CLONING AND CHARACTERISATION OF A GENE
ENCODING SEX-SPECIFIC TRANSCRIPTS IN DROSOPHILA
MELANOGASTER

Colin N. MacDougall

THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
The University of Edinburgh
1996
DECLARATION

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

Colin N. MacDougall.
July 1996
Sex determination of somatic tissue in *Drosophila melanogaster* is mediated by a hierarchy of regulatory genes controlled by the gene *Sex-lethal* (*Sxl*). Male differentiation is the default state, with *Sxl* being activated only in females. The Sxl protein is a RNA splicing modulator which regulates the splicing of both the *Sxl* transcript and the transcript of the sex-transforming gene *transformer* (*tra*). In the absence of Sxl protein, the *Sxl* and *tra* transcripts are spliced to produce transcripts incapable of producing active protein. Thus, active Tra protein is produced only in females, where it acts in concert with the product of the *transformer-2* (*tra-2*) gene to modulate splicing of the transcript from the *doublesex* (*dsx*) gene. In this way, sex-specific Dsx proteins are produced (Dsx^M and Dsx^F) which are differentially active, transcriptionally regulating sex differentiation genes, such as the *yolk protein* (*yp*) genes. Although the genes involved in sex determination have been well characterised, little is known about the genetic factors which play a part in sex-specific differentiation of the determined state. This thesis describes the cloning and characterisation of a gene which is a good candidate for a sex differentiation gene.

In an attempt to identify genes involved in differentiation of sex-specific characteristics, a non-gonadal differential screen was performed. Sex-specific, radiolabelled cDNA was used to screen a bacteriophage λ genomic DNA library and a genomic recombinant, fs(1), was selected for further analysis on the basis of preferential hybridisation to female non-gonadal cDNA. *In situ* hybridisation to *Drosophila* 3rd instar larva salivary gland polytene chromosomes indicates that the genomic DNA contained within fs(1) is located distally, on the left arm of chromosome 3, at position 061C1-3.

Two cDNAs, a 4.5kb cDNA (cDNA11) and a 3.0kb cDNA (cDNAa), were isolated on the basis of hybridisation to sequences contained within fs(1). Both cDNAs were fully sequenced and found to encode a novel OPA-repeat-containing
serine/threonine-specific protein kinase. cDNAa and cDNA11 both contain the entire open reading frame (ORF) which encodes this predicted protein, and differ only in untranslated regions. The cDNAa ORF was subcloned into a fusion-protein expression vector and fusion protein was successfully expressed in bacterial cells, as shown by Western blot analysis using antibodies specific to the vector-derived fusion protein. Genomic DNA, containing the entire cDNA11 and cDNAa transcription units, was isolated. The precise intron/exon structure of both cDNAs was determined by Southern blotting, DNA sequencing and PCR analysis.

cDNAa hybridises with four transcripts on Northern blots; a 3.0kb testis-specific, 3.5kb ovary-specific, 4.5kb female carcass-specific and a common transcript of around 4.7kb. The cDNA11-specific 3' UTR hybridises with the 4.5kb and common transcripts, but not with the 3.5kb or 3.0kb transcripts. Thus, cDNAa and cDNA11 are likely to represent the testis-specific and female carcass-specific transcripts, respectively. Both cDNAs contain translational control elements which are found in transcripts under male germline-specific translational control. The presence of these elements, together with Northern blot and whole-mount testis *in situ* hybridisation evidence, suggests that the testis-specific transcript is in fact germline-specific. Both cDNAs also contain *d sx*-like 13-nucleotide repeat elements which are required for Tra/Tra-2-mediated *d sx* splicing regulation. This, together with Northern blot evidence using *tra* and *tra-2* mutant flies, suggests that production of the female carcass-specific transcript is under direct control of *tra* and *tra-2*. Production of the testis-specific transcript is *tra/tra-2*-independent, as shown by Northern blots using *tra* and *tra-2* mutant flies and phenotypic analysis of these mutants.

The implications of these findings, and the possible functions of the protein encoded by cDNAa and cDNA11 are discussed.
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Mary Bownes, for her help and guidance throughout the project. A heartfelt thank you to my fellow lab workers; Diane, Dot, Neil, Claudia, Angela (for her irrepressible happiness), Bryce (for his encyclopaedic knowledge), Debbie and Janis (for impromptu counselling), Elaine, Kathleen (for looking after my flies so well), Roger (for TVR discussions), Simone, Craig, Debiao and Wu-min. Thank you also, to all the ladies involved in media preparation and washing-up.

Above all, I want to thank my parents, Neil and Johanna, and my sister, Marie. Without their unfailing and unconditional support (both financial and emotional), I would never have made it through all the ups and downs of the last four years. This thesis is dedicated to them.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine-5'-triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>°C</td>
<td>Degrees centigrade</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>Ci</td>
<td>Curies</td>
</tr>
<tr>
<td>cm</td>
<td>centimetre(s)</td>
</tr>
<tr>
<td>(d)dATP</td>
<td>2' (3'-di) Deoxyadenosine-5'-triphosphate</td>
</tr>
<tr>
<td>(d)dCTP</td>
<td>2' (3'-di) Deoxycytosine-5'-triphosphate</td>
</tr>
<tr>
<td>(d)dGTP</td>
<td>2' (3'-di) Deoxyguanosine-5'-triphosphate</td>
</tr>
<tr>
<td>(d)dTTP</td>
<td>2' (3'-di) Deoxythymidine-5'-triphosphate</td>
</tr>
<tr>
<td>dUTP</td>
<td>2' Deoxyuridine-5'-triphosphate</td>
</tr>
<tr>
<td>UTP</td>
<td>Uridine-5'-triphosphate</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleotide-5'-triphosphate</td>
</tr>
<tr>
<td>dH₂O</td>
<td>Distilled water</td>
</tr>
<tr>
<td>DEAE</td>
<td>Diethylaminoethyl</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNAase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Diaminoethanetetra-acetic acid</td>
</tr>
<tr>
<td>FSB</td>
<td>Formaldehyde sample buffer</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione-S-transferase</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactoside</td>
</tr>
<tr>
<td>kDa</td>
<td>KiloDalton(s)</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase(s)</td>
</tr>
<tr>
<td>pfu</td>
<td>Plaque forming units</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>RACE</td>
<td>Rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>pH</td>
<td>-Log10 (hydrogen ion concentration)</td>
</tr>
<tr>
<td>PolyA⁺ RNA</td>
<td>Polyadenylated ribonucleic acid</td>
</tr>
<tr>
<td>PMSF</td>
<td>Phenylmethylsulphonyl fluoride</td>
</tr>
<tr>
<td>psi</td>
<td>Pounds per square inch</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAse</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>³⁵S</td>
<td>β-emitting isotope of sulphur</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SDS PAGE</td>
<td>Sodium dodecyl sulphate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>TEMED</td>
<td>NNN'N'-tetra-methyl-1,2-diamino-ethane</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)-amino-methane</td>
</tr>
<tr>
<td>Triton X-100</td>
<td>Octylphenoxypolyethoxyethanol</td>
</tr>
<tr>
<td>Tween-20</td>
<td>Polyoxyethylene sorbitan monolaurate</td>
</tr>
<tr>
<td>TCA</td>
<td>Trichloroacetic acid</td>
</tr>
<tr>
<td>U</td>
<td>Unit(s)</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>μCi</td>
<td>Microcurie(s)</td>
</tr>
</tbody>
</table>
ug  Microgram(s)
ul  Microlitre(s)
µM  Micromolar
µmol  Micromole(s)
V  Volt(s)
v/v  Volume per volume
w/v  Weight per volume
Klenow  Large fragment of DNA polymerase I
Krpm  Kilorevolutions per minute
λ  Lambda
L  litre
M  Molar
mA  Milliampere(s)
mCi  Millicuries
mg  Milligram(s)
ml  Millilitre(s)
mm  Millimetre(s)
mM  Millimolar
mmol  Millimole(s)
min  Minute(s)
MOPS  Morpholinopropanesulphonic acid
mRNA  Messenger ribonucleic acid
MWt  Molecular weight
ng  Nanogram(s)
nmol  Nanomole(s)
OLB  Oligo labelling buffer
OD  Optical density
$^{32}\text{P}$  $\beta$-emitting isotope of phosphorus
pers. comm.  Personal communication
PEG  Polyethylene glycol
%  Percentage
pg  Picogram(s)
A  Alanine
R  Arginine
N  Asparagine
D  Aspartate
C  Cysteine
Q  Glutamine
E  Glutamate
G  Glycine
H  Histidine
I  Isoleucine
L  Leucine
K  Lysine
M  Methionine
F  Phenylalanine
P  Proline
S  Serine
T  Threonine
W  Tryptophan
Y  Tyrosine
V  Valine
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TITLE PAGE</td>
<td>i</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>vi</td>
</tr>
<tr>
<td>CONTENTS</td>
<td>x</td>
</tr>
</tbody>
</table>

## CHAPTER 1: INTRODUCTION

### 1.1 SEXUAL DIMORPHISM IN DROSOPHILA MELANOGASTER

1.1.1 GROSS MORPHOLOGICAL DIMORPHISM.  
1.1.2 NEUROLOGICAL DIMORPHISM.  
1.1.3 THE GONADS.  
  1.1.3.1 Oogenesis.  
    1.1.3.1.1 The *chorion* Genes.  
    1.1.3.1.2 The *yolk protein* genes.  
    1.1.3.1.3 The Maternal Effect Genes.  
  1.1.3.2 Spermatogenesis.  
    1.1.3.2.1 Generation of Haploid Spermatids.  
    1.1.3.2.2 Spermiogenesis.  
    1.1.3.2.3 Genetic Factors Involved in Sperm Production.

### 1.2 DETERMINATION OF THE SEX OF SOMATIC TISSUE.

1.2.1 DOSAGE COMPENSATION AND *SEX-LETHAL*.  
1.2.2 *SEX-LETHAL* DETERMINES SEX.  
  1.2.2.1 Activation of *Sex-Lethal* : Zygotic and Maternal Factors.  
  1.2.2.2 Activation of *Sex-Lethal* is a Two Part Process.  
  1.2.2.3 The Next Step - transformer and transformer-2.  
  1.2.2.4 *doublesex* and *intersex* - The Sex Determination Effector Genes.

### 1.3 SEX DETERMINATION OF THE GERMLINE

54
1.3.1 CELL-AUTONOMOUS AND INDUCTIVE FACTORS 54
1.3.2 THE FEMALE INDUCTIVE SIGNAL- PRODUCTION AND TARGETS 57
1.3.3 ACTIVATION OF SXL IN THE FEMALE GERMLINE 60
1.3.4 FUNCTIONS OF TRA-2 IN THE MALE GERMLINE 62
1.3.5 SUMMARY 63

1.4 FACTORS INVOLVED IN SOMATIC SEX DIFFERENTIATION 65
1.4.1 THE YOLK PROTEIN GENES 65
1.4.2 DSX AS A GLOBAL REPRESSOR OF SEX DIFFERENTIATION GENES 67
1.4.3 BRANCHES IN THE HIERARCHY ABOVE DSX 70
1.4.4 CONTINUAL DEPENDENCE UPON THE HIERARCHY 73
1.4.5 OTHER GENES WITH SEX-SPECIFIC EXPRESSION PATTERNS 75

CHAPTER 2: MATERIALS AND METHODS 77

2.1 MATERIALS 78
2.1.1 SOLUTIONS 78
2.1.2 BACTERIOPHAGE LIBRARIES 83
2.1.2.1 Stratagene Drosophila Genomic DNA library. 83
2.1.2.2 Adult Body cDNA Library. 83
2.1.3 PLASMID VECTORS AND BACTERIAL HOSTS 84
2.1.4 DROSOPHILA STRAINS 85

2.2 METHODS 86
2.2.1 GENERAL METHODS 86
2.2.1.1 Phenol extraction. 86
2.2.1.2 Butanol extraction. 86
2.2.1.3 Precipitation of nucleic acids. 86
2.2.1.4 Quantitation of nucleic acid solutions. 87
2.2.1.5 Agarose gel electrophoresis. 87
2.2.1.6 DEAE purification of DNA fragments from agarose gels. 88
2.2.1.7 Qiaex purification of DNA fragments from agarose gels. 89
2.2.1.8 Restriction endonuclease digestion of DNA. 89
2.2.1.9 Ligation of DNA molecules. 89
2.2.1.10 Transformation of DNA into CaCl₂-competent *E. coli*. 90
2.2.1.11 Transformation of *E. coli* by electroporation. 90
2.2.1.12 Preparation of genomic DNA. 91
2.2.1.13 Small scale preparation of plasmid DNA. 91
2.2.1.14 Large scale preparation of plasmid DNA via Qiagen columns. 92
2.2.1.15 Large scale preparation of plasmid DNA via CsCl gradient. 92
2.2.1.16 Preparation of total cellular RNA. 93
2.2.1.17 Preparation of PolyA⁺ RNA. 94
2.2.1.18 Southern blot analysis. 95
2.2.1.19 Northern blot analysis. 95
2.2.1.20 Radiolabelling of DNA by random priming. 96
2.2.1.21 Determination of radiolabelling efficiency. 96
2.2.1.22 Autoradiography. 97
2.2.1.23 Labelling of DNA with digoxigenin-dUTP. 97
2.2.1.24 Labelling of RNA with digoxigenin-UTP. 97
2.2.1.25 Detection of digoxigenin-labelled probes. 98
2.2.1.26 Determination of DIG-dUTP labelling efficiency. 98

2.2.2 DNA SEQUENCING 99
2.2.2.1 Manual dideoxynucleotide DNA sequencing. 99
   2.2.2.1.1 Preparation of single-stranded DNA template. 99
   2.2.2.1.2 Denaturation of double-stranded DNA. 100
   2.2.2.1.3 Annealing of template and primer. 100
   2.2.2.1.4 Termination reactions. 100
   2.2.2.1.5 Polyacrylamide gel analysis of DNA sequencing reactions. 101
2.2.2.2 Automated dideoxynucleotide DNA sequencing. 101

2.2.3 MANIPULATIONS WITH BACTERIOPHAGE LAMBDA 102
2.2.3.1 Plating bacteriophage lambda. 102
2.2.3.2 Screening bacteriophage lambda libraries. 102
2.2.3.3 Preparation of bacteriophage lambda DNA. 103
CHAPTER 4: REGULATION OF SEX-SPECIFIC TRANSCRIPTS
PRODUCED FROM A NOVEL GENE IN DROSOPHILA MELANOGASTER.

4.1 INTRODUCTION

4.2 RESULTS

4.2.1 THE MALE-SPECIFIC TRANSCRIPT IS TESTIS-SPECIFIC
4.2.2 TRA-2 REGULATION OF THE TESTIS-SPECIFIC TRANSCRIPT
4.2.3 TRA REGULATION OF THE TESTIS-SPECIFIC TRANSCRIPT
4.2.4 DSX REGULATION OF THE TESTIS-SPECIFIC TRANSCRIPT
4.2.5 THE TESTIS-SPECIFIC TRANSCRIPT CONTAINS TCE ELEMENTS
4.2.6 REGULATION OF THE FEMALE CARCASS-SPECIFIC TRANSCRIPT
  4.2.6.1 The female carcass-specific transcript contains dsx-like 13-nt repeats.
  4.2.6.2 Ira-2 regulation of the female carcass-specific transcript.
  4.2.6.3 Ira regulation of the female carcass-specific transcript.

4.3 DISCUSSION

CHAPTER 5: SEQUENCE AND BACTERIAL EXPRESSION ANALYSIS
OF THE PREDICTED PROTEIN FROM A DROSOPHILA MELANOGASTER
GENE ENCODING SEX-SPECIFIC TRANSCRIPTS.

5.1 INTRODUCTION

5.2 RESULTS

5.2.1 PREDICTED PROTEIN SEQUENCE ANALYSIS.
5.2.2 BACTERIAL EXPRESSION STUDIES.
  5.2.2.1 Codon preference analysis of the stk61 open reading frame.
  5.2.2.2 Construction of pGEX\textsuperscript{STK61}.
  5.2.2.3 Bacterial expression of STK61/GST fusion protein.
  5.2.2.4 Western blot analysis of STK61/GST fusion protein.
  5.2.2.5 Kinase assays of crude protein extracts containing STK61/GST protein.
  5.2.2.6 Solubility of STK61/GST protein in bacterial cells.
  5.2.2.7 Purification of STK61/GST and GST proteins.

5.3 DISCUSSION

5.3.1 PROTEIN KINASES IN SIGNAL TRANSDUCTION
CHAPTER 1

INTRODUCTION
1.1 SEXUAL DIMORPHISM IN *DROSOPHILA MELANOGASTER*

The difference between one cell type and another can be defined by two major criteria; the genes expressed by the cell and the morphology of the cell. This distinction is to some extent artificial, as the morphological development of the cell will be directed along specific pathways dependent upon the genes which are expressed in the cell. However, not all cells of the same morphology will have exactly the same patterns of gene expression. In the fruit fly, *Drosophila melanogaster*, the body tissues can be divided into groups dependent upon their morphology. Some of these tissues are present in both males and females and may appear to be identical. Some of these apparently identical tissues will, however, represent different cell types as defined by their patterns of gene expression, due to the fact that there are genes active in the tissue which define it as either male or female. Some tissues are present only in one sex. Thus we can divide the fly up into three major tissue types. These are sexually-dimorphic somatic tissues (present in both sexes), sex-limited somatic tissues (present in only one sex) and germline tissues.

There are essentially two components of the developmental process which lead to differences between these tissues. These are determination and differentiation. The available evidence shows us that somatic tissues are determined to be male or female via the action of a hierarchy of regulatory genes (reviewed in Baker 1989; Slee & Bownes, 1990; Burtis, 1993; Cline 1993, Ryner & Swain, 1995). The sex determination of the germline, while requiring some products of the same hierarchy is ultimately brought about by a set of different genes (reviewed in Steinmann-Zwicky, 1992).

Once a tissue or cell has been determined as male or female it must then use this information as a framework within which to interpret the various positionally and temporally restricted signals which tell any given cell or tissue where it is in the fly and what function it must perform. In this way, a specific pattern of gene expression
will be implemented which will eventually lead to the correct development of the
tissue or cell. In order to understand how the sex determination decision is
transduced into sex-specific characteristics, it is necessary to study the regulation and
function of genes which are modulated by the sex determination decision. The
isolation and characterisation of such sex differentiation genes is the subject of the
work described in this thesis. It is first necessary to review the available evidence
concerning the sex determination signal which is responsible for the development of
sex-specific features. We will first consider the tissues of *Drosophila* which are
morphologically or genetically sex-specific.

1.1.1 GROSS MORPHOLOGICAL DIMORPHISM.

In *Drosophila melanogaster* sexual dimorphism is manifest in a number of structures
(figure 1.1). Females are larger than males and contain seven segments instead of
six. The posterior segments are darkly pigmented in the male but not in the female.
The male has a muscle in the fifth abdominal segment which is not present in the
female (Muscle of Lawrence or MOL). Until recently this muscle was thought to be
involved in curling of the abdomen during copulation. However, certain mutants of
the *fruitless* gene lack this muscle but are still able to copulate (Taylor *et al.*, 1994).
At present it is not clear exactly what role this muscle plays. The foreleg of the male
exhibits a short row of thick bristles known as the sex comb. The genitalia and analia
are also clearly distinguishable. The analia is derived from a single group of imaginal
cells which will develop down either a male or female pathway, while the genitalia
develop from two distinct populations of cells, with only one group developing in

Although the fat bodies of males and females appears largely similar, the female fat
body cells express the *yolk protein* (*yp*) genes which are required for oogenesis
(reviewed in Bownes, 1994). Thus, this tissue is dimorphic at the genetic level.
Figure 1.1
Gross morphological sexual dimorphism in *Drosophila* (taken from MacDougall *et al.*, 1995) (a) Males are smaller than females. (b) Males have a row of thickened bristles on the foreleg, the sex combs (SC). (c) There is a male-specific muscle in fifth abdominal segment, the muscle of Lawrence (MOL). (d) In males, the posterior abdomen is darkly pigmented. Genitalia and analia are also dimorphic. (e) Gonads are clearly dimorphic. A=Accessory gland, T=Testis.
1.1.2 NEUROLOGICAL DIMORPHISM.

The central nervous system of *Drosophila* exhibits several dimorphic characteristics. In the mushroom bodies of the brain, there are a greater number of cells in the female, including increased numbers of Kenyon fibres (Technau, 1984). Also, the terminal abdominal neuroblasts of the male undergo extra divisions which do not occur in the female (Taylor and Truman, 1992). It is not known what part these developmental differences play in determining the behavioural responses of the adult. However, it is interesting to note that perturbation of certain male courtship behaviours can result from feminisation of mushroom body tissue under the action of a transgene expressing the female from of a sex-determining gene (Ferveur *et al.*, 1995; O'Dell *et al.*, 1995). This will be discussed in more detail later in this chapter.

There are also sex-specific genital neurons which develop from the genital imaginal disc (Laugé, 1980). A dimorphic pattern of axon wiring and difference in the number of gustatory receptors is seen in the adult foreleg (Possidente & Murphey, 1989). It is also interesting to note that the development of the muscle of Lawrence depends not upon the sexual identity of the muscle itself, but of the innervating axons (Lawrence & Johnston, 1984; 1986). Thus, the axons which innervate this muscle must be sexually dimorphic, at least at the genetic level.

1.1.3 THE GONADS.

Gynandromorph studies indicate that the somatic component of the gonads develops from the same group of progenitor cells in both sexes (Szabad & Nöthiger, 1992). Clearly, the male and female gonads are highly structurally dimorphic, as would be expected (Figure 1.2). At the genetic level, there are also a large number of genes expressed sex-specifically in this tissue which are mainly involved in the differentiation of the germ cells into functional gametes. Gametogenesis has been well characterised in *Drosophila* and is discussed below.
1.1.3.1 Oogenesis.

The appearance of the *Drosophila* oocyte throughout oogenesis is shown in figure 1.3, and the major features of each stage are summarised in table 1.1 (Mahowald & Kambysellis, 1980).

Oogenesis begins in the germarium at the tip of the ovariole. A single oocyte progenitor cell, or oogonium, undergoes a mitotic division to produce two cells. The telophase of this division does not occur to completion, since cytokinesis is only partial and the cells remain joined by a cytoplasmic bridge. Each of the two cells undergoes another incomplete mitotic division, and so on until an interconnected 16-cell cyst is formed. Since cytoplasmic bridges only exist between mitotic 'partners', the 16-cell cyst contains only two cells which are linked by four cytoplasmic bridges. It is one of these two cells which will go on to complete meiosis and form the oocyte. The other 15 cells will develop as nurse cells, providing components to the oocyte which are required for successful oogenesis. One such nurse cell-supplied product is the mRNA from the gene *bicoid*. 
(Nüsslein-Volhard et al., 1987). In a wild type oocyte, bicoid mRNA is supplied to the anterior of the oocyte by the nurse cells, where it is localised. The resulting gradient of Bicoid protein is responsible for determining the polarity of the embryo. In embryos mutant for the gene dicephalic, the oocyte is placed centrally in the follicle, with nurse cells at either end (Lohs-Schardin, 1982). This results in bicoid mRNA being supplied to both ends of the oocyte which causes the embryo to develop with bilateral symmetry, having a head at both ends.

