulated RNA splicing and *in-vitro* Sex-lethal binding " DEVEL. BIOL. 161 : 302 - 312

SOUTHERN, E.M. (1975) "Detection of specific sequences among DNA fragments separated by gel electrophoresis" MOL. BIOL. 98 : 503 - 517

STEINHAUER, W.R., R.C. WALSH, and L.J. KALFAYAN (1989) " Sequence and structure of the *Drosophila melanogaster ovarian tumor* gene and generation of an antibody specific for ovarian tumor protein" MOL. & CELL. BIOL. 9(12) : 5726 - 5732

STEINHAUER, W.R., and L.J. KALFAYAN (1992) " A specific *ovarian tumor* protein isoform is required for efficient differentiation of germ cells in *Drosophila* oogenesis " GENES & DEVEL. 6 : 233 - 243

STEINMANN-ZWICKY, M., H. SCHMID, and R. NÖTHIGER (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. (1988) " Sex determination in *Drosophila* : the X-chromosomal gene *liz* is required for *Sxl* activity " EMBO J. 7(12) : 3889 - 3898

STEINMANN-ZWICKY, M. (1992) " Sex determination of *Drosophila* germ cells " SEM. IN DEVEL. BIOL. 3 : 341 - 347

STEINMANN-ZWICKY, M. (1994) " Sex determination of the *Drosophila* germ line : *tra* and *dsx* control somatic inductive signals " DEVELOPMENT 120 : 707 - 716

STEPHEN, D., C. JONES, and P.J. SCHOFIELD (1990) "A rapid method for isolating high quality plasmid DNA suitable for sequencing" NUCLEIC ACIDS RESEARCH 18(24) : 7463 - 7464

STURTEVANT, A.H. (1921) " The North American species of *Drosophila* " CARNEGIE INST. WASH. YEARBOOK 301

STURTEVANT, A.H. (1945) " A gene in *Drosophila melanogaster* that transforms females into males " GENETICS 30 : 297 - 299

TAMURA, T., C. KUNERT, and J. POSTLETHWAIT (1985) " Sex- and cell-specific regulation of yolk polypeptide genes introduced into *Drosophila* by P-element-mediated gene transfer " PROC. NATL. ACAD. SCI. USA. 82 : 7000 - 7004

TAUTZ, D., and C. PFEIFFLE (1989) "A nonradioactive *in situ* hybridisation method for the localisation of specific RNA's in *Drosophila* embryos reveals a translational control of the segmentation gene *hunchback*" CHROMOSOMA (Berl.) 98: 81 - 85

TAYLOR, B.J. (1992) " Differentiation of a male-specific muscle in *Drosophila melanogaster* does not require the sex-determining genes *doublesex* or *intersex* " GENETICS 132 : 179 - 191

TAYLOR, B.J., and J.W. TRUMAN (1992) " Commitment of abdominal neuroblasts in *Drosophila* to a male or female fate is dependent on genes of the sex-determination hierarchy " DEVELOPMENT 114 : 625 - 642

THOMAS, P.S. (1980) "Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose" PROC. NATL. ACAD. SCI. U.S.A. 77 : 5201 - 5205

THOMPSON, P.E., and J.S. BOWEN (1972) " Interactions of differentiated primary sex factors in *Chironomus tentans* " GENETICS 70 : 491 - 493

TIAN, M., and T. MANIATIS (1993) " A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA " CELL 74 : 105 - 114

TOMPKINS, L., and S.P. McROBERTS (1989) " Regulation of behavioural and pheromonal aspects of sex determination in *Drosophila melanogaster* by the *Sex-lethal* gene " GENETICS 123 : 535 - 542

TORRES, M., and L. SÁNCHEZ (1989) "The *scute* (T4) gene acts as a numerator element of the X:A signal that determines the state of activity of *Sex-lethal* in *Drosophila* " EMBO J. 8(10) : 3079 - 3086

TORTIGLIONE, C., and M. BOWNES (SUBMITTED) "Conservation of tissue- but not sex-specific gene regulation in the Diptera"

TOWNER, P. (1991) "Purification of DNA" In 'Essential Molecular Biology - A Practical Approach, Volume I' IRL Press (Ed. T.A. Brown)

TURNER, B.M., A.J. BIRLEY, and J. LAVENDER (1992) " Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei " CELL 69 : 375 - 384

ULLERICH, F. (1984) " Analysis of sex determination in the monogenic blowfly *Chrysomya rufifacies* by pole cell transplantation " MOL. GEN. GENET. 193 : 479 - 487