**Figure 1.3**
Light microscopy photographs of *Drosophila* oocytes throughout oogenesis (taken from Mahowald & Kambysellis, 1980). Stages 1-14 are indicated. Stages 1-7 @ X 380 magnification, using phase contrast microscopy. Stages 8-14 @ X 100 magnification, using bright field microscopy. M=Muscle sheath, G=Germarium, F=Follicle cells, N=Nurse cells, ON=Oocyte nucleus.
<table>
<thead>
<tr>
<th>Stage</th>
<th>Duration (hours)</th>
<th>Size (µm)</th>
<th>Major Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.56</td>
<td></td>
<td>16-cell cyst in germarium.</td>
</tr>
<tr>
<td>3</td>
<td>9.56</td>
<td></td>
<td>Oocyte nucleolus disappears.</td>
</tr>
<tr>
<td>4</td>
<td>9.16</td>
<td></td>
<td>Bulbous chromosomes in nurse cells.</td>
</tr>
<tr>
<td>5</td>
<td>2.61</td>
<td></td>
<td>Bulbous chromosomes disappear.</td>
</tr>
<tr>
<td>6</td>
<td>8.45</td>
<td></td>
<td>Cessation of follicle cell divisions</td>
</tr>
<tr>
<td>7</td>
<td>8.69</td>
<td>145</td>
<td>Elongation apparent. Follicle cells begin polyploidisation and become columnar over oocyte.</td>
</tr>
<tr>
<td>8</td>
<td>5.21</td>
<td>190</td>
<td>Yolk formation begins.</td>
</tr>
<tr>
<td>9</td>
<td>5.61</td>
<td>275</td>
<td>Most follicle cells over oocyte. Secretion of vitelline membrane begins. Border cells migrate.</td>
</tr>
<tr>
<td>10</td>
<td>5.13</td>
<td>430</td>
<td>Follicle cells are columnar over oocyte and squamous over nurse cells.</td>
</tr>
<tr>
<td>11</td>
<td>0.4</td>
<td>490</td>
<td>Nurse cells empty contents into oocyte. Chorion secretion begins.</td>
</tr>
<tr>
<td>12</td>
<td>1.9</td>
<td>540</td>
<td>Nurse cells shrunken. Formation of chorionic appendage begins.</td>
</tr>
<tr>
<td>13</td>
<td>1.79</td>
<td>650</td>
<td>Nurse cell nuclei begin to disappear.</td>
</tr>
<tr>
<td>14</td>
<td>1.98</td>
<td>700</td>
<td>Mature oocyte.</td>
</tr>
</tbody>
</table>

**Table 1.1**
Summary of the major features of stages 1-14 of oogenesis, as shown in figure 1.3.

At stage 1 of oogenesis, the follicle is present in the germarium and consists of the 16-cell cyst surrounded by somatic follicle cells. As the follicle develops into a mature oocyte, it passes down the ovariule until it is present in the oviduct. It is then fertilised by sperm stored in the spermatheca, and passes into the uterus to be laid as a fertilised egg.

At stage 2, the nurse cells begin to become polyploid, in preparation for the large amounts of maternal products which will be laid down in the developing oocyte. By stage 5 the follicle cells have divided 4 times and still completely cover the oocyte.
and nurse cells. At stage 6, the follicle cells stop dividing and start to enlarge (stage 7) and become polyploid. Yolk starts to be deposited at stage 8 and by stage 9 the vitelline membrane begins to be laid down over the oocyte by the follicle cells. It is also at around stage 9 that the oocyte, which has been transcriptionally silent until then, begins some transcription and a cluster of cells migrate through the nurse cells and associate with the anterior of the oocyte to form the border cells. By stage 10 the follicle cells have enlarged and are squamous where they cover the nurse cells, but columnar over the oocyte. These columnar follicle cells begin to secrete the chorion at stage 11, and the nurse cells begin a rapid emptying of their contents into the oocyte. By stage 12, the now shrunken nurse cells form the 'anterior cap' of the oocyte, and the formation of the dorsal chorionic appendage begins. At stage 14, the chorion is completely formed and the oocyte is mature and ready for fertilisation.

1.1.3.1.1 The chorion Genes.

The chorionic membrane consists of around 20 proteins exported from the overlying follicle cells (Mahowald & Kambysellis, 1980). Genes have been cloned which encode six chorionic proteins; s36, s38, s15, s16, s18 and s19 (Spradling, 1981). The genes encoding s36 and s38 are X-linked, while the genes encoding the other four proteins are present in a cluster on chromosome 3. Within the follicle cells, these genes undergo 4-6 rounds of bi-directional replication which presumably is necessary to enable the production of sufficient chorion proteins. This replication event does not appear to be required for correct temporal regulation of chorion gene expression, but may act to regulate the levels of each chorion protein produced (Parks et al., 1986).

1.1.3.1.2 The yolk protein genes.

The three yolk proteins YP1, YP2 and YP3 are synthesised in the female fat body and in the ovarian follicle cells (Bownes & Hames, 1978; Brennan et al., 1980). The YP's synthesised in the fat body are carried to the ovary via the haemolymph, where
both the fat body-synthesised and follicle cell-synthesised YPs are taken into the oocyte via receptor mediated endocytosis (Bownes, 1986; Giorgi et al., 1979).

Cloning of the genes encoding the YP proteins has revealed that all three genes are X-linked (Barnett et al., 1980). The yp1 and yp2 genes are close together and divergently transcribed, while the yp3 gene lies several kilobases downstream. The sex-specific regulation of these genes will be discussed later in this chapter.

Female flies which are starved after eclosion show greatly reduced levels of yp gene transcription (Bownes et al., 1988). When these flies are fed, yp transcription levels return to normal. Transgenic studies delimited a 890bp fragment of DNA between yp1 and yp2 which is sufficient to confer this nutrient-dependent expression.

Expression of the yp genes is also regulated by the hormones juvenile hormone and ecdysone (Bownes et al., 1993). Injection of 20-hydroxyecdysone into male flies results in transient expression of the yp genes in the fat body (Bownes et al., 1983) and an analogue of the ecdysone receptor (ponasterone A receptor) binds to a DNA restriction fragment which lies between the yp1 and yp2 genes (Pongs, 1989, in Bownes, 1994). Fragments of DNA lying upstream, downstream and within yp3, and between yp1 and yp2, have been isolated which confer 20-hydroxyecdysone-dependent induction on transgenic reporter gene constructs (Bownes et al, 1996).

Although ecdysone is able to stimulate yp expression in the fat body, a similar effect is not seen in the ovary follicle cells. A juvenile hormone analogue, however, is able to stimulate yp expression in both tissues (Jowett & Postlethwait, 1980). As yet, no specific DNA sequences have been identified which may be involved in the juvenile hormone response, and it is not known how this signal is transduced into regulation of the yp genes. Juvenile hormone analogue-dependent induction of transgenic reporter genes is only seen with yp-encoding constructs, suggesting that juvenile hormone may act at regulatory elements within the yp transcription unit, or may have
a role in post-transcriptional yp regulation (Bownes et al, 1996). It is not clear exactly what role hormone regulation has in the sex-specific patterns of yp gene expression seen in wild type flies, since natural levels of both juvenile hormone and ecdysone appear to be fairly constant in both sexes (Bownes, 1989).

1.1.3.1.3 The Maternal Effect Genes.

Both the chorion and yp genes are expressed maternally, with their protein products being supplied to the developing oocyte. Genetic analysis has revealed a large number of other 'maternal effect' genes, mutation of which results in oogenic/embryogenic defects. In particular, groups of maternal effect genes have been isolated which result in polarity defects such as ventralisation or dorsalisation in the developing embryo.

Cytoplasmic transplantation experiments indicate that there are factors localised at the anterior and posterior poles of the egg which established morphogen gradients, defining thoracic-head and abdominal structures respectively (Nüsslein-Volhard et al., 1987). The genes responsible for these activities have been identified by screening for mutations which cause deletions of thoracic and abdominal regions. This analysis reveals genes which fall into three groups according to the areas of the embryo in which they specify polarity. These are the anterior, terminal and dorso-ventral maternal effect genes. Together, these genes lay down the anterior-posterior and dorsal-ventral polarity of the oocyte (and therefore the embryo), providing a framework of positional cues for the zygotically-expressed gap, pair-rule and segment polarity genes which determine the positional identity of embryonic cells within each developing segment.

The anterior polar centre is organised by the 'anterior group' of maternal effect genes. These are the genes bicoid (bcd), exuperantia (exu), swallow (swa), bicaudal (bic), bicaudal-C (bic-C) and bicaudal-D (bic-D) (St Johnston & Nüsslein-Volhard, 1992). Bicoid is a homeobox-containing protein which acts as a morphogen to define the
polarity of the anterior of the embryo (Berleth et al., 1988). This is clearly shown by injection of bicoid RNA into ectopic sites of the embryo which results in head structures forming at the injection site (Driever et al., 1990). bicoid transcript is localised to the anterior pole of the egg dependent on the gene products of the exu, swa and staufen genes (Berleth et al., 1988; Stephenson et al., 1988; St Johnston et al., 1989). At fertilisation of the egg, translation of the RNA begins and gives rise to a gradient of Bicoid protein. That Bicoid is indeed a true morphogen is shown by the increase in anterior structures in response to increasing maternal bcd gene copies or injections of anterior cytoplasm (Struhl et al., 1989). The mechanism by which these concentration-dependent decisions are made is shown by the presence of six bcd binding sites upstream of the hunchback gene (Driever & Nüsslein-Volhard, 1989). Three of these sites are of high affinity and three of low affinity. In this way the gene is activated in the anterior half of the embryo in response to bcd concentrations above a certain threshold level. Other possible targets of bcd include the gap genes orthodenticle, empty spiracle and button head which lead to anterior deletions when mutant (in St Johnston & Nüsslein-Volhard, 1992).

The 'posterior group' maternal effect genes direct the identity of the posterior region of the embryo and include the genes nanos (nos), pumilio (pum), oskar (osk), vasa (vas), tudor (tud), valois (val) and mago nashi (St Johnston & Nüsslein-Volhard, 1992). Also required are the staufen (stau) gene and the zygotically-required genes cappuccino (cap) and spire (spi). cap and spi are involved in the localisation of the maternal products to the posterior pole of the egg (Manseau & Schüpbach, 1989; St Johnston et al., 1991). This localisation takes place in a stepwise manner with each step dependant on the previous one. These maternal products go to make up part of the polar granules; densely staining structures which form at the posterior pole of the embryo and are incorporated into the pole cells at cellularisation. Maternal mutations which cause a lack of polar granules result in the production of sterile offspring due to a lack of germ cells. The mutant effect of all of the posterior group genes apart from pum and nos probably results from the inability to form these polar granules. Nanos protein is the likeliest candidate for the morphogen in this system, since in
nos mutants the polar granules are able to form but the posterior cytoplasm has no morphogen activity (Lehmann & Nüsslein-Volhard, 1991). Indeed, injection of nanos RNA into embryos from mothers mutant for all posterior group genes rescues the mutant phenotype (Wang & Lehmann, 1991). In pum mutants, polar granules form and rescuing activity is present. Thus, it seems likely that pum product is involved in the release of the nos morphogen (Lehmann & Nüsslein-Volhard, 1987). The only function of Nos appears to be the destablisation of maternal hunchback RNA which allows the activity of hunchback-repressed gap genes in the posterior half of the embryo (Irish et al., 1989).

The 'terminal group' of maternal effect genes are responsible for the specification of the acron and telson (head and tail structures) of the embryo and their mode of action is very similar to that of the dorso-ventral genes which set up the dorso-ventral axis in the embryo. Both these sets of genes provide a spatially restricted ligand which is released from the follicle cells and recognised by a receptor which is present in the plasma membrane of the developing oocyte (see figure 1.4).

![Diagram showing models for the maternal effect genes which are involved in setting up the terminal polarity and dorsal-ventral polarity of the Drosophila embryo. Major genes and proteins involved are indicated. See text for details.](image-url)
The terminal group maternal effect genes are torso (tor), torsolike (tsl), l(l)pole-hole (l(l)ph), fs(1)pole-hole (fs(1)ph), fs(1)Nasrat (fs(1)N) and trunk (trk) (St Johnston & Nüsslein-Volhard, 1992). Cloning of the gene encoding Torso protein has revealed that it has homology to receptor tyrosine kinases (RTK's) and that the protein is localised all over the egg plasma membrane (Sprenger et al., 1989; Casanova & Struhl, 1989). Constitutive gain-of-function alleles of this gene have shown a pattern of epistasis where the genes trk, fs(1)N, fs(1)ph and tsl lie upstream of tor and the gene l(l)ph lies downstream of tor (Ambrosio et al., 1989; Stevens et al., 1990). The somatic and germline specificity of the above mutants was shown by pole cell transplantations between embryos mutant for different genes. Thus trk, fs(1)N and fs(1)ph are required in the germline and tsl is required in the follicle cells but not the germline. Also, an X-ray induced tsl/tsl clone in the posterior polar follicle cells results in the loss of the posterior filzkorper structure. Thus, the current model suggests that tsl encodes the ligand for Tor and is released by the follicle cells at either pole of the embryo (figure 1.4). The polar identity of the follicle cells is believed to be defined by processes involving the genes notch (N) and Delta (Dl), since hypomorphemic alleles of these genes can result in mislocalisation of bicoid to the posterior of the embryo as well as the anterior (Ruohola et al., 1991). trk, fs(1)N and fs(1)ph which are released from the oocyte, may be required for the activation of Tsl which then binds to and activates Torso. Torso protein may then act to phosphorylate L(1)ph which in turn would go on to direct transcription/regulation of various downstream genes such as the terminally-expressed tailless and huckebein gap genes (Weigel et al., 1990).

The 'dorso-ventral group' maternal effect genes act to determine the dorsal-ventral polarity of the embryo and include the genes gastrulation-defective, snake, easter, spatzle, toll, tube, pelle, dorsal, cactus, pipe, nudel and windbeutel (St Johnston & Nüsslein-Volhard, 1992). The genes gastrulation-defective, snake, easter, spatzle, toll, tube, pelle and dorsal belong to a group of genes which cause a dorsalising of the embryo when hypomorphemic and a ventralising effect when hypermorphemic. Mutant alleles of the genes cactus, pipe, nudel and windbeutel ventralise the embryo when
hypomorphic or amorphic and dorsalise it when hypermorphic. *toll* encodes a transmembrane receptor with homology to the interleukin-1 (IL-1) receptor and is present ubiquitously in the oocyte plasma membrane (Hashimoto et al., 1988; 1991). Again, constitutive gain-of-function alleles of *toll* have enabled the epistatic relationship between these genes to be elucidated (Anderson et al., 1985). The somatically-required genes *pipe, nudel* and *windbeutel* together with the germline-required genes *gastrulation-defective, snake, easter* and *spatzle* lie upstream of *toll* while the germline-required genes *tube, pelle, cactus* and *dorsal* lie downstream of *toll*. Injection of perivitelline fluid from *toll/toll* flies induces the embryo to develop with its ventral side at the site of injection, if the mother is mutant for *pipe, nudel* or *windbeutel* (Stein et al., 1991). *toll/toll* mothers must be used to provide this fluid as otherwise, the inducing ligand is sequestered from the fluid by the Toll receptor. This is in agreement with a role for these genes in the production of ligand from ventral follicle cells and explains their mutant phenotypes.

The nature of the Toll ligand is not known, but injection of perivitelline fluid can rescue the mutant phenotypes of embryos from *snake, easter* and *spätzle* mutant mothers (Stein & Nüsslein-Volhard, 1992, in St Johnston & Nüsslein-Volhard, 1992), suggesting that these three genes may encode the ligand or have a role in ligand production. The observation that the *snake* and *easter* genes encode putative serine proteases suggests a possible role for these gene products in activation of the ligand by proteolytic cleavage (DeLotto & Spierer, 1986; Chasan & Anderson, 1989). The end result of the functions of all of these gene products is the setting up of a nuclear gradient of Dorsal protein such that ventral nuclei contain higher concentrations of Dorsal than dorsal nuclei (Roth et al., 1989). The use of *cactus/dorsal* double mutants to produce intermediate nuclear concentrations of dorsal has indicated that the downstream targets of dorsal are *zerknült* (*zen*) and *decapentaplegic* (*dpp*) which are active at low Dorsal concentrations and *twist* and *snail*, which are active at high Dorsal concentrations (Roth et al., 1989). This also shows that Dorsal is a true morphogen, producing different concentration-dependent effects. The homology of Cactus to IκB protein and the homology of Dorsal to
NF-κβ suggests a model for Cactus, Dorsal and Toll protein action based on the function of the homologous proteins as shown in figure 1.5 (St Johnston & Nüsslein-Volhard, 1992). This model has been shown to be essentially correct (Norris & Manley, 1996) and will be discussed further in chapter 5.

**Figure 1.5**
Diagram showing mode of action of II-1, Iκβ and NF-κβ, (a), and a proposed model for Toll, Cactus and Dorsal proteins, (b), based on the shared homology of these proteins. Other factors may lie between Toll and Cactus- no direct phosphorylation of Cactus is indicated.

1.1.3.2 Spermatogenesis.

In oogenesis, each meiosis gives rise to only one haploid oocyte, with the other two products of the meiotic divisions becoming polar bodies which eventually degenerate. However, in the testes, four functional haploid spermatozoa are produced from each meiosis.

1.1.3.2.1 Generation of Haploid Spermatids.

Spermatogenesis begins at the very tip of the testis (see figure 1.2) with a small number of diploid stem cells. These stem cells undergo four mitoses before they enter meiosis and spermatogenesis (reviewed by Lindsley & Tokuyasu, 1980). The first mitotic stem cell division gives rise to a second stem cell, which remains at the tip of the testis, and a primary spermatogonial cell, which completes the further 3 mitoses and eventually gives rise to 64 spermatozoa. The 64 progeny cells from each primary spermatogonial cell develop synchronously, moving down the testis towards the testicular duct as they develop. The primary spermatogonial cell is surrounded by
two somatic cyst cells which remain with the germ cells throughout their divisions. The initial mitotic divisions of the primary spermatogonial cell produce a cyst containing 16 primary spermatocytes. These cells remain connected by intercellular bridges (ring canals), in the same way as the 16 cells produced from the initial stem cell divisions of oogenesis. The diploid primary spermatocytes undergo a 90 hour growth phase during which their volume increases approximately 25 times. This growth phase is followed by two meiotic divisions, producing 64 haploid spermatids which are still linked by ring canals.

1.1.3.2.2 Spermiogenesis.

The process by which a spermatid differentiates into a spermatozoon is called spermiogenesis, which lasts 5-6 days and comprises 3 main phases known as elongation, individualisation and coiling (reviewed by Lindsley & Tokuyasu, 1980).

Prior to the main phase of elongation a pre-elongation phase occurs. During pre-elongation the mitochondria of the spermatid assemble into a bipartite laminate sphere known as the nebenkern. During the second meiotic division, the centriole does not replicate, leaving a single centriole present in the spermatid. This centriole forms the 'basal body', at the base of the nucleus. The axoneme begins to extend from the basal body and is surrounded by plasma membrane to form a cilium. The acrosome begins to be formed from Golgi body which repositions at the anterior pole of the nucleus.

During elongation the spermatid cyst extends to the point where it occupies 80-90% of the length of the testis. The axoneme greatly extends and becomes associated with cylindrical rods formed from the nebenkern. The nucleus metamorphoses from a sphere into a needle-like structure.
During individualisation, the ring canals linking the spermatids are broken down. A swelling, known as the cystic bulge, moves down the length of the sperm bundle, sweeping away the ring canals as well as any excess organelles and cytoplasm.

Coiling is the final process in spermiogenesis. One of the two cyst cells becomes associated with the membrane at the base of the testis and the sperm bundle moves down the testis as it is coiled. The coiled spermatozoa then separate from the cyst cells and move to the seminal vesicle.

1.1.3.2.3 Genetic Factors Involved in Sperm Production.

Much less is known about the genetic control of sperm formation than about oogenesis. However, it has been known for some time that the Y chromosome is required for correct spermatogenesis, since XO males develop normally in all respects other than in spermatogenesis, where severe meiotic disruption is observed (Lifschytz & Hareven, 1977). Such males are sterile. Six Y-linked complementation groups (known as fertility factors) are responsible for the role of the Y chromosome in spermatogenesis (Kennison, 1981; Hazelrigg et al., 1982; Gatti & Pimpinelli, 1983). No functional sperm are produced in males lacking either the whole Y chromosome or one or more of the fertility factors; ks-1, ks-2, kl-1, kl-2, kl-3 and kl-5. Elongation does occur, but spermatids degenerate before reaching maturity. The three fertility factors ks-1, kl-3 and kl-5 correspond to thread-like structures seen in spermatocytes which are thought to represent large Y chromosome loops at these loci which may have some structural role in spermatogenesis (Bonaccorsi et al., 1988).

Approximately 20% of EMS-induced X-linked male sterile mutants exhibit some level of meiotic disruption in spermatogenesis. In a characterisation study of male-sterile mutants, Lifschytz & Hareven, 1977, described 3 categories of spermatogenic defect which result in male sterility; namely, timing mutations (nebenkern formation occurs before, rather than after, meiosis), spindle structure...
mutations and primary spermatocyte deformation mutations. Of the first group, the
ms(1)413, ms(1)682 and ms(1)RD11 X-linked lesions cause the nebenkern to form
prior to meiosis and result in a perturbation of its structure. Nebenkern stability also
appears to be affected, since disaggregation is seen during elongation. The ms(1)516
X-linked lesion characterises the second group. This mutation causes a perturbation
of the spindles seen during the second meiotic division where only one set of
centrioles is seen, rather than two. Of the third group, the ms(1)401 X-linked lesion
appears to cause primary spermatocytes to become stalled prior to meiosis. An
accumulation of primary spermatocyte late stages is seen, although some elongation
does take place. In addition, the stalled spermatocytes show structural abnormalities
such as fragmentation of nucleoli.

In a screen performed by Schäfer, 1986a, five genes with male gonad-specific
transcripts were isolated. Only one of these transcripts, encoded by the Mst87F gene
(originally mst(3)gl-9), was not detected in germline-deficient males, indicating that
this transcript is germline-specific. The other four transcripts were shown to be
localised to accessory gland tissue. The diploid primary spermatocytes are highly
transcriptionally active during their growth phase (Olivieri & Olivieri, 1965). After
this stage however, virtually no transcription is seen to occur. Clearly, all of the
mRNA required to make protein during spermiogenesis of the haploid spermatids
must be stored in the cells prior to meiosis. We might expect translational control of
these transcripts to be required to coordinate the expression of spermiogenic
proteins. Translation from the Mst87F transcript has been shown to be under
translational control during spermiogenesis (Kuhn et al., 1988). Although the
Mst87F transcript is present in the primary spermatocyte, Mst87F protein does not
appear until late in spermatid elongation. Transgenic studies have shown that a 24bp
section of the Mst87F 5' UTR is sufficient to confer this translational control upon a
reporter gene (Schäfer et al., 1990). This translational control element (TCE) is the
subject of further discussion in chapter 4. The Mst87F gene belongs to a family of
genes known as the Mst(3)CGP genes (Schäfer et al., 1993). They all encode
proteins with cysteine/glycine/proline repeats, all contain a TCE element at an
invariant position and are all translationally regulated. The Mst(3)CGP genes encode structural sperm tail proteins which may form part of the axoneme (Kuhn et al., 1988).

The UTR sequences of the anterior group maternal effect gene exuperantia (exu) are also implicated in transcript regulation in the male germline (Hazelrigg & Tu, 1994). The exu gene function is required in the male germline for correct spermatogenesis since exu mutant males are sterile, exhibiting disfunctional non-motile sperm (Hazelrigg et al., 1990). Both testis-specific and ovary-specific exu transcripts contain the same coding potential, but differ at their 5' and 3' ends, with the testis-specific transcript having an extended 3' UTR (Hazelrigg & Tu, 1994). The exu allele, exu<sup>PP3</sup>, results in production of a testis-specific exu transcript which lacks most of the testis transcript-specific 3' UTR and causes male sterility but has no effect on oogenesis (Hazelrigg & Tu, 1994). Deleting specific regions of the exu testis transcript-specific 3' UTR results in male sterility and reduced levels of exu mRNA (Crowley & Hazelrigg, 1995). Thus, the function of the extended 3' UTR may be to stabilise the exu transcript in a testis-specific manner.

The sex determining gene tra-2 is also required in the male germline for correct spermatogenesis, as mutant males are sterile, producing non-functional sperm (Belote & Baker, 1983). However, this effect is at least partially indirect, since tra-2 has been shown to be required for efficient production of the male testis-specific exu transcript (Hazelrigg & Tu, 1994).

Some progress has been made in elucidating the transcriptional regulation of primary spermatocyte-expressed genes by the identification of specific regulatory elements in the β2 tubulin gene promoter (Michiels et al., 1989). The β2 tubulin protein is the major tubulin isotype which forms part of the primary spermatocyte cytoskeleton and is expressed only in the testis. Transgenic analysis shows that 53bp of β2 tubulin gene upstream sequence together with the first 23bp of the transcription unit is sufficient to confer primary spermatocyte-specific expression upon a reporter gene.
Further deletion of the 53bp upstream sequence shows that a 14bp section which has been called the β2 upstream element 1 (β2UE1) is absolutely required for this transcriptional regulation. Identification of other genes containing the β2UE1 in their promoters will hopefully enable further elucidation of the genetic processes required for primary spermatocyte development and spermiogenesis.

We would expect that a process such as spermatogenesis would be highly dependent upon precise control of the cell division cycle and this is indeed the case. A simplified model showing the major processes involved in cell cycle regulation is shown in figure 1.6. Sigrist et al., 1995, have shown that a separate CDC25 phosphatase, the Twine/CDC25 phosphatase, is required in male germ cells to drive the cells into M-phase. It is possible that this protein is required in male germ cells to compensate for the lack of the String/CDC25 phosphatase which is maternally supplied to female germ cells. Flies mutant for twine or cdc2 (encoding the p34 kinase) still go through spermiogenesis even though the chromosome segregation and cytokinesis of meiosis is incomplete (White-Cooper et al., 1993; Sigrist et al., 1995). This indicates that cell-cycle progression is to some extent independent of spermiogenesis. However, four genes have been identified, in which mutations cause the primary spermatocytes of mutant male flies to stall in growth phase, which is in fact an extended G2-phase (Lin et al., 1996). These genes are spermatocyte arrest (sa), cannonball (can), always early (aly) and meiosis 1 arrest (mia). None of these genes are required for female fertility. The authors suggest that the process of primary spermatocyte entry into meiosis and subsequent spermiogenesis can be broken down into two separate pathways; those which are twine-dependent (such as meiotic chromosome condensation and cytokinesis) and those which are twine-independent (entry into spermiogenesis). The persistence of Cyclin A protein in sa, can, aly or mia mutant spermatocytes suggests that the products of these genes may be involved in the ubiquitin-mediated pathway which usually degrades Cyclin A in a cell cycle-regulated manner.
Figure 1.6
Diagram showing the major gene products involved in cell division cycle regulation (reviewed in Pines & Hunter, 1990; Woodgett, 1991; Enoch & Nurse, 1991; MacNeill et al., 1991; Glover, 1991). A great deal of this work was carried out on the fission yeast S. pombe, but homologous genes have been found in a number of organisms, including humans, and have been shown to be functionally interchangeable in a number of cases. S=S-phase (DNA synthesis phases), M=M-phase (Mitotic division phase), G1=Gap-phase 1, G2=Gap-phase 2.

The CDC2/p34 Serine/Threonine (S/T) kinase is thought to be the main 'workhorse' of cell cycle regulation, acting to drive the cell through the divisions between the phases of the mitotic cycle, dependent upon its phosphorylation state and association with the cell cycle-regulated Cyclin proteins. Dephosphorylation of CDC2 in the CDC2/CyclinB complex in G2-phase enables transition from G2 into M-phase. Dephosphorylation of CDC2 is brought about by the CDC2-activator, CDC25 phosphatase. The Weel protein (and its redundant homologue Miki) is a Serine/Threonine/Tyrosine (S/T/Y) kinase which inhibits cell cycle progression and may phosphorylate CDC2. The Nimi S/T kinase is thought to negatively regulate Weel via phosphorylation. Suc1 protein physically associates with CDC2 and regulates G2/M transition (negatively) and mitotic progression (positively).

It is clear that Drosophila exhibits a large number of sexually dimorphic characteristics. In the following section the role of genetic factors in determining the sex of the developing fly is discussed.
1.2 DETERMINATION OF THE SEX OF SOMATIC TISSUE.

The origin of the sex determination signal was discovered by C.B. Bridges, 1921, via studies of the progeny of matings between triploid females and diploid males. During oogenesis in triploid flies one haploid set of chromosomes is distributed to the oocyte pronucleus and one haploid set to each of the three polar bodies. This still leaves one haploid set of chromosomes unaccounted for. The remaining chromosomes are randomly apportioned to either the oocyte or the polar bodies. When such flies are mated to normal males progeny can be generated with an unusual number of X chromosomes relative to numbers of sets of autosomes (the X:A ratio). This work revealed that the sex of flies is dependent upon the X:A ratio, such that flies with a ratio of 4X:4A, 3X:3A, 2X:2A or 1X:1A (an X:A ratio=1) are female while flies with a ratio 1X:2A (an X:A ratio=0.5) are male. It was also observed that flies with an X:A ratio lying between these two values such as 2X:3A (an X:A ratio=0.67) developed neither as complete males nor complete females but as intersexes, consisting of mosaic patches of male and female tissue (Bridges, 1925). The Y chromosome was seen to play no part in somatic sex determination but it is required for the development of fully functional sperm.

The mechanism by which the X:A ratio is assessed has been very well characterised. Several genes which are involved in the transduction of the X:A ratio have been identified on the basis of their mutant phenotypes as shown in table 1.2 (reviewed in Baker & Belote, 1983; Slee & Bownes, 1990). This genetic analysis revealed two major types of sexual transformation. Firstly, the conversion of females to pseudomales. Pseudomales exhibit male characteristics displaying male pigmentation and cuticular structures and possessing rudimentary testes. They are, however, of female size and infertile due to non-functional sperm. Triploid (2X:3A) intersexes appear to represent a decision made early on in development by each cell to follow either the male or female pathway of development. This is indicated by the variability of patch size in these mosaics with a large patch representing a decision made early on in development and thereafter followed autonomously by the progeny.
of that cell. Conversely, \( dsx \) mutants are true intersexes with the phenotype seeming to result from an attempt at the individual cell level to express both male and female functions. Thus, true intersexes are non-mosaic flies with each cell being neither male or female but of an intermediate phenotype.

### Table 1.2

Summary of phenotypes caused by alleles of the sex-determining genes *Sex-lethal*, *transformer*, *transformer-2*, *doublesex* and *intersex*.

<table>
<thead>
<tr>
<th>GENE</th>
<th>MUTATION</th>
<th>PHENOTYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sex-lethal</em></td>
<td>Recessive</td>
<td>Female lethal. XX cells masculinised</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male lethal. XO cells feminised</td>
</tr>
<tr>
<td><em>transformer</em></td>
<td>Recessive</td>
<td>Females transformed to pseudomales</td>
</tr>
<tr>
<td><em>transformer-2</em></td>
<td>Recessive</td>
<td>Females transformed to pseudomales</td>
</tr>
<tr>
<td><em>doublesex</em></td>
<td>Recessive</td>
<td>Females transformed to intersexes</td>
</tr>
<tr>
<td></td>
<td>Recessive</td>
<td>Males transformed to intersexes</td>
</tr>
<tr>
<td></td>
<td>Recessive</td>
<td>Both sexes transformed to intersexes</td>
</tr>
<tr>
<td></td>
<td>Dominant</td>
<td>Females transformed to intersexes</td>
</tr>
<tr>
<td><em>intersex</em></td>
<td>Recessive</td>
<td>Females transformed to intersexes</td>
</tr>
</tbody>
</table>

Evidence from a number of studies has shown that the genes discussed above operate in a hierarchy to determine sex. The epistatic relationship between these genes was first shown by mating flies which were homozygous for either *dsx* or *tra* mutant alleles, resulting in progeny homozygous mutant for both genes (Baker & Ridge, 1980). These *tra/tra, dsx/dsx* flies exhibited a phenotype typical of *dsx/dsx* mutants, indicating that *tra* and *dsx* gene products act in the same pathway and that *dsx* is epistatic to *tra*. In another set of experiments, *tra/tra* flies which were also mutant for other genes in the hierarchy were crossed to a strain containing a P-element construct with the hsp70 constitutive promoter fused to the female-specific *tra*
coding sequence (McKeown, et al., 1988). Ectopic tra expression was shown to cause development down the female pathway in tra/tra flies lacking Sxl function but the phenotypes of tra-2 and dsx mutations were unaffected. This shows that tra is epistatic to Sxl and that tra-2 and dsx are epistatic to tra. In addition, the transforming activity of this tra construct was shown to be dependent upon the integrity of the ix gene, which shows that ix product is required for the correct progression of female differentiation. A pattern of epistasis was established which suggested that these genes act in a hierarchical pathway in the following way; Sxl→tra→tra-2→dsx. Since the function of the dsx gene is dependent upon the ix gene, ix may lie downstream of dsx, or may act in conjunction with dsx (Baker & Ridge, 1980).

The epistatic relationship described above is based mainly on the genetic evidence, but is confirmed by the molecular evidence that active Sxl is required for production of relevant tra transcripts and that tra and tra-2 are required for production of relevant dsx transcripts (Nagoshi et al., 1988). However, this evidence does not rule out the possibility that tra-2 actually lies on a side-branch of the hierarchy, acting in conjunction with tra and upstream of dsx. Cloning and molecular characterisation has elucidated the mechanistic relationship between Sxl, tra, tra-2 and dsx, and is discussed below.

1.2.1 DOSAGE COMPENSATION AND SEX-LETHAL

In Drosophila melanogaster the fact that males have one X chromosome while females have two, is compensated for by a change in the transcription rate of the X-linked genes (Mukherjee & Beermann, 1965). Incorporation of tritiated uridine into hnRNA being transcribed from polytene chromosomes indicates that the single male X chromosome is hypertranscribed.

The lethality of mutations at the X-linked Sxl locus indicates that this gene must serve functions other than sex determination. Such lethality cannot be entirely
explained by incomplete development due to transformation of sex in such mutants, since tra/tra pseudomales still require active Sxl product to be viable. It has been observed that in XX larvae carrying hypomorphic alleles of Sxl, the rate of gene expression of the X chromosome is increased (Lucchesi & Skripsky, 1981). It was also observed that heterozygosity for a Sxl hypomorph can partially rescue the mutant phenotype of a hypomorphic allele of the X-linked segmentation gene runt in XX larvae, indicating that a reduction in Sxl product leads to an increase in runt gene product (Gergen, 1987). This evidence points towards a female-specific role for Sxl in deactivation of X-chromosome hypertranscription.

If the above model is correct then it follows that Sxl should be inactive in this respect in males and this is shown to be the case by the following evidence. Females homozygous for a Sxl hypomorph die as embryos but this mutation has no effect on males (Cline, 1984). Indeed, males carrying a deletion covering the Sxl locus develop normally. However, males with a hypermorphic Sxl allele are embryonic lethal while females homozygous for this allele are unaffected. The lethality of a Sxl hypomorph in the female must be due to the presence of two X chromosomes rather than the absence of the Y as the allele is also lethal in XXY flies.

Further evidence for a role for Sxl in dosage compensation as well as sex determination is seen in X/X, Sxl/Sxl X-ray induced clones of cells which, as well as being male-like in character, are also much smaller than male-like clones caused by other transforming mutations (Sánchez & Nöthiger, 1982). This reduced viability in Sxl/Sxl clones is proposed to be due to over expression of the X chromosomes. The same inviability is seen in the XO cell clones of gynandromorphs heterozygous for a hypermorphic Sxl allele (Cline, 1979). In such flies, the XO clones which carry the Sxl mutation are much smaller than those that do not. Thus, it would appear that Sxl is active only in females to turn off X chromosome hypertranscription.

Four other genes have been proposed to be the targets of Sxl in the dosage compensation pathway. These are the male-specific lethal genes 1, 2 and 3 (msl-1, 2,
and 3) and the gene *maleless* (*mle*) which kill only males when homozygous mutant (Belote & Lucchesi, 1980a; 1980b; Lucchesi et al., 1982). These genes' involvement in the dosage compensation pathway and their epistasis to *Sxl* is indicated by the partial rescue of the phenotype of \(X/X, Sxl/Sxl\) flies by mutant alleles of *msl-1, 2, 3* and *mle* (Skripsky & Lucchesi, 1982). Since the mutant phenotypes of *msl-1, msl-2* and *mle* mutant flies are similar to the phenotype observed when all three are mutant at the same time, it is likely that these genes all act in the same pathway (Bachiller & Sánchez, 1989; Baker et al., 1994). The *mle* gene bears homology to RNA and DNA helicases, supporting a role for this protein in increase of transcription rate as part of the transcription complex (Kuroda et al., 1991).

Dosage compensation of the male X is a global process, affecting the whole chromosome, as shown by rearrangements which bring genes from elsewhere in the genome under the dosage compensation control of this chromosome and also the loss of dosage compensation of genes translocated from the X chromosome to autosomes (Lucchesi & Manning, 1987; Sass & Meselson, 1991, in Henikoff & Meneely, 1993). Immunofluorescence using anti-*mle* antibody showed that Mle protein is associated with hundreds of sites along the male X chromosome but not the female X chromosome (Kuroda et al., 1991). This same association is also seen with Msl-1 and Msl-3 proteins and is associated with an increase in acetylated histone H4 (Baker et al., 1994). This is in agreement with a direct role for these proteins in hypertranscription. Further support for this comes from the observation that active *Sxl* protein prevents Mle protein binding to the female X and that *msl-1, msl-2* and *msl-3* are all required for this binding to take place in the male (Gorman et al., 1993).

It has also been proposed that Mle may act to up regulate the transcription of other groups of autosomal genes in a non sex-specific manner. This is based on evidence which identifies a temperature sensitive allele, *nap\(^t\)*, as an allele of *mle* (Kernan et al., 1991). At the restrictive temperature, *nap\(^t\)* causes paralysis in male and female flies. It is thought to do this by causing a down regulation of transcription from the
para locus which encodes a sodium channel protein believed to be involved in the creation of action potentials. This function of mle is genetically separate from its function in dosage compensation and may represent a redundant function which is primarily carried out by other homologous genes, but the localisation of Mle protein to autosomal polytene arms indicates that there may be an alternative function for Mle in regulating the expression of certain autosomal loci (Kuroda et al., 1991; Kernan et al., 1991; Gorman et al., 1993).

Based on this evidence it seems likely that a function of active Sxl in the female is to inactivate X chromosome transcriptional hyperactivation by inactivating the pathway which includes the products of the msl-1, msl-2, msl-3 and mle genes. Analysis of msl-2 transcripts has shown that, in females, active Sxl protein acts to prevent the removal of a leader sequence in the msl-2 transcript which may prevent active Msl-2 protein from being produced, since Msl-2 is only produced in male flies (Zhou et al., 1995). In contrast, the transcripts from the msl-1, msl-3 and mle genes are not sex-specific (Palmer et al., 1994; Baker et al., 1994). Therefore, msl-2 is the target of Sxl in its dosage compensation function.

1.2.2 SEX-LETHAL DETERMINES SEX.

Clones homozygous for a Sxl null allele can be introduced into flies by use of X-ray induced mitotic recombination during embryogenesis (Sánchez & Nöthiger, 1982). As shown in table 1.2, the cells comprising these clones become masculinised despite their XX genotype. It was also observed that XO clones in gynandromorphic flies can be transformed from male-like to female-like by the presence of a hypermorphic Sxl Allele (Cline, 1979). Taken together with the fact that male flies show no requirement for Sxl, these results show that active Sxl product is responsible for the determination of cells to follow the female pathway of differentiation and that in males the absence of active Sxl directs the embryo down the male (default) pathway. Since the X:A ratio is the primary signal which determines sex and since
Sxl lies at the head of the hierarchy of sex determining genes, it seems likely that transduction of the X:A ratio occurs via the activation of the Sxl gene.

1.2.2.1 Activation of Sex-Lethal: Zygotic and Maternal Factors.

Sxl activation was originally proposed to require the presence of X-linked 'numerator' elements. These would exert either an activatory or repressive effect upon Sxl. The presence of an equal number of X chromosomes to sets of autosomes would then modulate this effect to permit Sxl activation. This could occur in a number of ways. Initially it was thought possible that these elements were non-coding and acted as binding sites for autosomally-encoded repressors (or 'denominators') which were 'watered down' by the extra X-linked sites provided in a fly with an X:A ratio=1 (Chandra, 1985). However, it is also feasible that the numerator elements encode proteins which act in some way to suppress the negative effects of autosomally encoded repressors. Also, a role for activation of Sxl by virtue of an extra copy of Sxl together with co-operative effects and possible autoregulation of the numerators would be feasible.

Two X:A numerators were identified on the basis of duplications and deletions which are sex-specific lethal (Cline, 1986; 1988). These are the sisterless loci, sis-a and sis-b. The sis-a gene encodes a protein with homology to the basic leucine zipper (bZip) family of transcription factors (Erickson and Cline, 1993). Initially, sis-a was uncovered as a recessive allele which was selectively lethal to females and caused an increase in the amount of male tissue present in 2X:3A mosaic intersexes. It was also found that a duplication of this locus was lethal to males but only if they carried two wild-type copies of Sxl. The sis-b gene encodes a protein containing a helix-loop-helix amphipathic dimerisation domain and a basic DNA-binding region (basic HLH domain, or bHLH) (Cline, 1988; Torres & Sánchez, 1989). This locus was isolated as a duplication which was also lethal to Sxl/Sxl males and resulted in an increase in the female character of 2X:3A intersexes. It was observed that the two loci had a synergistic relationship and that deletion of one copy of each of these loci
resulted in female lethality. These double heterozygote flies exhibit sterility due to ovarian tumours. This phenotype is also seen in germline clones homozygous for a Sxl null allele (Schüpbach, 1985). The above observations indicate that these elements act in a dosage dependant manner to activate the Sxl gene. Another numerator element, sis-c, has also been identified but has much weaker effects than either sis-a or sis-b (Cline, 1993).

Another locus was found to have a role in Sxl activation, being required both maternally and zygotically (Steinmann-Zwicky, 1988). This is the liz locus (previously named both fs(1)1621 and sans-fille). In gynandromorphic clones of XO cells carrying a liz duplication, an increase in the female character of the cells was observed (Steinmann-Zwicky & Nöthiger, 1985). The opposite was seen in such clones carrying a deletion covering this locus. The effect of deleting one copy of liz in female flies heterozygous for a Sxl null allele is to cause a degree of masculinisation and also to decrease the viability of the flies by an amount dependent upon the maternal genotype at this locus. This indicates that maternal liz is required for activation of the dosage compensation function of Sxl. Zygotic liz, however, would appear to have its role in the activation of the sex-determining function of Sxl. The protein encoded by liz bears extensive homology to certain human SnRNP proteins and so is likely to be involved in the positive regulation of Sxl-mediated RNA processing (Flickinger & Salz, 1994). The proposal that Sxl is epistatic to these elements is further supported by two pieces of evidence. Firstly, the masculinisation and lethality of females double heterozygote for both liz and Sxl null alleles is rescued by a constitutive Sxl allele and secondly, the male lethality of a sis-a and sis-b duplication is rescued by a Sxl null allele (Steinmann-Zwicky, 1988; Cline, 1988).

In addition to the zygotic requirement for numerator elements, there is also a maternal requirement for Sxl activation. As well as maternal liz, two genes have been identified as temperature sensitive mutations that selectively kill only the daughters of female flies carrying the allele. These have been named daughterless (da) and
Daughter-killer (Dk) (Bell, 1954; Steinmann-Zwicky et al., unpub. in Slee & Bownes, 1990).

Daughters from Dk mothers do reach adulthood if they are of the genotype 2X:3A, presumably due to the alleviation of dosage compensation problems that this genotype permits. These daughters do, however, consist of much more male-like tissue than is normally the case, indicating a reduction in the efficiency of the female sex determination pathway. That this reflects a loss of Sxl activation is shown by the fact that lethality is rescued by a Sxl constitutive gain-of-function allele. However, Dk may well act via other genes, since temperature shift experiments show that the requirement for Dk product ends before the end of oogenesis when Sxl is still dependent upon positive regulation from other genes. The effects of Dk are pleiotropic as it can also cause lethality in males (Steinmann-Zwicky et al., unpub. in Slee & Bownes, 1990).

Triploid (2X:3A) embryos from da/da mothers have much more male tissue than is normally seen in intersexes of this type (Cline, 1983). It appears that this is caused by a reduction in efficiency of Sxl activation, as demonstrated by the following evidence. Firstly, daughters from da/da mothers have unusual Sxl gene expression patterns, exhibiting a greatly increased concentration of male-specific Sxl transcripts (Maine et al., 1986). In addition, temperature shifts using a temperature-sensitive da allele have shown that da product is required from the mother until the end of the blastoderm stage when Sxl is thought to commence autoregulation (Gergen, 1987). Perhaps the most conclusive evidence for a role for da in Sxl activation comes from the observation that daughters of da/da mothers no longer die if the daughters carry a Sxl constitutive gain-of-function allele (Cline, 1983). Also, the male lethality caused by a sis-a and sis-b duplication, which can be rescued by a null Sxl allele, can also be rescued by a maternal da/da phenotype (Cline, 1988). Thus, Sxl is epistatic to da and Dk which are required for Sxl activation in females.
The *da* gene encodes a number of overlapping transcripts with apparently redundant functions (Cronmiller *et al.*, 1988). Sequence analysis of cloned cDNA's has revealed that the predicted *da* protein bears homology to proteins of the helix-loop-helix (HLH) class, notably *myc*, *MyoD1* and the *T3, T4* and *T5* *achete-scute* complex gene products (Caudy *et al.*, 1988; Murre *et al.*, 1989). These proteins contain a HLH amphipathic helix dimerisation domain and in some cases a basic DNA binding domain. The Da protein is of the bHLH type. The *myc* gene product has been implicated in both transcriptional regulation and RNA processing.

The bHLH proteins are believed to bind to DNA as dimers where they act to modulate gene expression. The potential for a great deal of regulatory variety is inherent in this type of protein as it may function as a dominant negative by having a non-functional or inappropriate DNA binding domain or it may serve several different functions by interacting with other proteins to form heterodimers which then go on to modulate gene expression in different ways. In this respect, it is interesting to note that zygotic *da* is also a proneural gene where it acts to determine between epidermal and neuronal cell fates by positive and negative interactions with the HLH proteins encoded by the genes *extramacrochaetae (emc)*, *hairy (h)* and the genes of the *achete-scute* complex. Also, it has been determined that the *sis-b* numerator element is, in fact, the *T4* locus of the *achete-scute* complex (Erickson & Cline, 1991). In early neurogenesis, the *T4* gene product is acted upon negatively by both the *h* and *emc* gene products, and the *achete scute* and *da* gene products positively regulate each other (Vaessin *et al.*, 1990). This clearly lends weight to the model of the numerator elements as encoding positive regulators of *Sxl* which act together with maternal *da* protein to overcome repressive activities encoded by autosomal denominators. It has been shown that Sis-b protein forms complexes with Da protein in the *Drosophila* embryo (Deshpande *et al.*, 1995). The same study showed that Sis-b protein is expressed prior to *Sxl* activation and that Sis-b enters that nucleus just before *Sxl* transcription begins. In addition, by use of a hsp70/*sis-b* fusion it was shown that *sis-b* is lethal to males due to expression of *Sxl* in these flies (Torres & Sánchez, 1991). Neither *T3* nor *T5* *achete-scute* loci appear to
function as strong numerators, as shown by anomalous expression of these loci. Indeed \( T4 \) appears to be the major transcript of the complex, since \( T3 \) and \( T5 \) are only poorly expressed by comparison and only activate \( Sxl \) very weakly (Parkhurst et al., 1993).

If *hairy* is ectopically misexpressed at around the same time as \( Sxl \) initiation, female lethality results (Parkhurst et al., 1990). The observation that this lethality can be rescued by duplications of the \( sis-a \) or \( sis-b \) elements points towards the possibility that autosomally encoded HLH proteins may act to antagonise the \( sis-b \) function in \( Sxl \) activation, since this is the type of protein-protein interaction in which Hairy is involved during neurogenesis. The pan neural gene *deadpan* (*dpn*) has been proposed as such a denominator (Younger-Shepherd et al., 1992). *deadpan* encodes a bHLH protein which is dependent upon zygotic \( da \) product for its expression in neuroblasts. It was seen that an inappropriately low dosage of \( dpn \) loci relative to \( sis-b \) loci caused male lethality while the opposite imbalance caused female lethality. The non-HLH protein encoded maternally by the *groucho* gene interacts with Dpn and these two proteins may act together as \( Sxl \) repressors (Paroush et al., 1994). It was also seen that male viability was reduced by maternal mutations of *emc*, another HLH protein-encoding gene which is expressed at around the blastoderm stage. The male lethalities discussed above were rescued by a \( Sxl \) null allele, showing that \( dpn \) and *emc* act to repress \( Sxl \). This suggests that maternal *emc* in the male performs the converse function to maternal \( da \) in the female, acting to repress \( Sxl \).

The model for \( Sxl \) activation which presents itself is one whereby positively acting maternal and numerator-encoded HLH proteins interact with repressive autosomal denominator-encoded HLH proteins to assess the X:A ratio and either activate or repress \( Sxl \). However, the observation that the segmentation gene *runt* also acts as a numerator shows that not all proteins involved in this process are of the HLH class (Duffy & Gergen, 1991). The mode of action of *runt* is different from that of the other numerators. Female embryos which are mutant for *runt* still express \( Sxl \) at the termini of the embryo. It has been proposed that Runt may act indirectly upon \( Sxl \) to
regulate its expression in a spatially restricted manner. However, the mode of action of the other two numerators, sis-a and sis-b is most likely to be the transcriptional activation of Sxl.