VALCÁREL, J., R. SINGH, P.D. ZAMORE, and M.R. GREEN (1993) " The protein *Sex-lethal* antagonises the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA " NATURE 362 : 171 - 175

VAN DEUSEN, E.B. (1976) " Sex determination in germ line chimeras of *Drosophila melanogaster* " J. EMBRYOL. EXP. MORPH. 37 : 173 - 185

VANOSSI ESTE, S., and C. RAVOTI (1982) " Inheritance of the arrhenogenic factor *Ag* of *Musca domestica* " BOLL. ZOOL. 49 : 269 - 278

VILLARES, R., and C.V. CABRERA (1987) " The *achaete-scute* gene complex of *D. melanogaster* : Conserved domains in a subset of genes required for neurogenesis and their homology to *myc* " CELL 50 : 415 - 424

VILLAROYA, A., and E. JUAN (1991) "ADH and phylogenetic relationships of *Drosophila lebanonesis* (Scaptodrosophila) " J. MOL. EVOL. 32 : 421 - 428

VINSON, C.R., P.B. SIGLER, and S.L. McKNIGHT (1989) " Scissor-grip model for DNA recognition by a family of Leucine zipper proteins " SCIENCE 246 : 911 - 916

Von HEIJNE, G. (1983) "Patterns of amino acids near signal-sequence cleavage sites" EUR. J. BIOCHEM. 133 : 17 - 21

WANG, J., and L.R. BELL (1994) " The *Sex-lethal* amino terminus mediates co-operative interactions in RNA binding and is essential for splicing regulation " GENES & DEVEL. 8 : 2072 - 2085

WAGONER, D.E. (1969) " Presence of male determining factors found on three autosomes in the housefly, *Musca domestica* L. " NATURE 233 : 187 - 188

WATANABE, T.K. (1975) " A new sex-transforming gene on the second chromosome of *Drosophila melanogaster* " JAP. J. GENET. 50 : 269 - 271

WEI, G., B. OLIVER, and A.P. MAHOWALD (1991) " Gonadal dysgenesis reveals sexual dimorphism in the embryonic germline of *Drosophila* " GENETICS 129 : 203 - 210

WENT, D.F., and R. CAMEZIND (1980) " Sex determination in the Dipteran insect *heteropeza pygmae* " GENETICA 52/53 : 373 - 377

WIESCHAUS, E., and R. NÖTHIGER (1982) " The role of the *transformer* gene in the development of the genitalia and analia of *Drosophila melanogaster* " DEVEL. BIOL. 90 : 320 - 334

WILLIAMS, J.L., and M. BOWNES (1986) " Reduced stability of RNA coding for yolk polypeptide 3 in *Drosophila melanogaster* ovary" EUR. J. BIOCHEM. 161 : 95 - 101

WILSON, T.G. (1982) " A correlation between juvenile hormone deficiency and vitellogenic oocyte degeneration in *Drosophila melanogaster* " ROUX'S ARCH. ENTW. MEC. ORG. 1919 : 257 - 263

YAN, Y.L., C.J. KUNERT, and J.H. POSTLETHWAIT (1987) " Sequence homologies among the three yolk polypeptide (*Yp*) genes in *Drosophila melanogaster* " NUCL. ACIDS RES. 15(1) : 67 - 85

YAN, Y.L., and J.H. POSTLETHWAIT (1990) " Vitellogenesis in *Drosophila* : Sequestration of a yolk polypeptide/ invertase fusion protein into developing oocytes " DEVEL. BIOL. 140 : 281 - 290

YOUNG, R.A., and R.W. DAVIS (1983) "Efficient isolation of genes by using antibody probes" PROC. NATL. ACAD. SCI. U.S.A. 80 : 1194 - 1198

YOUNGER-SHEPHERD, S., H. VAESSIN, E. BIER, L.Y. JAN, and Y.N. JAN (1992) " *deadpan*, an essential pan-neural gene encoding an HLH protein, acts as a denominator in *Drosophila* sex determination " CELL 70 : 911 - 922

ZHOU, S., Y. YANG, M.J. SCOTT, A. PANNUTI, K.C. FEHR, A. EISEN, E.V. KOONIN, D.L. FOUTS, R. WRIGHTSMAN, J.E. MANNING, and J.C. LUCCHESI (1995) " *Male-specific lethal-2*, a dosage compensation gene of *Drosophila*, undergoes

sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster " EMBO J. 14(12) : 2884 - 2895