The zinc finger-containing protein encoded by the hermaphrodite (her) locus also appears to be required maternally for Sxl activation (Pultz et al., 1994; Pultz & Baker, 1995; Ryner & Swain, 1995). The her locus was identified as a recessive mutation which caused reduced viability and reciprocal transformations in both sexes, with the severity of both of these effects dependent upon the maternal her genotype (reviewed in Baker & Belote, 1983). Zygotic her mutations have no effect upon dsx expression and cannot be rescued by ectopic expression of either Sxl or tra. This suggests that the zygotic function of her lies either downstream of, or parallel to dsx.

1.2.2.2 Activation of Sex-Lethal is a Two Part Process

In order to understand the mechanism of Sxl activation, it is helpful to first appreciate that it is a two-step process yielding two slightly different sets of transcripts-the early and late/adult transcripts (Bell et al., 1988). The first set appears early on in the embryo at around nuclear division 8 and disappears after cellularisation. Sxl alleles which carry a late transcript-specific deletion show adult Sxl function defects but can be rescued by a Sxl allele which specifically lacks early Sxl functions (Maine et al., 1986; Cline, 1986). This suggests that the early Sxl transcripts may direct dosage compensation but that the late transcripts are required for the sex determining function of Sxl. The transcripts which lead to sex determination appear before the disappearance of the early transcripts and are present throughout the life of the fly (Samuels et al., 1991). Maps of the exons included in these transcripts are shown in figure 1.7.
Figure 1.7
Exon/Intron structure of Ssy early and late transcripts. Exon numbers are shown. Sizes of spliced transcripts are indicated in kilobases. Open reading frames are shown as shaded areas and the number of amino acids encoded by each reading frame is also shown. Transcripts which are sex-specific are indicated.
The first part in the process to activate Sxl is a result of the reading of the X:A ratio by the numerator/denominator elements and maternally-encoded factors such as Da and Dk. That Da is a factor in the production of the early Sxl transcripts is shown by low levels of these transcripts in XX embryos from da/da mothers (Maine et al., 1986). These transcripts are entirely female-specific. As can be seen in figure 1.7, the early transcripts are initiated from a separate promoter (P_E) which is downstream of the late promoter (P_L) from which the late transcripts are initiated. Lac-Z reporter gene fusions driven by P_E provide stage-specific expression of β-galactosidase in wild type embryos but not in embryos from da/da mothers (Keyes et al., 1992). Similar reporter gene studies have shown that the products of the da, sis-a, sis-b and runt genes all activate transcription from P_E, while the product of the deadpan gene is a P_E-repressor (Estes et al., 1995). The P_E promoter contains 19 repeats of the CANNTG consensus known to act as a binding site for bHLH proteins. A cluster of six of these repeats at the proximal end of P_E is essential for reporter gene expression in the assay described above (Estes et al., 1995).

At around cellularisation, transcription form the P_L promoter begins and this is maintained throughout development and into adulthood (Maine et al., 1986). It is not known what is responsible for expression of the late transcripts but it is not dependent upon the sex determination hierarchy as it is also transcribed in the male slightly after transcription begins in the female. What is under the control of the hierarchy is the sex-specific splicing of exon 3 from the late transcripts (Bell et al., 1988). This exon contains translational stop codons and is not spliced out of the mRNA in the male which results in the production of a non-functional truncated protein. In the female, the splicing out of this exon allows production of a large, functional Sxl protein containing two domains which bear homology to RNA binding proteins (Bell et al., 1988). It has been observed that temperature sensitive Sxl alleles which lack the 5' end of the gene (including the late promoter) can complement another Sxl allele which lacks sex determining function and usually only produces male-specific late transcripts (Keyes et al., 1992). This indicates that
the production of the sex determining transcripts (i.e. the female-specific late transcripts) depends upon the early transcripts produced initially in the embryo.

The factor responsible for the splicing out of Sxl exon 3 in the female is in fact Sxl protein itself (Sakamoto et al., 1992). Drosophila cell culture cotransfection experiments using constructs producing active Sxl protein and minigene constructs containing Sxl genomic DNA between exons 2 and 4 have shown that Sxl protein is responsible for the exclusion of exon 3 in the processed RNA molecule. Various Uridine sequences have been identified in the vicinity of exon 3 (figure 1.8).

![Diagram showing approximate locations of uridine-rich sequences around the sex-specific splice site of the Sxl transcript.](image)

Figure 1.8
Diagram showing approximate locations of uridine-rich sequences around the sex-specific splice site of the Sxl transcript.

It has been observed that deletion of several of these U-rich motifs prevents the splicing function of Sxl and this function can be restored by replacing the deleted sequences with synthetic uridine oligonucleotides (Sakamoto et al., 1992). Sxl protein has been purified (Samuels et al., 1994). This protein has been used for gel-shifts, UV crosslinking and footprinting experiments which directly demonstrate that Sxl binds to polyU runs in RNA. Wang and Bell, 1994, have shown that binding of Sxl to polyU runs is a cooperative process, mediated by the amino-terminus of the protein.

Germline transformation with Sxl minigenes has been used to study the precise sequences which are required for Sxl-mediated splicing of exon 3 (Horabin &
Schedl, 1993a; 1993b). Somewhat surprisingly, deletion of the polyU runs downstream of exon 3 disrupted exon3-splicing much more than deletion of the upstream runs. In addition, when the donor splice site at the 3' end of exon 3 is deleted, exon 3 is not removed. This suggests that the major site of Sxl-mediated 'blocking' is at the donor splice site of the intron between exons 3 and 4. However, a model which involves the sequestration of the whole region between exons 2 and 4, mediated by cooperative Sxl binding to polyU runs in the introns, cannot be ruled out at this stage.

The above system enables us to explain the observations previously made with respect to the timing and cell autonomous nature of the sex determination decision, as described below.

Sex of the fly is not determined until cellularisation. This is shown by the fact that in gynandromorphs, XO clones develop correctly with correct dosage compensation (i.e. hypertranscribed) and sex (i.e. male), only if the second X chromosome is lost before cellularisation (Sánchez & Nöthiger, 1983). XO clones which are generated by loss of the X chromosome after cellularisation exhibit dosage compensation and sexual characteristics characteristic of XX cells, which causes these clones to be small and of low viability. However, these clones can be rescued by a Sxl null allele, indicating that the source of the autonomy lies with the Sxl gene (Sánchez & Nöthiger, 1983). Perhaps the most definitive demonstration of this comes from XX flies which have been manipulated such that Sxl is less likely to be activated, where patches of tissue develop which are either male or female in nature but never a mixture of the two (Cline, 1985).

The autonomy of the sexual choice at the cellular blastoderm stage is demonstrated by several phenomena. A number of structures such as the male foreleg sex combs have been examined in triploid intersexes and it has been determined that a decision is made at the level of the individual cell to produce either the male or female version of the structure and that this decision is then carried through to completion.
(Hannah-alava & Stern, 1957). The variation in clone size seen in triploid intersexes clearly reflects the precise stage at which the progenitor cell of the clone becomes determined, in the same way as the earliness of the loss of the ring X in a gynandromorph is in proportion to the size of the resulting XO clone (Cline, 1984).

The above phenomena can be explained as follows. In females, the X:A ratio is transduced by the numerator, denominator and maternal loci, as discussed above, such that the proportions of these gene products, together with the double dose of the Sxl gene, allows initiation of transcription from the Sxl P_E promoter. It is highly likely that the bHLH Da and Sis-b proteins, together with the Sis-a, Sis-c and Runt, act directly to overcome the repressive effect of Deadpan and Groucho and enhance expression from this promoter. Initiation from this promoter prevents the inclusion of exon 3 which allows the production of active Sxl protein. Thus, when transcription from the late promoter commences, Sxl protein produced from the early transcripts prevents the inclusion of exon 3 into Sxl late transcripts. In this way, an autoregulatory loop is set up that ensures that Sxl protein is continually produced and inherited by progeny cells. Thus, the sex determination decision is set to the female pathway in each of these progeny cells. In the male, active Sxl protein is not produced, due to the fact that the initial transcription of Sxl from P_E is not activated due to the X:A ratio not permitting a sufficient concentration of the relevant activator proteins with respect to copies of the Sxl locus. A summary of this model for Sxl sex-specific splicing is shown in figure 1.9(a) and 1.9(b) (upper panels).
Figure 1.9 (a)
Sex-specific alternate splicing of Sxl, tra and dsex transcripts in the female (taken from MacDougall et al., 1995). Exons are indicated by boxes with open reading frames shaded. Proteins involved in the splicing regulation and produced by it, are shown. Of the splicing apparatus, only the U2AF splicing factor is shown. DSX=female-specific protein, DSXM=male-specific protein, SXL=active Sxl protein, TRA=active Tra protein, TRA-2=Tra-2 protein.

In the female, Sxl prevents inclusion of exon 3 in the Sxl mRNA, allowing more active Sxl protein to be produced. Sxl also blocks the acceptor splice site at the 5' end of exon 2 in the tra transcript, causing splicing which produces a tra mRNA encoding a full-length active Tra protein. Tra and Tra-2 promote usage of the dsex exon 4 acceptor site, resulting in a female-specific dsex transcript.

(Figure 1.9 (b) overleaf.)
Figure 1.9 (b)

Sex-specific alternate splicing of Sxl, tra and dsx transcripts in the male (taken from MacDougall *et al*., 1995). Exons are indicated by boxes with open reading frames shaded. Proteins involved in the splicing regulation and produced by it, are shown. Of the splicing apparatus, only the U2AF splicing factor is shown. DSX\(^{\text{f}}\)=Dsx female-specific protein, DSX\(^{\text{m}}\)=Dsx male-specific protein, SXL=active Sxl protein, TRA=active Tra protein, TRA-2=Tra-2 protein.

In males, lack of early Sxl protein enables default splicing which results in Sxl and tra mRNAs which encode non-functional, truncated proteins and a male-specific Dsx protein.
The cell autonomy of the sex determination decision is analogous to that seen in position effect variegation (PEV) in *Drosophila*. In PEV, chromosomal rearrangements bring certain genes close to terminal or centric heterochromatin. This results in a mosaic expression pattern reflecting a cell-autonomous decision to either express or not express the gene (Tartof *et al.*, 1989; Henikoff, 1990). In the case of the *white* gene, this results in the eyes of the fly having a mottled appearance caused by the mosaic expression of the gene in this tissue. Perhaps the most likely model for this effect is that the ON/OFF decision is affected by a flowing of heterochromatin across the locus in some cells but not in others. The state of the gene would then be 'locked' at some point in development via the action of certain non-histone chromosomal proteins (NHC's). Evidence for this includes the reduction of PEV clone by the addition of heterochromatic material (i.e. extra copies of the Y chromosome) to the cell, presumably due to a sequestering of NHC's by this material (Dimitri & Pisano, 1989). Also several suppressor mutations of PEV have been isolated and have been identified as encoding proteins important in the formation and maintenance of heterochromatin, such as the *suvar(2)5* gene which encodes a protein containing the 37aa 'chromodomain' which has been implicated in the binding of proteins such as Heterochromatin associated protein-1 (HP-1) to chromatin (Paro & Hogness, 1991).

Another phenomenon similar to the cell-autonomy of the sex determination decision is seen in the homoeotic gene system. Here, the arrangement of the genes in the antennapedia complex (ANT-C) and bithorax complex (BX-C) of homoeotic loci is collinear with the segments of the fly in which they are expressed. It is proposed that this pattern of expression is brought about by a spreading of heterochromatin along the chromosome to an extent determined by the segment of which each cell is part (Gaunt & Singh, 1990). This pattern of heterochromatic spreading is proposed to be locked into position by the action of the polycomb-group and trithorax-group genes some of which show homology to the proposed NHC's which are involved in PEV (Paro, 1990). Notably, the Polycomb protein itself also contains the chromodomain discussed above (Paro & Hogness, 1991).
Another system which shows a similar two-step initiation of an autoregulatory loop leading to cell autonomous inheritance of a determined state is the maintenance of the lysogenic state by bacteriophage lambda (Ptashne, 1987). The first stage is the production of the transcription factor CI (perhaps analogous to sis-a, sis-a, da and Dk). CI initiates transcription from the $P_{RE}$ promoter (analogous to the $Sxl P_E$ promoter) which drives transcription of the $cl$ repressor gene (analogous to $Sxl$) and from $P_1$ which drives transcription from the lysogenic gene $Int$. As well as shutting down all other genes, the CI repressor also maintains transcription of itself by driving transcription from a second promoter $P_{RM}$ (analogous to $Sxl$ protein splicing out the translational stop codon containing exon from its own hnRNA). It is perhaps unsurprising that systems such as these are used (even in such evolutionary divergent systems) to provide maintenance of a determined state.

It is also interesting to note that in the case of CI repressor protein, cooperativity of binding to the adjacent target sequences lends the system a high degree of sensitivity to fairly small changes in protein concentration thus enabling the lytic/lysogenic 'switch' to be 'thrown' in response to small concentration changes; a quality that is essential to the temperate character of phage $\lambda$ (Ptashne, 1987). The HLH nature of Sis-b and Da present us with the possibility that a similar cooperativity may be occurring here, via dimerisation, to enable the two-fold change in dosage of the $Sxl$ locus to effect a large increase in the ability of these proteins to drive transcription from $P_E$.

The next question that presents itself is; what happens after $Sxl$? We know that $tra$ and $tra-2$ are epistatic to $Sxl$ and so we would expect the regulation of these proteins to be dependent upon $Sxl$. 

43
1.2.2.3 The Next Step - \textit{transformer and transformer-2}.

\textit{tra} and \textit{tra-2} mutant alleles which cause XX flies to be transformed into intersexes have no effect on the soma of XY flies (table 1.2), indicating that these genes are only active in the female soma.

In females, both \textit{tra} and \textit{tra-2} are likely to be required throughout development to enhance female functions and repress male characteristics. This is shown by studies involving the introduction of homozygous-mutant cell clones via mitotic recombination, showing that \textit{tra} and \textit{tra-2} are needed until early pupation in order for the correct female pigmentation to be seen in segments 5 and 6 (Baker & Ridge, 1980). These genes are also required for correct female differentiation of the foreleg. However, in this case the requirement ends at around two days before pupariation. It is thought that the different temporal requirements reflect the different rates of cell division in the progenitor cells of these two tissues. The \textit{tra-2} gene has also been shown to function at different times within the same structure to direct different aspects of female differentiation. For example, in the foreleg, the female number of certain bristles is determined by \textit{tra-2} up to a day before the determination of the female-specific structural aspects of the same bristles (Belote & Baker, 1982). This type of observation may represent the interaction of \textit{tra-2} product with other proteins which are temporally restricted.

The cloning of \textit{tra-2} has revealed that it encodes four transcripts which are alternately spliced as shown in figure 1.10 (Mattox & Baker, 1991).
The *tra-2* transcripts potentially encode proteins with a common C terminus which contain homology to known RNA binding proteins (Goralski *et al*., 1989). We can see from figure 1.10 that two of the transcripts encode 226 amino acid proteins containing two RS (arginine/serine-rich) domains and one RNP (ribonucleoprotein) domain. The RS domain is common to splicing factors such as U2AF (Zamore *et al*., 1992; Zhang *et al*., 1992), SF2/ASF (Ge *et al*., 1991) and Suppressor of white apricot (Su(Wa); Chou *et al*., 1987) and is vital for the splicing function of such factors (e.g. Valcárcel *et al*., 1993; Li & Bingham, 1991). The RNP domain is also common to splicing factors and mediates RNA binding (Amrein *et al*., 1994). The RS motif is thought to mediate protein-protein interactions between splicing factors (Wu & Maniatis, 1993). The 264 amino acid Tra-2 isoform also contains all three of these motifs. However, the male germline-specific 179 amino acid isoform contains only the RNP domain and RS2. The fact that there are no female-specific somatic transcripts and that no epistasis has been established between *tra* and *tra-2*, together with the fact that *tra-2* has a positive role in spermatogenesis suggests that *tra-2* actually lies on a side branch of the sex determination hierarchy.
The cloning of \textit{tra} has shown that it is indeed part of the sex determination hierarchy, lying directly downstream of \textit{Sxl} (Boggs \textit{et al.}, 1987). As would be expected for a gene regulating determination of tissues, maximal \textit{tra} expression occurs during the pupal stage. Two transcripts are present in the fly as shown in figure 1.11.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure11.png}
\caption{Diagram of the sex-specific transcripts produced from the \textit{tra} locus. Exons are indicated by boxes with open reading frames shaded. Sizes and sex-specificities are indicated. See text for details.}
\end{figure}

It is apparent that both of the \textit{tra} transcripts are initiated from the same promoter. We can also see that \textit{tra} is again subject to sex-specific RNA processing. In the female, the choice of a downstream splice acceptor site prevents the inclusion of a translational stop codon which enables the production of the 211aa active Tra protein. The observation that the area around the non sex-specific splice site contains a uridine octamer sequence indicates that \textit{tra} may be directly under the control of \textit{Sxl}. Transformation experiments have indeed shown that \textit{Sxl} is responsible for the sex-specific splicing of \textit{tra} primary RNA and that the uridine octamer is required for this regulation (Sosnowski \textit{et al.}, 1989). This was shown by the use of various constructs containing regions of \textit{tra} genomic DNA, driven by the hsp70 promoter, which were introduced into \textit{tra/tra} flies via germline transformation. Deletion of the non sex-specific splice site led to a degree of \textit{Sxl}-independent feminisation of the male as would be expected if the function of \textit{Sxl} is to prevent the use of the deleted splice site. Deletion of the sex-specific splice site had similar effects to the deletion
of the U-octamer, namely female accumulation of unspliced RNA, a lack of female-specific RNA and an inability of this construct to either rescue tra/tra females or to transform males.

The mechanism by which Sxl blocks the non sex-specific splice site has been shown to be via the antagonism of the essential splicing factor U2AF which binds to the same U-rich sequences as Sxl protein. Sxl protein, however, lacks the splicing 'effector' sequences (arginine-serine repeats, or RS repeats) which are present in U2AF (Zamore et al., 1992; Zhang et al., 1992). If this effector domain is introduced into the Sxl protein, it becomes constitutively active as a splicing factor, causing splicing from the same splice site which it would normally blocks (Valcárcel et al., 1993). A summary of Sxl-mediated tra sex-specific splicing is shown in figure 1.9(a) and 1.9(b) (middle panels).

The pleiotropic viriliser (vir) gene has been implicated in Sxl-mediated functions (Hilfiker & Nöthiger, 1991). A number of recessive alleles of this locus exist, some of which are lethal to both sexes and some of which cause XX flies to develop as true intersexes. Also, males carrying a constitutive gain-of-function Sxl allele which would normally be lethal, are rescued to viable males (i.e. without sex transformation) by the female-lethal vir12 allele. Thus, vir is epistatic to Sxl and so is likely to lie downstream of Sxl, playing a role in both sex determination and dosage compensation. This is supported by the fact that certain msl and mle mutations partially rescue the lethality of vir12. Vir does, however, lie upstream of tra as shown by the rescue of XX flies from vir12phenotypes, by ectopic expression of the female form of tra. XX/vir12 flies (viability achieved via a mle null allele) produce the male-specific transcripts of both tra and Sxl, indicating that Vir may have a role in Sxl-mediated splicing regulation (Hilfiker et al., 1995). XX/vir2f flies which also carry a constitutive gain-of-function Sxl allele, while viable, are largely male-like, morphologically. These flies splice both Sxl and tra transcripts predominantly in the male-specific mode. This suggests that Vir is required for Sxl to perform its splicing-modulation function. It is believed that the true intersexual phenotype of
some vir alleles reflects an intermediate dosage of gene products from genes below vir in the hierarchy, causing both Dsx\(^M\) and Dsx\(^F\) to be produced in the same cell.

The genes fl(2)d and liz also appear to be important in regulation of Sxl-mediated splicing (Granadino \textit{et al.}, 1990; Albrecht & Salz, 1993). XX flies carrying mutations for either of these genes splice Sxl transcript in the male-specific mode. A constitutive gain-of-function Sxl allele rescues the mutant phenotype of both of these genes (Steinmann-Zwicky, 1988; Salz, 1992; Granadino \textit{et al.}, 1992). There is also a Sxl-independent role for fl(2)d as shown by the non sex-specific lethality of certain alleles (Granadino \textit{et al.}, 1991). This latter function would appear to only be required early on in development as indicated by the fact that adult males carrying a fl(2)d temperature sensitive allele are unaffected by temperature shifts. Also, X-ray induced fl(2)d-mutant clones are totally viable in the male, when these clones are induced in larvae, indicating that the gene must act before then to perform its vital non sex-specific function.

The next gene in the hierarchy is dsx. It is attractive to suppose, bearing in mind the above evidence, that tra and tra-2 products may act upon dsx to cause a sex-specific splicing pattern which in turn leads to the production of proteins which implement the female differentiation pathway. In support of such a model for tra function, the predicted Tra protein has the RS splicing effector domain seen in U2AF (Boggs, \textit{et al.}, 1987). If the RS domain from the splicing factor suppressor of white apricot (Su(W\(^a\))) is replaced by the Tra RS domain, the Su(W\(^a\)) protein retains its RS domain-dependent function (Li & Bingham, 1991). This suggests that Tra protein may also function as a splicing factor.

1.2.2.4 doublesex and intersex - The Sex Determination Effector Genes.

As stated previously, flies mutant for the dsx and ix genes are true intersexes which is to say each individual cell is of intermediate sexual phenotype. This in itself indicates that these genes are involved in the activation/repression of those genes
which cause differentiation of male and female characteristics. This is further suggested by their epistasis to every other gene of the hierarchy. In other words, it is thought that it is at the stage of \( dsx \) and \( ix \) action that determination of the sexual state becomes differentiation of that state.

It is not clear exactly where \( ix \) fits into the hierarchy. It may lie downstream of \( dsx \), as \( ix/ix \) flies still produce the active female \( dsx \) transcript and both \( ix \) and \( dsx \) are required for correct female development (Baker & Ridge, 1980; Chase & Baker, 1995). However, the possibility that \( ix \) lies on a side branch cannot be discounted. This model is supported by the proposal that \( ix \) may be active in males to direct some aspect of pheromone production. This is based on the observation that \( ix/ix \) males appear to attract other males (Tompkins, 1986). However, \( ix \) male-specific function has not been shown conclusively.

The fact that null \( dsx \) alleles affect both sexes while certain other alleles affect only males or females, tells us that \( dsx \) must be differentially active in both sexes. The cloning of the gene has enabled a developmental profile of transcripts produced from \( dsx \) that is in agreement with this mode of action (Baker & Wolfner, 1988). This developmental profile is shown in table 1.3.

<table>
<thead>
<tr>
<th></th>
<th>Embryonic &amp; Larval Stages</th>
<th>Pupa</th>
<th>Adult Fly</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>1.65kb</td>
<td>2.9kb</td>
<td>2.9kb</td>
</tr>
<tr>
<td></td>
<td>2.8kb</td>
<td>3.9kb</td>
<td>3.9kb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.7kb</td>
</tr>
<tr>
<td>FEMALE</td>
<td>1.65kb</td>
<td>3.5kb</td>
<td>3.5kb</td>
</tr>
<tr>
<td></td>
<td>2.8kb</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1.3**
Developmental profile of \( dsx \) transcripts. Sizes of \( dsx \) transcripts detected throughout development are shown in kilobases.
The 1.65kb and 2.8kb transcripts share 3' and 5' ends but differ in the intervening sequence, showing that they are derived from the same primary RNA. The region covered by these transcripts represents a minimum of 40kb of genomic DNA (Baker & Wolfner, 1988). This is comparable to the 23kb Sxl locus but is much greater than the very small 2kb tra gene (Boggs et al., 1987).

The cloning of dsx cDNA's representing the 3.9kb and 3.5kb transcripts (Burtis & Baker, 1989) has demonstrated how the dsx gene can be differentially active in both sexes (figure 1.12). The transcripts shown in figure 1.12 are differentially spliced and polyadenylated but are both capable of producing large proteins. Hence it would be quite feasible for there to be differential activity in both sexes.

![Figure 1.12](image)

**Figure 1.12**
Intron/exon structure of the two major adult sex-specific dsx transcripts. Exons are shown as boxes, with open reading frames shaded. Transcript and predicted protein sizes are indicated.

It has recently been shown that there is a common region of DsxM and DsxF which is responsible for binding to the regulatory regions of downstream genes which are regulated directly by Dsx (Erdman & Burtis, 1993). The common N-terminal domain of these proteins carries a putative α-helical domain which is required for binding activity as shown by gel-shift assays with Dsx proteins lacking this region. Site directed mutagenesis has shown that certain cysteine residues next to this region are also critical for DNA binding and that this binding is EDTA sensitive. Also, spectrophotometric detection has shown that observed binding of divalent metal cations is reduced when these residues are mutant. The binding site has been classed
as a novel zinc finger type as it bears no homology to previously identified zinc finger motifs.

It would be a natural progression to suppose that the production of the Dsx<sup>F</sup> transcript is under the control of the Tra and Tra-2 proteins. In fact this turns out to be the case. Germline transformation using the female-specific tra cDNA fused to the hsp70 promoter transforms the male soma towards that of the female (Boggs et al., 1987). Direct evidence has been published that shows that tra and tra2 cause the female-specific transcript to be produced (Hoshijima et al., 1991; Ryner & Baker, 1991). Plasmid containing dsx genomic DNA from exon 3 to exon 5 (see figure 1.12) was transfected into Drosophila culture cells either alone or together with hsp70-driven female-specific tra cDNA or with hsp70-driven tra-2 cDNA, or with both (Hoshijima et al., 1991). It was observed that if Tra and Tra-2 proteins were not present, male-specific dsx RNA was produced and practically no female-specific transcript was detected. If Tra or Tra-2 was present then much more female-specific dsx RNA was produced. Finally, if Tra and Tra-2 were present then only female-specific dsx RNA was produced. A construct containing the region from exon 3 to exon 4, but not containing the female-specific polyadenylation signal, was capable of being directed by Tra and Tra-2 to utilise the female-specific splice acceptor. However, a construct lacking this splice site was spliced in a Tra/Tra-2-independent fashion. Thus, the only possible model to account for this is that Tra and Tra-2 act at the female-specific acceptor splice site, to positively promote the usage of this site.

Transgenic studies have shown that all three Tra-2 protein isoforms (figure 1.10) are capable of regulating the splicing of dsx, with the 226 amino acid and 264 amino acid proteins being the most efficient (Mattox et al., 1996). The transcripts encoding all three isoforms, under the control of their cognate promoter elements, were introduced into tra-2/tra-2 females via germline transformation and subsequent crosses. XX/tra-2 flies expressing either of the two larger Tra-2 isoforms developed as normal females and expressed the female-specific dsx transcript. The smaller 179
amino acid isoform, which lacks the RSI motif, was only capable of partially rescuing the pseudomale phenotype of XX/tra-2 flies and both male-specific and female-specific dsx transcripts were produced.

Six 13-nucleotide repeats (dsx 13-nt repeats) contained in exon 4 have been identified as being important for Tra/Tra-2-mediated dsx splicing. Deletion of these repeats causes a loss of female-specific product in the cotransfection system described above (Hoshijima et al., 1991). Also, substitution of some of the non-canonical purines present in the polypyrmidine (polyY) tract of the female-specific acceptor site causes female-specific splicing independent of Tra and Tra-2. This indicates that this site is not used in males because of its non-standard polypyrmidine stretch and that in females Tra and Tra-2 act to stabilise the splicing apparatus at this site and thus promote its use. Tra2 protein binds to RNA containing one of the dsx 13-nt repeats and a 540bp sequence containing the six 13-nt repeats is necessary and sufficient to direct Tra/Tra-2-mediated splicing of the dsx male-specific splice acceptor (Ryner & Baker, 1991).

In addition, arginine-serine repeat-containing splicing factors can be recruited by Tra and Tra-2 to the region containing the 13-nt repeats (Tian & Maniatis, 1993), and Tra and Tra-2 proteins have been shown to physically interact both with each other and with the splicing factor, SF2 (Amrien et al., 1994). The six 13-nt repeats of the dsx gene have been shown to interact synergistically with a purine-rich element (purine rich enhancer, or PRE) in the same region, to promote use of the female-specific splice site (Hedley & Maniatis, 1991; Lynch & Maniatis, 1995).

A summary of the RNA splicing regulation of the Sxl, tra and dsx genes is shown in figure 1.9 and the whole sex determination hierarchy is summarised in figure 1.13. This hierarchy controls the sex determination of the soma. However, the determination of the germline does not require the action of all these genes and it is this tissue which is discussed next.
Figure 1.13
Summary of the sex determination hierarchy of *Drosophila*. See figure 1.9 for mechanistic details.

Activation of Sxl is dependent upon the X:A ratio, due to the action of X-linked transcriptional repressors (numerators) in balance with autosomal transcriptional repressors (denominators).

In males, Sxl remains inactive and has no function. The msl-2 transcript is spliced to produce active mRNA. Msl-1, Msl-2, Msl-3 and Mle act in a single pathway to promote generalised hypertranscription of the X chromosome. The dsx primary RNA is spliced in the default mode to produce DsxM protein.

In females, Sxl is activated and is locked 'ON' by positive autoregulation. Sxl protein acts upon the msl-2 and tra primary transcripts to modulate their splicing, such that an inactive msl-2 mRNA and an active tra mRNA are produced. Tra and Tra-2 proteins act in concert to modulate the splicing of the dsx primary transcript, such that a female-specific Dsx protein, DsxF, is produced.
1.3 SEX DETERMINATION OF THE GERMLINE

1.3.1 CELL-AUTONOMOUS AND INDUCTIVE FACTORS

Initial investigations into sex determination of germ cells were done by pole cell transplantation between male and female embryos (Van Deusen, 1976). It was seen that XX gametes only developed in XX gonads and XY gametes only developed in XY gonads. This indicates that germ cells cannot develop in the gonad of the opposite sex. This means that the pole cells must, at the time of transplantation, be at least partly determined by their chromosomal sex to become either spermatogenic or oogenic, since they cannot be subverted to the developmental pathway of the opposite sex by the presence of the somatic gonad tissue of that sex. It also indicates that there must be some communication between the germ cells and the somatic cells of the gonad which allows the germ cells to differentiate male gonad somatic cells from those of the female. The death of germ cells transplanted into embryos of the inappropriate sex may to some extent be due to inappropriate X chromosome dosage. In support of this, the X-hypertranscription gene mle has been shown to have some function in the male germline (Bachiller & Sánchez, 1986). Thus, unlike somatic sex determination which involves purely cell autonomous decisions, sex determination in the germline involves a mixture of cell autonomous and inductive factors.

The autonomous nature of XY germ cells is shown by pole cell transplantations into host embryos which have been genetically manipulated to contain no germline. Both XY and XO pole cells appear to enter spermatogenesis in the ovary of an XX host, as characterised by the large nuclei, nucleoli and mitochondrial clusters typical of spermatocytes (Steinmann-Zwicky et al., 1989). Crystal structures are seen in the abortive XO cells. When XX pole cells are transplanted into a germline-deficient XY embryo they again appear to enter spermatogenesis, forming abortive sperm-like cells with the same crystal structures. Initially, this evidence was taken to indicate that male and female pole cells are autonomous to the extent that they exhibit different 'competence' to be directed down the spermatogenic or oogenic pathways,
this competence being controlled by the X:A ratio. This was indicated by the fact that while 1X:2A (X:A=0.5) germ cells can only enter spermatogenesis, 2X:3A (X:A=0.67) germ cells appeared to be able to enter either spermatogenesis or oogenesis (Nothiger et al., 1989). However, more recent experiments have required a re-thinking of this hypothesis, as described below. This work centres on studies involving two genes which have roles in female germ cell development; ovarian tumour (otu) and ovo.

Mutant alleles of the ovo gene only affect females. Germ cells are sensitive to a lack of ovo product at around the late blastoderm-early gastrula stage and will die at this stage in ovo mutant flies, giving rise to empty ovaries (Oliver et al., 1987). Chimeric flies show that this reflects a cell autonomous requirement for ovo (Perrimon & Gans, 1983). Different phenotypes have been observed for different ovo alleles, indicating that ovo is not only involved in sex determination. Some allele combinations do produce male-like germ cells but it is more likely that ovo is involved in modulating more general developmental functions in the oocyte, since mutations in sex transforming genes do not generally cause such severe death of germ cells as ovo alleles (Oliver et al., 1990). The gene has been cloned and has been found to encode a putative transcription factor containing four zinc fingers (Mével-Ninio et al., 1991).

Mutant alleles of the otu gene produce very similar phenotypes to ovo mutations (Pauli et al., 1993). A set of recent experiments on the otu and ovo genes have provided further insights into the nature of a female-specific somatic signal required for development of XX germ cells (Nagoshi et al., 1995). Several lines of evidence indicate that the transduction of the female somatic inductive signal is, at least in part, mediated by otu. When XY flies are transformed into pseudofemales by ectopic tra expression, the gonads of the flies can by thought of as a 'pseudo-ovary', containing genetically male germ cells surrounded by genetically female gonadal somatic cells. The follicle cells of such 'pseudo-ovaries' appear essentially wild-type but the germ cells are highly abnormal, having egg chambers filled with large
quantities of small, tumourous, egg cysts (Nagoshi et al., 1995). When the germ cells of a pseudo-ovary are mutant for a severe *otu* allele, no germ cells are observed in the egg chambers. If the same *otu* allele is complemented by ectopic *otu* expression, the egg chambers are again seen to be filled with egg cysts. In other words, placing a XY germ cell in a female somatic environment imposes upon it a requirement for *otu*. The same requirement is not seen for *ovo*, with egg cysts being produced in pseudo-ovaries regardless of the integrity of the *ovo* locus. In the converse set of experiments, XX germ cells developing in 'pseudo-testes' were seen to require both the *otu* and *ovo* genes for proliferation to occur. Thus, the XX germ cells maintain a cell autonomous requirement for *otu* and *ovo*, irrespective of the sex of the surrounding soma. On the strength of the above evidence, Nagoshi *et al.*, 1995, propose that the female somatic signal induces *otu* expression in the XX germ cells which, together with cell-autonomous Ovo, allows the cells to proliferate and develop.

This evidence is in conflict with previous pole cell transplantation experiments which showed that XY pole cells were not 'competent' to enter oogenesis (see above). These pole cell transplantation experiments suggested that XY germ cells developed into spermatocytes, whether they developed in a testis or in an ovary (Steinmann-Zwicky *et al.*, 1989). The XY germ cells in a pseudo-ovary were morphologically very similar to the cells that were classified as spermatocytes in the pole cell transplantation experiments (Nagoshi *et al.*, 1995). However, since the female germline-required gene *otu* is necessary for production of these cells in the pseudo-ovary, it seems that they are undergoing some aspects of oogenic development. It appears that the problem is one of classification. Germline sex determination is not controlled via a linear hierarchy in the same way as somatic sex determination. It is much more likely to require a number of parallel pathways to achieve correct gametogenesis. Therefore, mutation of a single gene may not result in a simple 'switch' between male or female developmental pathways, but rather some dimorphic phenotype which is a mixture of male and female characteristics. Thus, the cells that were classified as spermatocytes in the pole cell transplantation
experiments, may in fact be neither spermatocytes nor oocytes, but a mixture of the two.

Nagoshi et al., 1995, propose that both male and female germ cells are subject to sex-specific signals from the somatic cells of the gonad. In XX cells, a follicle cell signal would result in activation of otu which, together with ovo, would mediate certain oogenic processes. It is proposed that an analogous system operates in the testis, such that an as yet unidentified testis-specific otu-counterpart would be activated in response to a male-specific signal from somatic testis tissue. This would account for all of the above observations. A XY germ cell in an ovary, or pseudo-ovary would be subject to the inappropriate female-specific signal and otu would be activated in the cell. This would result in otu-mediated processes being activated in conjunction with cell autonomous spermatogenic processes which may result in the abortive "spermatocytes" observed. The converse argument can be applied to XX germ cells in a testis, or pseudo-testis.

1.3.2 THE FEMALE INDUCTIVE SIGNAL- PRODUCTION AND TARGETS

At present, nothing is known about the possible origins or targets of the putative male-specific signal. However, the female-specific somatic signal is under the control of the sex determination hierarchy in the follicle cells, as shown by the following evidence. XX germ cells which are mutant for each of the somatic hierarchy genes; tra , tra-2 , dsx and ix develop correctly when introduced into a wild type XX embryo (Marsh & Wieschaus, 1978; Schüpbach, 1982). Thus, these genes are not required cell-autonomously for oogenesis. However, when XX pole cells were transplanted into an XY host which was feminised by ectopic expression of a female-specific tra cDNA, it was observed that the XX pole cells entered oogenesis and indeed formed complete eggs (Steinmann-Zwicky & Niederer, unpublished, in Steinmann-Zwicky, 1992). In addition, alteration of tra or dsx expression can affect germ cell morphology (Steinmann-Zwicky, 1994).
This model is complicated by the observation that in \( tra/tra \) or \( tra-2/tra-2 \) pseudomales, abortive oocytes as well as abortive and immotile sperm can be found (Nöthiger et al., 1989). If a feminising inductive factor is produced via the somatic hierarchy, then how can oogenesis occur in \( tra/tra \) or \( tra2/tra2 \) flies? It has been proposed that some initial signal under the control of the sex determination hierarchy directs the XX pole cells down the oogenic pathway very early in development. However, these cells can still revert to spermatogenesis unless the oogenic pathway is confirmed later on in development by a further hierarchy-independent signal from the surrounding soma. Similarly, pole cells which have started down the spermatogenic pathway can be subverted by the presence of the later signal to follow the later stages of the oogenic pathway. This model is based on the morphology of XX pseudomale pole cells throughout development, as described below (Steinmann-Zwicky, 1992).

Up until the first larval instar, male and female pole cells are indistinguishable. Upon hatching, however, the male gonad is much larger than the female one, containing more germ cells. Thus, by this stage the initial feminising signal must have played its role, as the pole cells have entered one or other of the developmental pathways. In the second and third larval instars, and during metamorphosis, spermatogenic structures and oogenic structures, as well as degenerating cells, are seen. The presence of these structures shows that complete determination has not taken place by the first larval instar and so further signals must be required. Examination of the development of XX germ cells in pseudomales lacking the sex determination function of \( Sxl \) shows that \( Sxl \) is somatically required for the production of later oocyte structures such as nurse cells, in these flies (Nöthiger et al., 1989). Thus, it seems likely that \( Sxl \) controls the production of both the early and late feminising signals but through different pathways, such that the early signal is somatic hierarchy-dependent while the late signal is not.

We can now ask the question; what is the effect that the early hierarchy-dependent signal has upon XX germ cells? It seems likely that the action of the signal is to
direct Sxl to produce germline-specific Sxl transcripts. There is a number of pieces of evidence that point to this model. Firstly, there is a delay between the appearance of Sxl protein in the soma and its detection in the pole cells, as might be expected if signaling was required between the two tissues (Bopp et al., 1991). Secondly, when X/X, Sxl/Sxl germ cells develop in an XX host, any oocytes formed are cystic in nature and even spermatogenic cells are produced (Schüpbach, 1985). Conversely, XX germ cells carrying a constitutive Sxl gain-of-function allele enter oogenesis even when transplanted into a male host (Steinmann-Zwicky et al., 1989). The germline Sxl product is different from the somatic Sxl product, as indicated by the presence of Sxl alleles which cause cystic ovaries but do not otherwise affect the development of the fly, and by the fact that the 1.9kb and one of the 3.3kb Sxl late female-specific transcripts are germline-specific (Samuels et al., 1991). The 4.2kb and other 3.3kb transcripts are found both in the soma and the germline. The tissue specificity of the two 3.3kb transcripts was determined by comparing abundance in wild type and germline-deficient strains.

The orb locus encodes RNA-binding proteins and produces male and female germline-specific transcripts which are alternately spliced (Lantz et al., 1992). Analysis of the splicing patterns of Sxl and orb in various mutant backgrounds has indicated that, while tra and dsx are required for certain aspects of oogenic development, only tra-2 is required in the soma for the production of the female-specific somatic signal which directs the female germ cells into oogenesis (Horabin et al., 1995). In XX flies which are mutant for dsx, the majority of Sxl and orb germline-specific transcripts are of the female-specific class and Sxl protein is clearly detectable. However, XX flies which carry a tra-2 null allele exhibit virtually exclusive male-specific patterns of orb and Sxl expression and very little Sxl protein is detected, if any. Surprisingly, tra does not appear to be required for this tra-2-mediated function, since XX flies carrying some combinations of tra alleles exhibit mostly female-specific expression patterns of Sxl and orb, with Sxl protein being produced. Thus, it may be that Tra-2 can act in conjunction with an, as yet unidentified, cofactor other than Tra to regulate the production of the female-specific
somatic inductive signal. It appears that, while *tra* and *dsx* are required in the female soma to direct certain aspects of female germ cell development, neither gene is involved in production of the signal which sets the germ cells upon the oogenic pathway.

1.3.3 ACTIVATION OF SXL IN THE FEMALE GERMLINE

It has been shown that while *da* is not required, the gene *liz* is required for oogenesis to take place (Saltz, 1992). Female flies homozygous-mutant for *liz* develop cyst-filled ovaries. This phenotype is completely rescued by a gain-of-function *Sxl* allele and reflects an autonomous requirement for *liz*, as female flies with *liz/liz* germ cells lay no eggs. Therefore, it would appear that *liz* is required cell-autonomously for activation of *Sxl* in the germline and is probably involved in splicing regulation as previously suggested. The gene *fl(2)d* is also required for *Sxl* activation in the germline (Steinmann-Zwicky, 1993). Also, the *otu* gene has been implicated in *Sxl* activation, since some *otu* alleles produce ovarian tumours typical of *Sxl* germline-specific alleles (Pauli *et al.*, 1993). *Sxl* protein is not produced in ovaries in flies carrying *liz* or *otu* mutations (Bopp *et al.*, 1993).

Oliver *et al.*, 1993, showed that several genes which cause ovarian tumours when mutant, are required cell-autonomously in the the female germline for female-specific splicing of the *Sxl* transcripts. These include the *otu*, *ovo* and *liz* loci. One other such gene is the *fused* locus, which encodes a serine/threonine kinase and as such may be involved in transducing the follicle cell signal into the female germline-specific *Sxl* splicing pattern.

Germline clones which are mutant for the genes *sis-a*, *sis-b* and *runt*, still produce oocytes in the female (Steinmann-Zwicky, 1993; Granadino *et al.*, 1993). In the soma, these mutations result in lethality due to an inability to activate *Sxl*. This could mean that *sis-a*, *sis-b* and *runt* play no role in activation of *Sxl* in the germline or that *Sxl* expression is already set up at the time of clone induction or transplantation.
The *bag of marbles* gene is required for gametogenesis in both sexes (McKearin & Spradling, 1990). Mutant alleles of the gene result in undifferentiated cyst-like cells in ovaries and testes. In oocytes, Sxl protein shows an abnormal localisation pattern in these mutants (Bopp et al., 1993).

If the activation of Sxl is the only effect which the feminising signal produces in the female, then we would expect XY pole cells to enter oogenesis if they carried a constitutive Sxl allele. This is not the case. XY germ cells remain spermatogenic even though they carry either of the constitutive gain-of-function Sxl alleles, Sxlp^M1 or Sxlp^M4, and develop in an XX host (Steinmann-Zwicky et al., 1989). How can this be explained? Either there is some other cell autonomous property of XX or XY pole cells which is Sxl-independent and limits the inductive competence of these cells, or the Sxl alleles do not direct sufficient Sxl production in the XY pole cells. There is some evidence for the latter explanation as it is apparent that in XY flies not all somatic cells are feminised by the Sxlp^M1 allele. However, Sxlp^M4 is known to express Sxl constitutively in somatic XX and XY cells and in XX germ cells (Steinmann-Zwicky, 1993). Therefore, it appears that while Sxl activation is critical to female germ cell development, there are a number of other cell-autonomous pathways acting in parallel which are required for correct oogenesis to occur. In support of this model, several of the genes involved in female germline development can be categorised according to their phenotypes. Mutant alleles of *otu* and *ovo* cause a range of phenotypes ranging from complete loss of female germline, to sexual transformations of the germ cells, while Sxl or *liz* mutations cause sexual transformations, but have no effect on germ cell viability (Oliver et al., 1993). This indicates that *otu* and *ovo* are involved in a viability pathway as well as the Sxl activation pathway which leads to sexual transformations.

The above evidence shows that Sxl function in the germline is different from its function in somatic cells, in that Sxl does not appear to control a hierarchy of germ cell sex-determining genes but is rather a part of a pathway of genes which is only one of a number of parallel pathways required for correct oogenesis. Further support
for this comes from studies of germline markers in germ cells which have perturbed Sxl function (Bae et al., 1994; Horabin et al., 1995). For example, the orb female-specific splicing pattern is observed exclusively in ovaries, even when they lack any detectable Sxl protein (Horabin et al., 1995).

1.3.4 FUNCTIONS OF TRA-2 IN THE MALE GERMLINE

As well as a female-specific somatic requirement for tra-2, this gene also functions cell-autonomously in the male germline, since non-functional sperm are produced in tra-2-mutant flies (Belote & Baker, 1983). Indeed, most abundant expression of tra-2 is seen in this tissue and two male germline-specific transcripts are produced by the tra-2 locus as shown in figure 1.10 (Mattox & Baker, 1991). In wild type flies, the removal of the M1 intron is normally an inefficient process such that the concentration of the M1-containing transcript is higher than that of the the completely spliced transcript. In male flies carrying a tra-2 null mutation, removal of the M1 intron occurs to completion (Mattox & Baker, 1991). Mutation of the M1 splice sites increases the levels of M1-containing transcripts. It has been shown by ectopic expression of M1 intron-containing cDNA under the control of the hsp70 promoter that this transcript is not sufficient to produce developed sperm (Mattox & Baker, 1991). Therefore, it would appear that Tra-2 protein in the male germline acts to increase levels of M1 intron-containing transcript. The fact that this transcript has no observable function, suggests that this acts as a type of negative feedback mechanism to limit the levels of active Tra-2 in the male germline. In support of this, Tra-2 ectopic transgenic expression experiments show that, of the two Tra-2 isoforms present in the germline, only the larger, 226 amino acid, protein is capable of promoting correct spermatogenesis in XY/tra-2 flies (Mattox et al., 1996). Thus, the 179 amino acid protein produced from the M1-containing transcript has no function in spermatogenesis. As well as restoring correct spermatogenesis to XY/tra-2 flies, the 226 amino acid Tra-2 protein also directs the production of M1-containing tra-2 transcript and testis-specific exon transcript (Mattox et al., 1996). It is interesting to note that sequences with similarity to the 13-nt repeats which are
required for Tra/Tra-2-mediated splicing of the \textit{dsx} transcript have also been found in both \textit{tra-2} and \textit{exu} cDNAs (Mattox & Hazelrigg, pers. comm.).

1.3.5 SUMMARY

The available evidence suggests that both male and female germ cells require an inductive signal from the somatic component of the gonad which, along with cell autonomous processes, is required for correct gametogenesis. It is not clear what genes are involved in these processes in male germ cells, but a model is emerging for female germ cells (figure 1.14). An inductive signal, under the control of \textit{tra-2}, is required from the follicle cells for correct oogenesis. This signal results in germline-specific splicing of \textit{Sxl} transcripts to enable production of active Sxl protein. The Fused, Liz, Otu and Ovo proteins are all required in the germline for \textit{Sxl} activation, along with a X:A ratio of 1.0. The Fused serine/threonine kinase may be involved of the transduction of the somatic signal and might modulate the activity of transcription factors, such as Ovo. The successful transduction of the inductive signal is dependent upon activation of the \textit{otu} gene. As well as promoting \textit{Sxl} activation, the Otu and Ovo proteins appear to be involved in a cell-autonomous pathway which is essential for female germ cell viability.

A recent screen for modifiers of the \textit{ovo}^D dominant allele has identified at least 4 regions containing \textit{ovo}^D suppressors and at least 6 regions containing \textit{ovo}^D enhancers (Pauli \textit{et al.}, 1995). The authors estimate that these regions contain around 20 \textit{ovo}-modifier loci. Further studies of these loci may enable the isolation of factors important in germline sex determination and lead to further elucidation of the processes involved.
Figure 1.14
Diagram of the factors known to play a part in germline sex determination and differentiation in *Drosophila*. In the female, an inductive signal is required for pole cells to be determined as female. The *tra-2* gene is required in the soma for this signal to be sent. Neither *tra* nor *dsx* appear to be required. *Tra-2* may act in concert with an, as yet unidentified, cofactor other than *Tra*. The S/T kinase encoded by the *fused* gene is required cell-autonomously in the female germline and may be involved in transducing the somatic signal. The *ottu* gene is required for pole cells to respond to their somatic environment and may lie downstream of the Fu kinase. The *Sxl* gene is only activated in female germ cells and the genes *ottu, ovo, liz* and *fl(2)d* are required for this activation. The *ottu* and *ovo* genes are involved in oogenic functions other than sex determination, as is the somatic *dsx* gene. In male germ cells, *tra-2* is required cell-autonomously for correct spermatogenesis. There may also be a requirement for somatic induction under hierarchy control, but it is not known which genes control this putative signal.
1.4 FACTORS INVOLVED IN SOMATIC SEX DIFFERENTIATION

A great deal is known about the genes and processes which are involved in determining the sex of a cell in the developing embryo of *Drosophila*. Much less is known at the molecular level about how this decision is transduced into a sexually dimorphic phenotype. To date, the only non-gonadal sex differentiation genes to be cloned are the *yolk protein* genes.

1.4.1 THE *YOLK PROTEIN* GENES

There are three *yolk protein* genes (*yp1*, *yp2* and *yp3*) and all are X-linked. The *yp1* and *yp2* genes are divergently transcribed as shown in figure 1.15. The *yp3* gene is situated over 1000kb away from *yp1* and *yp2*.

![Figure 1.15](image)

**Figure 1.15**
Factors involved in female fat body-specific expression of *yp1* and *yp2*. Protein binding sites, as shown by footprint analysis, are indicated. Directions of *yp* transcription are shown.

The regions which have so far been identified as important in the regulation of *yp1* and *yp2* are indicated in figure 1.15 (reviewed in Bownes, 1994). The differential
regulation of the yp’s in the female fat body and follicle cells is dependent upon separable regions within the intergenic region and in the first exon of yp2. Correct levels of follicle cell-specific yp expression are dependent upon the elements OE1 and OE2 which act synergistically to confer this tissue-specific expression pattern (Logan et al., 1989; Logan & Wensink, 1990). A 125bp orientation-dependent region named the fat body enhancer (FBE) was initially thought to be sufficient and essential for expression of the yp’s in the fat body (Garabedian et al., 1986). This is not in fact the case, as shown by Adh reporter gene constructs driven by regions of the intergenic spacer (Abrahamsen et al., 1993). The yp2 gene and 887bp of upstream DNA is sufficient to confer female fat body expression upon the reporter gene. Similarly, sequences upstream and downstream of, but not including, the FBE also confer this tissue and sex-specificity. Therefore, it is apparent that no one sequence is solely responsible for the fat body and sex-specific expression of yp1 and yp2. It has been demonstrated that sequences within 705bp of yp3 upstream DNA are sufficient to drive tissue and sex-specific expression of this gene (Liddell & Bownes, 1991) and that this region can be subdivided into separate elements which control either fat body-specific or ovary-specific expression (Ronaldson & Bownes, 1995).

Footprinting and binding studies have shown that DsxM and DsxF bind to the FBE in an identical fashion, as would be expected from the shared binding sites within the proteins (Burtis et al., 1991). Computer sequence analysis has shown that there are four Dsx binding sites within this region and 32 possible sites within the whole intergenic region (Abrahamsen et al., 1993).

The fat body of Drosophila is thought to be a homologous tissue to mammalian liver. For this reason, proteins which control liver-specific expression of the mammalian Adh gene were used in FBE footprinting experiments (Abel et al., 1992; Falb & Maniatis, 1992). Footprints of the AEF-1, C/EBP and BBP-2 proteins overlap the Dsx footprints as shown in figure 1.15. C/EBP is a transcriptional enhancer of the Adh gene in rat hepatocytes. AEF-1 is a repressor of C/EBP and so may have this
same function in \(yp\) expression. AEF-1 activates \(Adh\) expression in the fat body of \(Drosophila\).

1.4.2 DSX AS A GLOBAL REPRESSOR OF SEX DIFFERENTIATION GENES

Null \(dsx\) mutations cause flies to develop as intersexes (Baker & Ridge, 1980). The phenotype of these flies appears to result from simultaneous expression of both male and female sex differentiation. This suggests that the \(Dsx^M\) and \(Dsx^F\) proteins act purely as repressors of inappropriate sex differentiation genes.

The fact that \(XY/dsx\) null mutants express the \(yp\)'s at a high level in the fat body suggests that \(Dsx^M\) is a \(yp\) gene repressor and that \(Dsx^F\) has no role in this system (Bownes & Nöthiger, 1981). The dominant \(dsx^D\) allele constitutively expresses the male form of \(Dsx\) protein, \(Dsx^M\). \(XX/dsx^D\) flies develop as pseudomales which exhibit no \(yp\) expression in the fat body, indicating that \(Dsx^M\) is a repressor of the \(yp\) genes.

A purely repressive role for \(Dsx\) is further indicated by the fact that \(dsx\) null intersexes have mosaic genitals containing some male and some female structures, showing that both genital primordia have developed to some extent (Baker & Ridge, 1980). \(tra\) and \(tra-2\) are required throughout development for the repression of the male genital primordium (Wieschaus & Nöthiger, 1982; Belote & Baker, 1982; Epper & Bryant, 1983). Taken together these data suggest that \(Dsx^M\) represses the female genital primordium and \(Dsx^F\) represses the male genital primordium. There is however evidence that both \(Dsx\) proteins also have positive roles in activating transcription.

Ectopic expression of a hsp70-driven construct containing a cDNA encoding the \(Dsx^M\) protein leads to three novel phenotypes, indicating a positive role for \(Dsx^M\) (Jursnich & Burtis, 1993). Firstly, an abnormal pigmentation was seen in the third instar larva which was proposed to be an indication of a positive role for \(Dsx^M\) in male-specific abdominal pigmentation. This is supported by the observation that a
reduced level of adult pigmentation is seen in flies carrying \( dsx \) null mutations. The second phenotype is seen in the foreleg of the fly. The foreleg of \( Drosophila \) is sexually dimorphic in a number of ways. Males have more gustatory receptors than females and a male-specific row of large bristles known as the sex combs. It was observed that ectopic expression of the \( Dsx^M \) protein caused a large number of bristles on all the legs of both male and females to be transformed towards a sex comb-like phenotype. This was taken as evidence for a positive role for \( Dsx^M \) in this process also.

The results from the ectopic expression of \( Dsx^M \), although compelling, can be explained within a model that only allows \( Dsx^M \) a repressive function. If we imagine a group of male-specifically expressed genes which control the development of a male-specific characteristic such as the sex comb or abdominal pigmentation, then these genes would only be active in the tissue in which they are required. What inactivates these genes in the rest of the fly? It may be that some high-activity repressive factor which is expressed under the control of the homoeotic loci in a spatially restricted manner is responsible for this. The tissue in which the male-specifically expressed genes are required, would be instructed not to produce this repressor. In the equivalent tissue in the female, \( Dsx^F \) would repress the male specifically-expressed genes so that the absence of the homoeotic-controlled transcription factor would not enable their expression. If ectopic \( Dsx^M \) is present, it could compete with the homocotic-controlled repressor (and with \( Dsx^F \)) allowing some expression of male-specifically expressed genes throughout the fly and thus giving rise to ectopic male-like structures. This type of model may provide an explanation for the third observed phenotype seen in these studies, which is increased lethality. Such lethality may be caused by the extensive ectopic differentiation of male characteristics in a largely female background.

Further evidence for a positive role for \( dsx \) comes from the observation that \( Dsx^F \) can cause a four-fold increase in transcription of a lacZ reporter gene which is driven by sequences from the \( yp \) FBE (Coschigano & Wensink, 1993). The integrity of the \( dsx \)
binding sites in the FBE was required for this to take place. This represents a four-fold increase in expression over and above the basal level of expression exhibited by the same construct when it is introduced into flies which are homozygous for a \textit{dsx} null mutation. This would indicate that in order for expression of the \textit{yolk proteins} in the female fat body to be maximal, a positive regulation is required from the Dsx\textsuperscript{F} protein as well as the absence of the repressive Dsx\textsuperscript{M} protein. However, this extra increase in expression is small compared to the several hundred times increase seen between flies with and without Dsx\textsuperscript{M}, although it may still be physiologically relevant. While explanations for most of the above evidence can be made within a model which only allows for a repressive function for Dsx proteins, this direct evidence makes it much more likely that positive functions for Dsx proteins do exist.

Finally, a positive role for Dsx proteins is indicated by the observation that while certain abdominal neuroblasts undergo extra divisions in the male, no divisions are seen in either sex when the flies carry \textit{dsx} null mutations (Taylor & Truman, 1992). Conversely, both \textit{XX} and \textit{XY} flies carrying a \textit{tra} null allele exhibit the extra divisions. This suggests that Dsx\textsuperscript{M} positively causes these neuroblasts to enter the 'extra-divisions' pathway while Dsx\textsuperscript{F} positively causes the same cells to enter the 'no-extra-divisions' pathway.

The studies on the \textit{yp}'s described above indicate that Dsx\textsuperscript{M} and Dsx\textsuperscript{F} proteins exert opposite influences upon the genes while binding to exactly the same sites. How can this occur? These disparate functions must result from the regions of these proteins which are not common. It is likely that the differential effects are brought about by repressive and activatory functions, mediated by the non-common region of the Dsx proteins, upon other transcriptional activators which bind to promoters, in the same way as AEF-1, C/EBP and BBF-2 bind to the \textit{yp} FBE. In this way, the absence of either Dsx protein allows a basal rate of transcription from sex-specifically expressed genes under the control of non sex-specific activator proteins. The binding of the
appropriate Dsx protein would then boost the expression to its maximal level, while binding of the inappropriate Dsx protein would inactivate the gene.

Finally, a role for the Intersex protein can be proposed with reference to the binding of a yeast protein; the MATα2 homeodomain protein. This protein binds to its cognate site as a dimer with the aid of the MCM1 protein as shown in figure 1.16 (Wolberger et al., 1991). The genetic evidence discussed above shows that Ix is required for Dsx function. One model to account for this would be that Ix protein is required for high efficiency binding of Dsx proteins to their cognate binding sites in sex differentiation genes. Elucidation of this model awaits cloning of the ix gene.

![Figure 1.16](image_url)

**Figure 1.16**
Diagram showing mode of action of cooperative binding of the yeast MCM1 and MATα2 DNA-binding proteins.

### 1.4.3 BRANCHES IN THE HIERARCHY ABOVE DSX

It is apparent that not all sexual functions are controlled directly by dsx. The overall size of the fly is affected by Sxl mutations but not by mutations in genes epistatic to Sxl in the hierarchy. Thus, tra or tra-2 pseudomales are much larger than their XY counterparts. tra-2 is also required for processes where dsx is not. This includes the production of a somatic feminising signal to developing pole cells (see earlier discussion of germline sex determination) and the specification of the male-specific muscle of Lawrence (MOL) in the fifth abdominal segment. Transplantation of
myoblasts between male and female have shown that the identity of the MOL is not autonomous but is more likely to depend upon the sex of the innervating axons (Lawrence & Johnston, 1986). It is believed that tra and tra-2 have a role to play in this innervation, since tra or tra-2 mutations can allow the development of the muscle in chromosomal females while dsx mutations do not have this effect (Taylor, 1992).

There is also some evidence that there may be a branch of regulatory genes, leaving the hierarchy at tra/tra-2, which represses aspects of male courtship behaviour (Belote & Baker, 1987; Gailey et al., 1991; Taylor et al., 1994). Male courtship consists of tapping the female with forelegs, orienting towards and following her, vibrating a wing to produce courtship 'song', licking the female genitalia and copulation (Bastock & Manning, 1955). Initially, use of a tra-2 temperature sensitive allele showed that female repression of certain elements of male courtship behaviour are dependent upon tra-2 (Belote & Baker, 1987). It was later observed that XX flies carrying either a dsx null allele or the dsx<sup>D</sup> allele, which constitutively produces the male Dsx<sup>M</sup> protein, did not exhibit any aspects of male courtship behaviour (Taylor et al., 1994). This indicates that neither Dsx<sup>M</sup> nor Dsx<sup>F</sup> play a part in male courtship regulation. However, the possibility of a separate dsx-independent pathway which regulates male courtship is a contentious issue at present. This is due to the observation that XY/dsx flies court much less frequently than wild type flies, elicit higher levels of courtship, and have a disrupted courtship song (McRobert & Tompkins, 1985; Villella & Hall, 1996). Taylor et al., 1994, suggested that the CNS of XY/dsx flies could still be essentially male, with lower levels of courtship and higher levels of elicitation of courtship being due to generalised developmental defects caused by inappropriate expression of genes usually repressed by the Dsx proteins. Since young wild type males exhibit lower levels of courtship and higher levels of elicitation of courtship than adults, a retardation of development might cause these characteristics to be prolonged. In support of this theory, the courtship song of XY/dsx flies, although perturbed, was not nearly as discomposed as the song of gynandromorphic flies in which much of the thoracic ganglion has the XX
genotype (Taylor et al., 1994). This suggests that the CNS of XY/dsx flies may still be genetically male. In addition, XY/dsx flies are male-like in the nature of the pheromones which they produce (Jallon et al., 1988). However, recently more in-depth studies on the courtship behaviour of XY/dsx flies by Villella & Hall, 1996, have suggested that the elicitation of courtship by these mutants is not likely to be due to maturational factors. However, these same authors report that the courtship song of XY/dsx flies was "quite similar" to that of wild-type males. In general, the analysis of courtship behaviour is complicated by variations of particular behaviour seen, dependent upon which dsx allele is used and what methods are used to score the behaviour. For example, in two separate studies on XY/dsx null mutants, both using the dsx' allele, McRobert & Tompkins, 1985, reported elicitation of courtship by the mutants, while Jallon et al., 1988, reported no such effect. All of the above evidence is open to different interpretations. On balance, however, the evidence suggests that some aspects of male courtship behaviour are regulated by genes downstream of dsx, while other aspects are dependent upon tra/tra-2 and not upon dsx.

If there is a dsx-independent tra/tra-2-mediated regulatory pathway which represses certain aspects of male courtship behaviour, it may be that this pathway has shared components with the tra/tra-2-mediated pathway which controls repression of the muscle of Lawrence. Interestingly, certain alleles of the gene fruitless (fru) result in loss of the MOL and in non-specific courting behaviour, such that XY/fru flies court males just as much as females (Gailey et al., 1991; Taylor et al., 1994). Wild-type males usually only court adult females and very young males. It may be that the fru gene lies in the tra/tra-2-mediated pathways which regulate male courtship behaviour and MOL development. In support of such a model, ectopic expression of the female form of tra in the antennal lobes or mushroom bodies of the male brain causes similar non-specific courting behaviour as with some fru alleles (O'Dell et al., 1995; Ferveur et al., 1995). However, it has recently been observed that XY/dsx mutants can also display non-specific courting behaviour (Villella & Hall, 1996). Thus, it seems likely that this particular behaviour is not under the exclusive control.
of a $d_{sx}$-independent pathway. However, it may be that non-specific courting is a 'default' state which results from general perturbation of the genetic control of courtship behaviour. If this is the case, mutation of a number of courtship-determining genes may result in this phenotype even if they are not all part of a common pathway.

1.4.4 CONTINUAL DEPENDENCE UPON THE HIERARCHY

Analysis of $yp$ gene expression in flies carrying a temperature sensitive $tra-2$ allele shows that $yp$ expression remains under continual hierarchy control in the fat body, but not in the follicle cells of the ovary (Bownes et al., 1990). Regulation of the ovary-specific expression pattern of the genes which go to make up the chorion and vitelline membrane of the oocyte is independent of the genes of the sex determination hierarchy (Waring & Mahowald, 1979; Fargnoli & Waring, 1982; Kafatos et al., 1985). This suggests that the sex determination hierarchy genes act to continually regulate genes with non-gonadal sex-specific expression patterns, but are not required to maintain or repress expression of genes with gonad-specific expression patterns. Thus, while the sex determination hierarchy genes are required to determine the identity of the gonad, the maintenance of sex-specific expression patterns within the tissue appears to be under the control of tissue-specific factors. A number of studies support this hypothesis.

Two accessory gland-specific proteins are encoded by the $msP355a$ and $msP355b$ genes (Monsma & Wolfner, 1988; Monsma et al., 1990). The $msP355a$ gene encodes a protein with similarity to a hormone precursor and the $msP355b$ gene encodes a small acidic protein. Both proteins are transferred to the female during copulation. Transcripts from these genes are detected in accessory gland tissue of $XX/tra-2s^s$ pseudomales raised at the restrictive temperature. When these flies are shifted to the permissive temperature, the transcripts remain. Thus, in the same way as with the gonadal expression of the $yp$ genes, the $msP355a$ and $msP355b$ genes appear to
require hierarchy genes in order to determine the tissue in which they are expressed but do not exhibit continual dependence.

The screen performed by Schäfer, 1986a, in which the Mst87F gene was isolated (section 1.1.3.2.3), also uncovered four other genes with accessory gland-specific expression patterns. Northern analysis of one of these accessory gland-specific transcripts, using a tra-2 mutant, shows that determination of accessory gland tissue must occur for this transcript to be produced but does not indicate whether any continual regulation by tra-2 is occurring (Schäfer, 1986b).

A second differential screen performed by DiBenedetto et al., 1987, led to the isolation of one ovary-specific transcript, five testis-specific transcripts and one accessory gland-specific transcript. The accessory gland-specific transcript was shown not to be under the continual control of tra-2.

In XX gynandromorphs, the study of XO cell clones in different regions of the nervous system has enabled the identification of regions which are important for the production of certain male characteristics (Hodgkin, 1991). Correct male courtship requires a male brain and the fidelity of the male song depends upon the thoracic ganglion being chromosomally male. The continuing control of these tissues by at least one gene of the sex determination hierarchy is shown by adult female temperature shift experiments using the tra-2<sup>2</sup> allele. XX; tra-2<sup>2</sup> adults can exhibit aspects of male behaviour when shifted to the restrictive temperature. This means that at least some aspects of the dimorphic nervous system are continuously under the control of tra-2.

The sexual phenotype the genitalia and analia is also constantly dependant upon the expression of the tra-2<sup>2</sup> allele. This is shown by temperature shift experiments which shift the sex of the analia between the male and female fates (Belote & Baker, 1982; Epper & Bryant, 1983).
One system which does not fit in with the above model is the sex-specific pattern of abdominal neuroblast divisions discussed earlier (Taylor & Truman, 1992). Once these cells have entered either the male-specific or female-specific division pathway, temperature-shifts using hierarchy gene temperature sensitive alleles have no effect.

The glucose dehydrogenase (Gld) gene of Drosophila is expressed in very specific patterns in the somatic component of both male and female gonads (Feng et al., 1991). Unlike the gonad-specific transcripts discussed above, the expression pattern of Gld can be altered by hierarchy gene mutants (including dsx) in adults.

Thus, it seems that not all genes will fit the paradigm of the yp genes. However, the body of evidence suggests that the majority of genes with somatic gonad-specific expression patterns will be hierarchy gene-independent, having come under the control of tissue-specific factors, following hierarchy gene-dependent determination of gonad cells.

1.4.5 OTHER GENES WITH SEX-SPECIFIC EXPRESSION PATTERNS

Following copulation, ovulation and oviposition is stimulated in the female. Females also indicate unreceptivity, by extending their ovipositor towards any courting males. These responses can be induced by the injection of male accessory gland secretions into virgin female abdomens (Leahy & Lowe, 1967). The 36 amino acid sex peptide is synthesised in the accessory gland and is capable of eliciting the oviposition and rejection responses in a purified form (Chen et al., 1988).

The janus region contains two overlapping transcription units janA and janB (Yanicostas et al., 1989). The janA unit produces a non sex-specific transcript and two male-specific transcripts. The janB unit produces a single male-specific transcript. All of the male-specific transcripts appear to be germline-specific.
The *Andropin* gene encodes an antibacterial protein which is localised to the male ejaculatory duct (Samakovlis *et al.*, 1991). The regulation of this gene has not yet been examined.

A differential screen performed using cDNA prepared from pre-blastoderm and gastrula embryos uncovered the *yema* locus which encodes four nurse cell-specific transcripts (Ait-Ahmed *et al.*, 1987). Again, little is known about the regulation of this locus.

The *Dromsopa* gene produces a male-specific transcript encoding a protein which contains glutamine-rich OPA repeats (Grabowski *et al.*, 1991). It is not known precisely which tissue this transcript is localised to, or whether it is regulated by the sex determination hierarchy genes.

To date, no genes have been isolated which exhibit non-gonadal sex-specific expression patterns other than the *yp* genes and the genes of the sex determination hierarchy. Genes which would be involved in differentiation of non-gonadal sex-specific features would be expected to express non-gonadal sex-specific transcripts. Therefore, the elucidation of non-gonadal sex differentiation awaits the cloning of genes which express non-gonadal sex-specific transcripts. It is likely that such genes have not been isolated so far via differential screens because their transcripts are present at lower levels than abundant gonad-specific transcripts.

The following chapters describe the cloning and characterisation of a gene expressing non-gonadal sex-specific transcripts isolated via a differential screen designed to isolate candidate non-gonadal sex differentiation genes.
CHAPTER 2

MATERIALS AND METHODS
2.1 MATERIALS

2.1.1 SOLUTIONS

Chemicals were obtained from SIGMA, BDH and Aldrich. Modification enzymes, such as restriction enzymes, Taq polymerase, nucleic acid ligases and phosphatases were obtained from Pharmacia, Boehringer Mannheim, New England Biolabs (NEB), GIBCO BRL, USB and NBL. Radioactive isotopes were obtained from Amersham and ICN Pharmaceuticals.

Solutions were prepared in sterile double-distilled water, unless otherwise stated, and sterilised by autoclaving at 15 psi for 15 minutes.

The composition of solutions not described in the main text is given below;

<table>
<thead>
<tr>
<th>SOLUTION</th>
<th>COMPOSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td>1mM EDTA (pH8.0), 10mM Tris-HCl (pH7.5).</td>
</tr>
<tr>
<td>0.5M EDTA</td>
<td>0.5M Diaminoethanetetra-acetic acid (pH8.0)</td>
</tr>
<tr>
<td>10X Agarose-gel loading buffer</td>
<td>100mM EDTA (pH8.0), 0.1% (w/v) Bromophenol Blue, 0.1% Xylene Cyanol FF, 30% (v/v) Glycerol.</td>
</tr>
<tr>
<td>2X Protein gel loading buffer</td>
<td>100mM Tris-HCl (pH 6.8), 200mM DTT, 4% (w/v) SDS, 0.2% (w/v) Bromophenol Blue, 20% (v/v) Glycerol.</td>
</tr>
<tr>
<td>10X MOPS</td>
<td>0.2M Sodium-MOPS (pH7.0), 50mM Sodium Acetate, 10mM EDTA.</td>
</tr>
<tr>
<td>10X TBE</td>
<td>0.89M Tris-Borate, 0.89M Boric acid, 10mM EDTA.</td>
</tr>
<tr>
<td>20X SSC</td>
<td>3M Sodium Chloride, 0.3M Tri-Sodium Citrate (pH7.0).</td>
</tr>
<tr>
<td>20X SSPE</td>
<td>3.6M Sodium Chloride, 20mM Sodium Hydrogen Phosphate (pH 7.4), 20mM EDTA (pH 8.0).</td>
</tr>
<tr>
<td>10X PBS</td>
<td>0.1M Sodium Phosphate (pH 7.5), 1.3M Sodium Chloride.</td>
</tr>
</tbody>
</table>
Ringer's solution.............................. 1 litre: 3.2g Sodium Chloride, 3.0g Potassium Chloride, 1.8g Magnesium Sulphate, 0.69g Calcium Chloride, 1.79g Tricine, 3.6g Glucose, 17.1g Sucrose.

100X Denhardts Solution...................... 2% (w/v) Bovine Serum Albumin, 2% (w/v) Polyvinylpyrrolidone, 2% (w/v) Ficoll.

Samlon Sperm DNA.............................. 10mg/ml, sonicated and denatured.

OLB Solution O................................. 0.125M Magnesium Chloride, 1.25M Tris-HCl (pH 8.0).

OLB Solution A................................. 0.95ml Solution O, 18ol β-Mercaptoethanol, 25ol 20mM dATP, 25ol 20mM dTTP, 25ol 20mM dGTP.

OLB Solution B................................. 2M HEPES (pH 6.0).

OLB Solution C................................. Hexadeoxyribonucleotides @ OD=90 units/ml (Sigma).

OLB................................................. Mixture of OLB Solutions A, B and C, in the ratio 2:5:3, respectively.

FSB............................................... 50% (v/v) Formamide, 25% (v/v) Formaldehyde (@ 14.8% [w/v]), 25% (v/v) 10X MOPS.

NES............................................... 50% (v/v) FSB, 7.5% (v/v) 10X Agarose gel loading buffer, 42.5% (v/v) DEPC-treated dH2O.

DNA extraction buffer......................... 50mM Tris-HCl (pH 9.0), 0.1M EDTA (pH 8.0), 0.2M Sodium Chloride, 1mg/ml Ribonuclease A.

RNA extraction buffer.......................... 7M Urea, 350mM Sodium Chloride, 100mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0), 2% (w/v) SDS.

mRNA lysis/binding buffer..................... 100mM Tris-HCl (pH 8.0), 500mM Lithium Chloride, 10mM EDTA (pH 8.0), 1% (w/v) Lithium Dodecyl Sulphate, 5mM DTT.

mRNA washing buffer........................... 10mM Tris-HCl (pH 8.0), 0.15M Lithium Chloride, 1mM EDTA (pH 8.0), 0.1% (w/v) Lithium Dodecyl Sulphate.

mRNA elution solution.......................... 2mM EDTA (pH 8.0).

Magnetic bead recon. solution................ 0.1M Sodium Hydroxide.

Magnetic bead storage buffer.................. 250mM Tris-HCl (pH 8.0), 20mM EDTA (pH 8.0), 0.1% (v/v) Tween-20, 0.02% Sodium Azide.
Miniprep solution I: 50mM Glucose, 10mM EDTA (pH 8.0), 25mM Tris-HCl (pH 8.0).

Miniprep solution II: 0.2M Sodium Hydroxide, 1% (w/v) SDS.

Miniprep solution III: 3M Potassium Acetate, 11.5 % (v/v) Acetic Acid.

Southern Denaturing solution: 1.5M Sodium Chloride, 0.5M Sodium Hydroxide.

Southern Neutralising solution: 1.5M Sodium Chloride, 0.5M Tris-HCl (pH 7.2), 1mM EDTA (pH 7.7).

APS: 25% (w/v) Ammonium Persulphate.

Acid Phenol: Phenol (pH 4.3) was obtained from Sigma, pre-equilibrated with 0.1M Citrate buffer.

Phenol: Phenol (pH 8.0) was obtained from Sigma, pre-equilibrated with 0.1M Tris-HCl (pH 8.0).

Phenol/Chloroform: Phenol (pH 8.0) was mixed with Chloroform and iso-amyl alcohol in the ratio 25:24:1, respectively.

10mM dNTPs: 10mM each of dATP, dTTP, dGTP and dCTP.

10X PCR buffer: 200mM Tris-HCl (pH 8.4), 500mM Potassium Chloride.

5X 1st strand cDNA synthesis buffer: 250mM Tris-HCl (pH 8.3), 375mM Potassium Chloride, 15mM Magnesium Chloride.

5X T4 DNA Ligase buffer: 250mM Tris-HCl (pH 7.6), 50mM Magnesium Chloride, 5mM ATP, 5mM DTT, 25% (w/v) Polyethylene glycol (8000).

5X TdT buffer: 0.5M Potassium cacodylate (pH 7.2), 10mM Cobalt Chloride, 1mM DTT.

10X Transcription buffer: 400mM Tris-HCl (pH 8.0), 60mM Magnesium Chloride, 20mM Spermidine, 100mM Sodium Chloride, 100mM DTT, 1 unit/μl RNAase inhibitor.

DNAase I: 20mg/ml Deoxyribonuclease I in 50% (v/v) glycerol.

RNAase A: 20mg/ml Ribonuclease A in 50% (v/v) glycerol. DNAase contamination was removed by boiling and centrifugation.

Proteinase K: 20 mg/ml in 50% (v/v) glycerol.

Lysozyme: 8 mg/ml in dH₂O

10X SAP buffer: 200mM Tris-HCl (pH 8.8), 100mM Magnesium Chloride.
<table>
<thead>
<tr>
<th>Buffer/Compound</th>
<th>Concentration/Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP dilution buffer</td>
<td>50mM Tris-HCl (pH 8.0)</td>
</tr>
<tr>
<td>Ethidium Bromide</td>
<td>10mg/ml in dH₂O.</td>
</tr>
<tr>
<td>Phage storage buffer (SM)</td>
<td>100mM Sodium Chloride, 8.1mM Magnesium Sulphate, 50mM Tris-HCl (pH 7.5), 0.5% (v/v) Gelatin.</td>
</tr>
<tr>
<td>Phage adsorption buffer</td>
<td>10mM Calcium Chloride, 10mM Magnesium Chloride.</td>
</tr>
<tr>
<td>X-Gal</td>
<td>20mg/ml 5-bromo-4-chloro-3-indoyl-β-galactoside in dimethyl formamide.</td>
</tr>
<tr>
<td>IPTG</td>
<td>1M Isopropyl-β-D-thiogalactoside.</td>
</tr>
<tr>
<td>Ampicillin</td>
<td>50mg/ml in dH₂O.</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>10mg/ml in dH₂O.</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>5mg/ml in Ethanol.</td>
</tr>
<tr>
<td>Carbenicillin</td>
<td>50mg/ml in dH₂O.</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>34mg/ml in dH₂O.</td>
</tr>
<tr>
<td>Leupeptins</td>
<td>10mg/ml in dH₂O.</td>
</tr>
<tr>
<td>Pepstatin A</td>
<td>1mg/ml in Ethanol</td>
</tr>
<tr>
<td>PMSF</td>
<td>17.4mg/ml in Isopropanol</td>
</tr>
<tr>
<td>Lumigen Buffer 1</td>
<td>100mM Maleic Acid, 150mM Sodium Chloride, pH 7.5.</td>
</tr>
<tr>
<td>Lumigen Buffer 2</td>
<td>1% (w/v) Blocking Reagent (Boehringer Mannheim) in Buffer 1.</td>
</tr>
<tr>
<td>Lumigen Buffer 3</td>
<td>100mM Tris-HCl (pH 9.5), 100mM Sodium Chloride, 50mM Magnesium Chloride.</td>
</tr>
<tr>
<td>Lumigen Washing Buffer</td>
<td>0.3% (v/v) Tween-20 in Buffer 1.</td>
</tr>
<tr>
<td>DIG DNA labelling mix</td>
<td>1mM dATP, 1mM dCTP, 1mM.dGTP, 0.65mM dTTP, 0.35 mM DIG-dUTP, pH 7.5.</td>
</tr>
<tr>
<td>Qiagen L1</td>
<td>20mg/ml RNAaseA, 6mg/ml DNAaseI, 0.2mg/ml BSA, 10mM EDTA (pH 8.0), 100mM Tris-HCl (pH 7.5), 300mM Sodium Chloride.</td>
</tr>
<tr>
<td>Qiagen L2</td>
<td>30% (w/v) Polyethylene Glycol (6000), 3M Sodium Chloride.</td>
</tr>
<tr>
<td>Qiagen L3</td>
<td>100mM Tris-HCl (pH 7.5), 100mM Sodium Chloride, 25mM EDTA (pH 8.0).</td>
</tr>
<tr>
<td>Qiagen L4</td>
<td>4% (w/v) SDS.</td>
</tr>
<tr>
<td>Qiagen L5</td>
<td>2.55M Potassium Acetate (pH 4.8).</td>
</tr>
</tbody>
</table>
Qiagen P1.......................... 50mM Tris-HCl (pH 8.0), 10mM EDTA (pH 8.0), 0.1mg/ml RNAaseA.
Qiagen P2.......................... 200mM Sodium Hydroxide, 1% (w/v) SDS.
Qiagen P3.......................... 3M Potassium Acetate (pH 5.5).
Qiagen QBT........................ 750mM Sodium Chloride, 50mM MOPS, 15% (v/v) Ethanol, 0.15% (v/v) Triton X-100, pH 7.0.
Qiagen QC.......................... 1M Sodium Chloride, 50mM MOPS, 15% (v/v) Ethanol, pH 7.0.
Qiagen QF.......................... 1.25M Sodium Chloride, 50mM Tris, 15% (v/v) Ethanol, pH 8.5.
Qiaex QX1........................ 3M Sodium Iodide, 4M Sodium Perchlorate, 50mM Tris-HCl (pH 7.5), 0.1% (w/v) Sodium Sulphite.
Qiaex QX2........................ 8M Sodium Perchlorate.
Qiaex QX3........................ 70% (v/v) Ethanol, 100mM Sodium Chloride, 10mM Tris, 1mM EDTA, pH 7.5.
Boehringer restriction buffer A..... 33mM Tris-HCl, 10mM Magnesium Acetate, 66mM Potassium Acetate, 0.5mM DTT, pH 7.5.
Boehringer restriction buffer B..... 10mM Tris-HCl, 5mM Magnesium Chloride, 100mM Sodium Chloride, 1mM ß-Mercaptoethanol, pH 7.5.
Boehringer restriction buffer L..... 10mM Tris-HCl, 5mM Magnesium Chloride, 1mM Dithioerythritol, pH 7.5.
Boehringer restriction buffer M..... 10mM Tris-HCl, 10mM Magnesium Chloride, 50mM Sodium Chloride 1mM Dithioerythritol, pH 7.5.
Boehringer restriction buffer H..... 50mM Tris-HCl, 10mM Magnesium Chloride, 100mM Sodium Chloride, pH 7.5.
L-Broth.......................... 1 litre: 10g Difco Bacto Tryptone, 5g Difco Bacto Yeast Extract, 5g Sodium Chloride.
L-Agar.......................... 1 litre: 16g Difco Bacto Tryptone, 10g Difco Bacto Yeast Extract, 5g Sodium Chloride.
Top Agarose..................... 0.7% Agarose in L-Broth.
SOC............................. L-Broth supplemented with 2.5mM Potassium Chloride, 10mM Magnesium Chloride, 10mM Magnesium Sulphate, 20mM Glucose.
2.1.2 BACTERIOPHAGE LIBRARIES.

2.1.2.1 Stratagene *Drosophila* Genomic DNA library.

This library was purchased from Stratagene and utilises the Stratagene Fix™II bacteriophage lambda (λ) vector. The vector contains two adjacent *XhoI* sites in its multiple cloning site, which is flanked by convergent T3 and T7 transcriptional promoters. The vector was digested with *XhoI* and the resulting sticky ends were partially filled to create Sau3A-compatible ends. Vector was then added to a Sau3A digest of *Drosophila* genomic DNA and ligated. The vector is capable of accepting insert sizes of 9-23kb.

2.1.2.2 Adult Body cDNA Library.

This library was obtained from Prof. M. Ashburner (University of Cambridge), and was constructed in the bacteriophage lambda (phage λ) vector NM1149. RNA was extracted from *Drosophila* male and female carcasses (lacking gonads) and double-stranded cDNA synthesised. Internal restriction sites were protected by methylation and a linker sequence was ligated to both ends of the cDNA, which introduced an *EcoRI* site at the 5' end and a *HindIII* site and an *EcoRI* site at the 3' end (Dorssers & Postmes, 1987). In this way, *EcoRI/HindIII* double-digestion of the cDNA produced cDNAs with a *EcoRI*-compatible site at the 5' end and a *HindIII*-compatible site at the 3' end. The cDNA's were then cloned into the *EcoRI* and *HindIII* sites of the NM1149 vector.
2.1.3 PLASMID VECTORS AND BACTERIAL HOSTS

Details of all plasmid vectors and bacterial hosts used are given below;

<table>
<thead>
<tr>
<th>PLASMID</th>
<th>RELEVANT GENOTYPE</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBluescript</td>
<td><em>amp</em></td>
<td>Phagemid used for general cloning procedures and generating single-stranded DNA templates for DNA sequencing. Contains multiple cloning site inserted, in-frame, into β-galactosidase (LacZ) gene. The LacZ gene is stimulated by IPTG to produce β-galactosidase which catalyses the breakdown of X-Gal, creating a blue reaction product. Thus, blue colonies are produced, unless insert DNA is present in the multiple cloning site.</td>
</tr>
<tr>
<td>pGEX3x</td>
<td><em>amp</em></td>
<td>Glutathione-S-Transferase (GST) Fusion Protein expression vector, containing a multiple cloning site downstream of GST sequence and an IPTG-inducible Ptac promoter.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HOST</th>
<th>RELEVANT GENOTYPE</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>XL-1 Blue</td>
<td>*Rec A1, end A1, gyr A96, thi, hsdR17, supE44, rel A1, λ-, Δ(lac), [F', proAB, lacF', ZΔM15, Tn10(ter')]</td>
<td>Used to propagate pBluescript clones. Tet selection is required to maintain F plasmid which encodes factors required for blue/white colour selection and single-stranded DNA preparation from pBluescript phagemid.</td>
</tr>
<tr>
<td>C600</td>
<td><em>F', thi-1, leuB6, lacY', tonA21, supE44</em></td>
<td>Used to propagate both 'phage λ libraries described above.</td>
</tr>
<tr>
<td>BL21</td>
<td>DE3 LysS</td>
<td>Used as host for pGEX expression vectors. Chloramphenicol selection is required to maintain LysS plasmid, encoding factors preventing expression from Ptac promoter in absence of IPTG and enable efficient lysis of cells via freeze-thawing.</td>
</tr>
</tbody>
</table>
2.1.4 DROSOPHILA STRAINS

*Drosophila* stocks were maintained at 18°C or 25°C on cornmeal food (1 litre: 25g cornflour, 50g sugar, 17.5g yeast, 10g agar, boiled, cooled to 40°C and poured into bottles or vials to set). Nipagen fungicide was added to 4.5μg/l. Antibiotics were added only on an occasional basis to prevent selection for resistant bacterial strains. In the event of mite infestations, filter paper strips, soaked in 3% (v/v) Benzyl Benzoate in ethanol and allowed to air dry, were placed on the cooled food prior to introduction of any flies. The relevant details of *Drosophila* strains used in the following analysis are shown below. All mutant *Drosophila* stocks were obtained from Prof. R. Nöthiger (Zoologisches Institut der Universität, Zürich).

<table>
<thead>
<tr>
<th>STOCK</th>
<th>GENOTYPE</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oregon R (OrR)</td>
<td>Wild Type.</td>
<td></td>
</tr>
<tr>
<td>2-10</td>
<td>B<em>Y; <em>tra</em>2</em>s<em>1</em> bw /ln(2LR)CyO</td>
<td>*XY flies have bar eyes. Permissive temperature, 16°C. Restrictive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>temperature, 25°C. <em>tra</em>2<em>s</em>1* heterozygotes have curly wings. <em>tra</em>2<em>s</em>1*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>homozygotes have straight wings and brown eyes.</td>
</tr>
<tr>
<td>2-11</td>
<td>B<em>Y; <em>tra</em>2</em>a<em>2</em> bw / CyO</td>
<td>As 2-10, but the <em>tra</em>2<em>a</em>2* allele is less severe than <em>tra</em>2<em>s</em>1*, showing increased viability.</td>
</tr>
<tr>
<td>O3-10</td>
<td>B<em>Y; <em>dsx</em> p</em> / TM6; <em>ubx</em></td>
<td>*XY flies have bar eyes. <em>dsx</em> null homozygotes have pink eyes. The TM6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>balancer is homozygous lethal.</td>
</tr>
<tr>
<td>O13-1</td>
<td>B<em>Y; <em>dsx</em> p</em> / TM6</td>
<td>As O3-10, but a different <em>dsx</em> null allele.</td>
</tr>
<tr>
<td><em>tra</em> FEM</td>
<td>B<em>Y; Df(3L)82j7 <em>Ki</em> roe p</em> hs[<em>tra</em> FEM ] / <em>dsx</em>D</td>
<td>*XY flies have bar eyes. Heterozygotes exhibit kinked, stubbled bristles. <em>dsx</em>D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>homozygotes have stubbled bristles and ebony body colour. <em>tra</em> FEM homozygotes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>have kinked bristles and roughened, pink eyes.</td>
</tr>
</tbody>
</table>

Both *dsx* and *tra*2*s* homozygotes exhibit poor viability and trans-heterozygotes (e.g. *tra*2*s*1* / *tra*2*a*2*) were constructed to take advantage of hybrid vigour.
2.2 METHODS

2.2.1 GENERAL METHODS

2.2.1.1 Phenol extraction.

Nucleic acid solutions were deproteinised by vortexing with an equal volume of Tris-HCl-equilibrated phenol (pH 8.0). The aqueous and organic phases were separated by centrifugation in an eppendorf microcentrifuge (12Krpm for 1min) and the upper (aqueous) phase removed to a fresh tube. The nucleic acid solution was then re-extracted with an equal volume of phenol/chloroform and again with an equal volume of chloroform/iso-amyl alcohol (24:1) to remove any traces of phenol.

2.2.1.2 Butanol extraction.

Ethidium Bromide was removed from nucleic acid solutions by vortexing with an equal volume of dH₂O-saturated butanol. The aqueous and organic phases were separated by centrifugation in an eppendorf microcentrifuge (12Krpm for 1min) and the lower (aqueous) phase removed to a fresh tube. This was repeated until no traces of Ethidium Bromide remained in the aqueous phase.

2.2.1.3 Precipitation of nucleic acids.

Nucleic acids were precipitated by addition of 0.1X volumes of 3M Sodium Acetate (pH 4.8) and 2.5X volumes of either ethanol or isopropanol. The solution was mixed and incubated on ice for 15min. Nucleic acids were pelleted by centrifugation at 4°C in an Eppendorf microcentrifuge at 17Krpm for 15min. The supernatant was removed and the pellet washed in 80% (v/v) ethanol. The pellet was then dried in vacuo and resuspended in TE or dH₂O.
2.2.1.4 Quantitation of nucleic acid solutions.

The OD$_{260}$ of the nucleic acid solution was determined in a suitable spectrophotometer. DNA solutions of 0.05 mg/ml, and RNA solutions of 0.04 mg/ml, have an OD$_{260}$ of 1.0. Thus, the concentration of any nucleic acid solution, of known OD$_{260}$, can be calculated. Dilutions of nucleic acid solutions were used which produced OD$_{260}$ readings within the range 0.2-1.0, since readings outwith this range are unreliable. Concentrations of nucleic acid solutions were also determined by comparison with samples of known concentration on Ethidium Bromide-stained agarose gels. A protein-free nucleic acid solution is indicated by an OD$_{260}$/OD$_{280}$ reading in the range 1.8-2.0, with readings significantly lower than this indicating protein contamination.

2.2.1.5 Agarose gel electrophoresis.

For analysis of DNA samples, agarose was added to 1X TBE and dissolved by heating in a microwave oven. The gel mixture was allowed to cool to approximately 60°C and then poured into a gel forming apparatus. 0.1X vols of 10X agarose gel loading buffer was added to DNA samples and the samples were loaded into the wells, under 1X TBE, in the gel apparatus. Gels were run in 1X TBE at the voltages recommended by the apparatus manufacturer. DNA fragments were visualised either by addition of Ethidium Bromide (to 0.5ug/ml) to the gel mix prior to pouring, or by staining the gel after running, in 1X TBE containing 0.5ug/ml Ethidium Bromide. Gels were destained in 1X TBE, to remove excess Ethidium Bromide. For most applications, 1.0% (w/v) agarose gels were used. Lower agarose concentrations were used where the resolution of DNA fragments larger than around 2.0kb was required (e.g. 0.8% for 2-10kb fragments). Higher agarose concentrations were used to resolve small DNA fragments (e.g 1.2% for 0.2-1.0kb fragments). The molecular weights of DNA molecules were determined with reference to Gibco BRL '1kb DNA ladder' molecular weight markers.
For analysis of RNA samples, denaturing agarose gels were used. Agarose was added to RNAase-free dH₂O and dissolved by heating in a microwave oven. The gel mixture was allowed to cool to approximately 60°C and 10X MOPS added to a final concentration of 1X. Formaldehyde was then added to a final concentration of 2.2M. The mixture was then poured into gel forming apparatus and allowed to set. RNA samples were denatured by adding an equal volume of FSB and heating to 65°C for 15 minutes followed by snap-cooling on dry ice. Where visualisation of the samples was required, Ethidium Bromide was added to the samples to a final concentration of 0.01μg/μl. The samples were loaded and the gel run in 1X MOPS buffer, according to the voltages recommended by the gel apparatus manufacturer. The molecular weights of RNA molecules were calculated with reference to transcripts of known size (e.g. 18s small subunit ribosomal protein transcript, ribosomal protein 49 [rp49] transcript, ribosomal RNA transcripts, α-tubulin transcript and dsx transcripts).

2.2.1.6 DEAE purification of DNA fragments from agarose gels.

DNA was separated on 1X TBE agarose gels. A slit was cut in front of the required band to enable insertion of DEAE membrane (Schleicher & Schnell). A piece of DEAE membrane was cut to the width of the slit and to a depth slightly greater than that of the slit. The membrane was activated by soaking for 5 minutes in 10mM EDTA, followed by soaking for 5 minutes in 0.5M Sodium Hydroxide. The membrane was then washed six times in excess dH₂O and stored in dH₂O at 4°C until required (for up to 3 weeks). The membrane was inserted into the slit and the gel was run until the band could be seen, under UV light, to be immobilised upon the membrane. The membrane, containing the DNA band, was then removed, rinsed in low salt buffer (150mM Sodium Chloride, 0.1mM EDTA), and placed into an Eppendorf tube. DNA was eluted from the membrane by incubating the membrane at 65°C in 300μl high salt buffer (1M Sodium Chloride, 0.1mM EDTA, 20mM Tris-HCl [pH 8.0]) for 30 minutes. The eluate was removed to a fresh tube and the
elution step repeated. Both eluates were combined, phenol/chloroform extracted and ethanol precipitated.

2.2.1.7 Qiaex purification of DNA fragments from agarose gels.

DNA was separated on 1X TBE agarose gels. The required band was excised from the gel in a minimum gel slice and placed in an Eppendorf tube. 300\(\mu\)l of QX1 solubilisation buffer and 30\(\mu\)l of 1M mannitol were added per 100mg of gel slice. 10\(\mu\)l of Qiaex DNA-affinity matrix was added and the gel slice was solubilised at 50°C for 10 minutes. Qiaex was briefly pelleted in a microcentrifuge, washed twice with 500\(\mu\)l of QX2 wash buffer and twice with 500\(\mu\)l of QX3 wash buffer. The pellet was dried \textit{in vacuo} for 2 minutes. DNA was eluted in 20\(\mu\)l of TE at 50°C. A second elution step was carried out and the eluates were pooled.

2.2.1.8 Restriction endonuclease digestion of DNA.

In general, a 5-fold excess of restriction enzyme was used for digestion of DNA molecules (i.e. 5 units of enzyme per 1\(\mu\)g of DNA). Boehringer Mannheim restriction endonuclease reaction buffers were added to a final concentration of 1X. Restriction reactions were carried out at the recommended reaction temperature for the particular enzyme, for at least 1 hour.

2.2.1.9 Ligation of DNA molecules.

Where required, 5' phosphate groups were removed from the restriction endonuclease-digested ends using shrimp alkaline phosphatase, to prevent intramolecular religation. Generally, 0.1 units of shrimp alkaline phosphatase (SAP) was used per phosphatase reaction. SAP 10X reaction buffer (NEB) was added to a final concentration of 1X. Phosphatase reactions were carried out at 37°C for 60 mins, followed by a 10 min incubation at 70°C to inactivate the enzyme.
Typically, vector and insert fragments were separated on agarose gels and purified via DEAE membrane. In general, a vector:insert ratio of 1:3 was used for standard ligation reactions, with 10-100ng of insert DNA being used. Ligations were carried out in 10ul reactions containing 1X ligase buffer and 0.01 units of T4 DNA ligase. Ligations were carried out overnight at 4°C. An aliquot of the ligation reaction was then transformed into *E. coli* competent cells.

2.2.1.10 Transformation of DNA into CaCl₂-competent *E. coli*.

Calcium Chloride competent cells were prepared as follows. 100ml of L-broth was inoculated with a single colony of an appropriate bacterial strain and incubated at 37°C, with shaking, until the bacterial cells were in Log growth phase (OD₆₀₀ of 0.5-1.0). The cells were harvested by centrifugation (2 Krpm, 10 mins @ 4°C) and resuspended in an equal volume of ice-cold 50mM Calcium Chloride. The cells were incubated on ice for 20 mins, re-pelleted, and resuspended in 0.1X volumes of ice-cold 50mM Calcium Chloride. The competent cells were then stored overnight on ice to maximise their efficiency. A volume of ligation reaction not greater than 1/10th of the final volume was added to 100-200ul of competent cells in an Eppendorf tube. The cells/ligation mix was incubated on ice for 30 mins, heat-shocked at 42°C for 90 secs, and placed on ice for 1 min. 2XYT growth medium was added to a final volume of 1ml, and the mixture incubated at 37°C for 45 mins, to enable bacterial expression of newly-transformed antibiotic resistance genes. The transformed cells were pelleted by brief centrifugation in a microcentrifuge and resuspended in 100ul L-broth. The cells were then plated on petri dishes containing L-agar, supplemented with appropriate antibiotics. The plates were covered, inverted and incubated overnight at 37°C.

2.2.1.11 Transformation of *E. coli* by electroporation.

Electroporation-competent cells were prepared as follows. 100ml L-Broth was inoculated with a single colony of an appropriate bacterial strain and incubated at
37°C, with shaking, until the cells were in Log growth phase (OD$_{600}$ of 0.5-1.0). The cells were incubated on ice for 30 mins and pelleted by centrifugation at 4000xg for 15 mins at 4°C. The cells were resuspended in an equal volume of dH$_2$O, re-pelleted, and resuspended in 0.5X volumes of dH$_2$O. The cells were then pelleted, resuspended in 2ml 10% (v/v) glycerol, re-pelleted, and resuspended in 200μl 10% (v/v) glycerol. Ligation reaction was precipitated and resuspended in dH$_2$O, to remove any ions which could result in short-circuiting of the electroporation apparatus. 1-2μl of ligation reaction was added to 40μl of competent cells, mixed, and placed on ice for 1 min, in an electroporation cuvette. A 5 msec pulse of 2.5 KVols (25μF, 200Ω) was passed through the cells/ligation mix. 1ml of SOC medium was immediately added to the cuvette and mixed. The transformed cells were incubated for 45 mins at 37°C and plated as above.

2.2.1.12 Preparation of genomic DNA.

Approximately 5mg of Drosophila tissue was homogenised in a 1.5ml Eppendorf tube containing 50μl DNA extraction buffer. 50μl DNA extraction buffer, supplemented with 0.04% (w/v) SDS, was added. Proteinase K was added to a final concentration of 0.05 mg/ml and the homogenate incubated at 65°C for 60 mins. The mixture was then phenol extracted twice and nucleic acid was ethanol precipitated. The genomic DNA pellet was resuspended in TE.

2.2.1.13 Small scale preparation of plasmid DNA

1.5ml of plasmid-transformed overnight bacterial culture was pelleted by brief centrifugation in a microcentrifuge. The cells were resuspended in 100μl of ice-cold miniprep solution I and incubated on ice for 5 mins. 200μl of freshly made miniprep solution II was added and mixed by inverting the tube 6-10 times. 150μl of ice-cold miniprep solution III was then added, mixed by inversion, and incubated on ice for 5 mins. Bacterial debris and genomic DNA were pelleted by centrifugation in a microcentrifuge at 17 Krpm for 15 mins (4°C). The supernatant was removed to a
fresh tube, phenol extracted, phenol/chloroform extracted and chloroform extracted. Plasmid DNA was ethanol precipitated, washed in 80% (v/v) ethanol, dried in vacuo and resuspended in TE. This method typically yielded 1-10µg of plasmid DNA.

2.2.1.14 Large scale preparation of plasmid DNA via Qiagen columns.

Tip-100 Qiagen columns were used to prepare up to 100µg of plasmid DNA. Essentially, this method is a scaled-up version of the Birnboim & Doly-type method described above. Qiagen DNA-affinity columns are used to enable purification of high quality, RNA-free plasmid DNA. Plasmid-transformed cells were harvested from 100ml of overnight culture by centrifugation at 6000xg for 15 min at 4°C. The cells were resuspended in 4 ml of ice-cold buffer P1. 4ml of buffer P2 was added, mixed by inversion, and the mixture incubated at room temperature for 5 mins. 4ml of ice-cold buffer P3 was then added, mixed by inversion, and the mixture incubated on ice for 15 mins. Bacterial debris and genomic DNA were removed by centrifugation at >20,000xg for 30 min at 4°C. The supernatant was centrifuged again at >20,000xg for 15 min at 4°C, to remove any traces of cellular debris. The supernatant was then applied to a qiagen Tip-100 (previously equilibrated with 4ml of QBT buffer) and allowed to fully enter the column. The column was washed with 20ml of buffer QC, and the plasmid eluted in 5ml of buffer QF. 0.7X volumes of isopropanol were added, mixed by inversion, and the precipitate pelleted by centrifugation at >15,000xg for 30 min at 4°C. The pellet was washed in 80% (v/v) ethanol, dried in vacuo, and resuspended in TE.

2.2.1.15 Large scale preparation of plasmid DNA via CsCl gradient.

The Caesium Chloride gradient method of plasmid preparation enables purification of several mg of extremely pure plasmid DNA. Plasmid-transformed cells were harvested from 250ml of overnight bacterial culture by centrifugation at 6000xg for 15 min at 4°C. The cells were resuspended in 5ml miniprep solution I and incubated on ice for 5 min. 10ml of freshly made miniprep solution II was added and mixed by

92
inverting the tube 6-10 times. 7.5ml of ice-cold miniprep solution III was then added, mixed by inversion, and incubated on ice for 5 mins. Bacterial debris and genomic DNA were pelleted by centrifugation at >20,000xg for 30 mins (4°C). The supernatant was removed to a fresh tube, containing 15ml Isopropanol, mixed, and incubated at room temperature for 15 min. Nucleic acid was pelleted by centrifugation at >15,000xg for 10 min at 4°C, washed in 80% (v/v) ethanol, dried in vacuo and resuspended in 3 ml TE. 50μl of 10mg/ml Ethidium Bromide was added to the 3 ml of plasmid solution, and the total density of the solution adjusted to 1.59 g/ml (+/- 0.01 g/ml) by addition of solid Caesium Chloride. The solution was placed in ultracentrifuge tubes and centrifuged at 45 Krpm for 20 hours at 4°C, in a swing-out rotor. The resulting plasmid band was visualised under UV light and extracted from the ultracentrifuge tube using an 18-gauge syringe needle. Ethidium Bromide was removed from the plasmid solution by butanol extraction. Plasmid DNA was ethanol precipitated, pelleted, washed in 80% (v/v) ethanol, dried in vacuo and resuspended in TE. Typically 1-2 mg of plasmid DNA were obtained using this method.

2.2.1.16 Preparation of total cellular RNA.

All solutions required for preparation of RNA were either rendered RNAase-free by treatment with Diethylpyrocarbonate (DEPC), or prepared from RNAase-free chemicals and RNAase-free dH₂O. DEPC was added to a final concentration of 10% (v/v), mixed thoroughly, and the mixture incubated at 37°C overnight. DEPC was then removed from the solution by autoclaving at 15 psi for 15 min. All glassware was rinsed with Chloroform and DEPC H₂O, and all plastics were treated with 0.1M Sodium Hydroxide and rinsed in DEPC H₂O. Where possible, all steps were carried out on ice, dry ice or in liquid Nitrogen, as appropriate.

Up to 100mg of *Drosophila* tissue was placed in a 1.5 ml Eppendorf tube and frozen in liquid Nitrogen. 400μl of RNA extraction buffer was added to the tissue and the tissue was homogenised using a hand-held homogeniser. The homogenate was
phenol/chloroform extracted with 200μl of acid phenol and 200μl of chloroform/iso-amyl alcohol (24:1). The homogenate was re-extracted with an equal volume of chloroform/iso-amyl alcohol (24:1) until no residue was observed between the aqueous and organic phases. 1ml of ethanol was added and the RNA precipitated at -70°C for at least 30min. RNA was pelleted by centrifugation in a microcentrifuge at 17Krpm for 15min at 4°C, washed in 80% (v/v) ethanol, dried in vacuo and resuspended in DEPC H₂O. RNA solutions were either used immediately, or frozen in liquid Nitrogen and stored at -70°C.

2.2.1.17 Preparation of PolyA⁺ RNA.

Oligo (dT)₂₅ superparamagnetic polystyrene beads (Dynal) were used for preparation of PolyA⁺ RNA. 250μl of bead suspension (5mg/ml) was removed from the main stock and placed in a fresh 1.5ml Eppendorf tube. The beads were magnetically captured, the storage buffer removed and the beads gently resuspended in 200μl lysis/binding buffer. Up to 50mg of *Drosophila* tissue was placed in a 1.5ml Eppendorf tube and frozen in liquid Nitrogen. 1ml of lysis/binding buffer was added to the tissue and the tissue was homogenised using a hand-held homogeniser. Debris was removed from the homogenate by centrifugation at 17Krpm for 5min at 4°C. The homogenate was then passed three times through a 21-gauge syringe needle to reduce viscosity caused by genomic DNA and fine cell debris. The 200μl aliquot of magnetic beads was placed in a magnet to capture the beads, and the supernatant removed. The 1ml homogenate was then added to the beads, gently mixed, and incubated for 8min at room temperature. The beads were washed three times with 1ml of washing buffer, with the beads being magnetically captured after each wash. PolyA⁺ RNA was eluted from the beads in 20μl elution solution at 65°C for 2min. For Northern blot applications, PolyA⁺ RNA was eluted in NES. The RNA was then denatured, supplemented with loading buffer, and loaded on a denaturing gel, as described above.
2.2.1.18 Southern blot analysis.

DNA samples were electrophoresed in an agarose gel, as described previously. The gel was soaked in 0.25M HCl until 10min after both Bromophenol Blue and Xylene Cyanol dyes had changed colour. The gel was rinsed in dH₂O and then soaked in denaturation buffer until 15min after loading buffer dyes had reverted to their original colours, or for 30min in total, whichever was longer. The gel was then rinsed in dH₂O and soaked in neutralisation buffer for 15min. A tray was filled with 20X SSC and a raised platform placed in the fluid. A wick, made from three sheets of 3M blotting paper, was placed on the platform, with the edges of the wick extending into the fluid. The gel was placed on the wick and a sheet of Hybond-N⁺ nylon membrane layed on top of the gel. Three further sheets of 3M paper were placed on top of the Hybond-N⁺. Finally, a stack of absorbent towels were layed on top and were weighted down by a 1kg weight. Transfer was allowed to continue for at least 16hrs. Before removing the Hybond-N⁺ from the gel, the position of wells were marked on the membrane using a syringe needle. The membrane was then removed, rinsed in 2X SSC, dried at 65°C and the DNA fixed by UV cross-linking. The membrane was pre-hybridised in 5X SSPE, 5X Denhardt's solution, 0.5% (w/v) SDS, 0.1mg/ml salmon sperm DNA, 50% (v/v) formamide, overnight at 42°C. Denatured, labelled, probe was added and allowed to hybridise overnight at 42°C. Background hybridisation was removed by washing in 0.1X SSPE, 0.5% (w/v) SDS, at room temperature for 15min. If background signal persisted, further room temperature washes were carried out, followed by washes at increasing temperature. Generally, washing in 0.1X SSPE, 0.5% (w/v) SDS, at 65°C for 60min was sufficient to remove persistent background signal.

2.2.1.19 Northern blot analysis.

RNA samples were electrophoresed in an agarose-formaldehyde gel, as described above. RNA was transferred to Hybond-N membrane. As for Southern blotting, a capillary blot was set up using DEPC 20X SSPE as transfer buffer. No pre-treatment
of the gel was required. Transfer, fixation, pre-hybridisation, hybridisation and stringency washes were carried out as for Southern blotting.

2.2.1.20 Radiolabelling of DNA by random priming.

30-100ng of DNA, in a final volume of 32ul, was denatured by boiling for 3min and snap-cooled on dry ice for 1min. 10ul OLB, 1ul 20mg/ml BSA, 5ul $[^{32}P]$-dCTP (50uCi) and 2ul Klenow DNA polymerase (2 units) were added to the denatured DNA and mixed. The labelling reaction was carried out for 60min at 37°C. Unincorporated $[^{32}P]$-dCTP was removed by passing the labelling reaction through a Pharmacia gel-filtration 'Nick' column, according to the manufacturers instructions. Labelled probe was denatured at 95-100°C for 5min and snap-cooled on ice. The denatured, labelled probe could then be added to pre-hybridised Southern or Northern membranes.

2.2.1.21 Determination of radiolabelling efficiency.

1ul of radiolabelling reaction was diluted into 200ul dH$_2$O containing 10ug BSA. The mixture was cooled on ice for 10min, and tri-chloroacetic acid (TCA) added to 10% (w/v), to precipitate nucleic acid. Precipitate was collected by centrifugation and both pellet (representing incorporated label) and supernatant (representing un-incorporated label) were collected. The precipitated pellet was resuspended in 200ul dH$_2$O. Equal quantities of incorporated and un-incorporated label solutions were spotted onto separate Whatman GF/C glass filters and allowed to air dry. The filters were placed in scintillation vials, covered with scintillation fluid (6g/l butyl-PBD in toluene) and radioactivity quantitated in a scintillation spectrometer. Thus, the percentage incorporation could be calculated. Typically, 60-90% incorporation was achieved.
2.2.1.22 Autoradiography.

Detection of $^{32}$P radiolabelled probe on Nylon membranes was performed using Cronex 4 Dupont X-ray film in light-tight cassettes at -70°C. Films were developed in an Agfa-1 automatic film processor. $^{35}$S-labelled probes were detected using the same procedure, but were exposed to film at room temperature.

2.2.1.23 Labelling of DNA with digoxigenin-dUTP.

30-3000ng of DNA, in a final volume of 15ul, was denatured by boiling for 3min and snap-cooled on dry ice for 1min. 2ul hexanucleotide mix (Boehringer Mannheim), 2ul digoxigenin (DIG) labelling mix and 1ul Klenow DNA polymerase (1 unit) were added to the denatured DNA and mixed. The labelling reaction was carried out for 60min at 37°C. Unincorporated DIG-dUTP was removed by passing the labelling reaction through a Pharmacia gel-filtration 'Nick' column, according to the manufacturers instructions. Labelled probe was denatured at 95-100°C for 5min and snap-cooled on ice. The denatured, labelled probe could then be added to pre-hybridised Southern or Northern membranes.

2.2.1.24 Labelling of RNA with digoxigenin-UTP.

T7 transcriptional promoters present in pBluescript phagemid and FixII lambda vectors enable production of RNA molecules extending into insert sequences. These *in vitro* transcripts can be labelled with DIG-UTP. The following components were mixed in a 1.5ml Eppendorf tube; 14ul denatured DNA template (1ug), 2ul 10X Transcription buffer, 2ul DIG labelling mix, 2ul T7 RNA polymerase (20units/ul). The reaction was incubated at 37°C for 2-15min. Unicorporated DIG-UTP was removed as before, using Pharmacia 'Nick' gel-filtration columns. The labelled riboprobe was then denatured at 95-100°C for 5min, snap-cooled on dry ice, and added to pre-hybridised Northern or Southern membranes.
2.2.1.25 Detection of digoxigenin-labelled probes.

The chemiluminescent substrate Lumigen PPD (4-Methoxy-4-[3-phosphatephenyl] spiro[1,2-dioxetane-3,2'-adamantane; Boehringer Mannheim) was used to detect probes labelled with digoxigenin. Alkaline phosphatase catalyses the conversion of Lumigen PPD into an unstable intermediate which decomposes into light-emitting products. Thus, combination of Lumigen PPD with an anti-digoxigenin/alkaline phosphatase antibody conjugate (anti-DIG-AP) enables localised detection of digoxigenin via exposure to X-ray film. Following hybridisation of DIG-labelled probes to Southern or Northern filters, stringency washes were carried out as follows. The filter was washed 2 X 5min at room temperature with 2X SSC, 0.1% (w/v) SDS, 2 X 15min at 68°C with 0.1X SSC, 0.1% (w/v) SDS. The filter was then washed briefly in lumigen buffer 1 and blocked for 30min in lumigen buffer 2. The membrane was incubated for 30min in buffer 2 containing anti-DIG-AP at a final concentration of 75mU/ml. Excess antibody was removed by 2 X 15min washes in buffer 1 and the membrane was equilibrated for 5min in lumigen buffer 3. Lumigen PPD was added in buffer 3, at a final concentration of 0.1mg/ml and allowed to bathe the membrane for 10min. Excess fluid was blotted off with 3M paper and the membrane sealed in a plastic bag. The membrane was pre-incubated at 37°C for 15min and then exposed to X-ray film at room temperature for 15-25min. Films were developed as described above.

2.2.1.26 Determination of DIG-dUTP labelling efficiency.

Alkaline phosphatase catalyses the production of a coloured precipitate from the substrates X-phosphate (5-bromo-4-chloro-3-indolyl phosphate, toluidinium salt) and NBT (nitroblue tetrazolium salt). Thus, DIG-labelled probe can be directly visualised using anti-DIG-AP, and the quantity of labelled DNA can be directly compared to a control labelling reaction. Serial dilutions of an aliquot of the DIG labelling reaction (following removal of un-incorporated DIG-dUTP) were made. Equal quantities of each dilution were spotted onto Hybond N+ membrane, allowed to air dry and fixed
by UV cross-linking. A control filter was also made up, using control DIG-labelled DNA provided by Boehringer Mannheim. The filters were washed firstly in lumigen buffer 1 for 1min, and then blocked in lumigen buffer 2 for 30min. The membranes were incubated for 30min in buffer 2 containing anti-DIG-AP at a final concentration of 75mU/ml. Excess antibody was removed by 2 X 15min washes in buffer 1 and the membrane was equilibrated for 5min in lumigen buffer 3. 45ul of NBT solution (75mg/ml in dimethyl formamide) and 35ul of X-Phosphate solution (50mg/ml in dimethyl formamide) were added to 10ml buffer 3 and the resulting "colour solution" was added to the membranes. The membranes were incubated in the dark for 15min to allow the colour precipitate to form. Estimates of labelling efficiency were then made on the basis of comparison to the supplied control labelling reaction. Typically, labelling was at least as efficient as in the provided control reactions.

2.2.2 DNA SEQUENCING

2.2.2.1 Manual dideoxynucleotide DNA sequencing.

2.2.2.1.1 Preparation of single-stranded DNA template.

A single pBluescript-transformed XL-1 Blue colony was inoculated into 50ml 2XYT medium, supplemented with 50ug/ml ampicillin and 50ug/ml tetracycline. VCS helper 'phage (kanamycin resistant) was added to 10⁸ pfu/ml. The culture was incubated at 37°C for 2hrs, with shaking, and VCS-producing cells were selected by addition of kanamycin to 70ug/ml. The culture was then grown overnight at 37°C, with shaking. Bacterial cells were separated by centrifugation at 8,000xg for 10min at 4°C. The supernatant was placed in a fresh tube, and 5ml 20% (w/v) Polyethylene Glycol (8000), 2.5M Sodium Chloride added. The 'phage were precipitated on ice for 1hr, harvested by centrifugation at 12,000xg for 5min at 4°C, and resuspended in 400ul TE. The 'phage solution was phenol extracted twice, and the aqueous phase removed to a fresh tube. Single-stranded 'phagemid DNA was Ethanol precipitated,
washed in 80% ethanol, dried *in vacuo* and resuspended in TE. Typically, 1µg of single-stranded DNA was obtained from each 1ml of starting culture. This DNA was used directly in primer annealing reactions.

2.2.2.1.2 Denaturation of double-stranded DNA.

Double-stranded DNA denaturation was carried out by addition of 0.1X volumes of 2M Sodium Hydroxide, 2mM EDTA to 3-5µg of plasmid DNA. The solution was incubated for 30min at 37°C and neutralised by addition of 0.1X volumes of 3M Sodium Acetate (pH 4.8). The denatured DNA was precipitated with 2-4 volumes of ethanol and incubated for 30min at -70°C. The DNA was pelleted by centrifugation in a microcentrifuge at 17Krpm for 15min at 4°C, washed in 80% (v/v) ethanol and dried *in vacuo*. The pellet could then be resuspended in 7µl dH₂O and used in an annealing reaction exactly as for single-stranded template.

2.2.2.1.3 Annealing of template and primer.

Annealing and subsequent sequencing reactions were carried out using Sequenase® version 2.0 DNA sequencing kit. 7µl of template DNA (1µg single-stranded, 3-5µg denatured double-stranded) was mixed with 1µl of sequencing primer (~2pmoles) and 2µl of sequenase reaction buffer (USB). The annealing mix was heated to 65°C for 2min, left to cool slowly to below 35°C and placed on ice.

2.2.2.1.4 Termination reactions.

Four tubes were labelled T, C, G and A. 2.5µl of the appropriate termination mix was added to the labelled tubes (e.g. ddATP-containing mix added to 'A' tube) which were placed at 37°C. Sequenase DNA polymerase was diluted 1:8 in enzyme dilution buffer (USB), and sequenase 5X labelling mix was diluted to 1X. The following components were added to the annealed primer/template mix on ice; 1µl DTT (0.1M), 2µl 1X sequenase labelling mix, 0.5µl ³⁵S-dATP (5µCi) and 2µl diluted
sequenase DNA polymerase. The reaction components were mixed by pipetting and the labelling reaction incubated at room temperature for 2-5min. 3.5μl of labelling reaction was added to each of the four termination tubes, mixed by pipetting and the termination reactions incubated at 37°C for 5min. 4μl of stop solution/loading buffer was added to each termination reaction, mixed by pipetting and the reactions stored on ice until required for loading.

2.2.2.1.5 Polyacrylamide gel analysis of DNA sequencing reactions.

Sequencing gel mix was 6% (w/v) acrylamide (19:1 acrylamide:bis-acrylamide) in 7.7M urea (in 1X TBE). For a 50ml gel, 100μl 25% APS and 100μl TEMED was added to 50ml gel mix and the gel poured between taped, siliconised, glass plates. The 'shark's-tooth' type of comb was used. Gels were run in 1X TBE buffer. Prior to loading, gels were pre-warmed by running at 40 watts (for 40cm x 30cm x 0.4mm gels) for 30min. Termination reactions were denatured at 75°C for 2min and snap-cooled on ice. 3μl of each termination reaction was loaded, in the order T-C-G-A. Gels were run at 40 watts for 2-9hrs (40cm x 30cm x 0.4mm gels), or at 70 watts for up to 16hrs (80cm x 30cm x 0.4mm gels). Following electrophoresis, gels were fixed for 20min in 10% (v/v) Acetic Acid, 10% (v/v) methanol, transferred to blotting paper and dried under vacuum. Autoradiography was carried out to X-ray film, at room temperature, for 5-24hrs.

2.2.2.2 Automated dideoxynucleotide DNA sequencing.

Automated sequencing reactions were carried out using PRISM™ Ready Reaction DyeDeoxy™ dye-labelled terminator cycle sequencing kits (Perkin Elmer) and analysed using an ABI automatic sequencer. Template DNA (250-500ng double-stranded plasmid, 200ng PCR product) in a volume of 11μl, was mixed with 1μl sequencing primer (3.2pmoles) and 8μl Ready Reaction mix (Perkin Elmer, containing labelling mix, dye-labelled ddNTPs & thermo-stable DNA polymerase). The reaction was overlaid with mineral oil and placed in a PCR thermal cycler,
pre-heated to 96°C. Cycling was initiated as follows; 30sec @ 96°C, 15sec @ 50°C, 4min @ 60°C. 25 cycles were performed. The reaction was placed on ice and 2μl 3M Sodium Acetate (pH 4.6), 50μl 50% (v/v) ethanol added. Nucleic acid was precipitated on ice for 10min and pelleted in a microcentrifuge at 17Krpm for 30min at 4°C. The pellet was washed in 70% (v/v) ethanol, dried in vacuo and stored at 4°C. The pellet was later resuspended in loading buffer, denatured and analysed on an automatic sequencer.

2.2.3 MANIPULATIONS WITH BACTERIOPHAGE LAMBDA

2.2.3.1 Plating bacteriophage lambda.

Plating cells were prepared as follows. A single colony of host bacterial cells was inoculated into 50ml L-broth, supplemented with 0.4% (w/v) maltose, and grown overnight at 37°C with shaking. Cells were harvested at 2000xg for 10min, resuspended in 20ml 0.01M Magnesium Sulphate and stored at 4°C (up to 3 weeks). 1μl of phage stock (typically 1 plaque [~1 X 10^{10} pfu] stored in 1ml SM/50μl chloroform) was added to 100μl plating cells, 100μl adsorption buffer. This adsorption mix was incubated for 10min at room temperature and 20min at 37°C, to enable adsorption and injection. The adsorption mix was then added to 3ml 0.7% top agarose, at 47°C, swirled and poured evenly over L-agar. When set, the plate was covered, inverted and incubated overnight at 37°C.

2.2.3.2 Screening bacteriophage lambda libraries.

Libraries were plated on large square petri dishes, such that one plate contained around 1 X 10^5 plaques. Hybond N+ nylon membrane was cut to the size of the plate, lowered over the surface of the plate and its position marked with a syringe needle. After 1min, the membrane was peeled off the top agarose, and lowered, plaque side up, into a tray of denaturing solution. After 7min, the membrane was transferred to a tray of neutralising and soaked for 3min. The membrane was rinsed carefully in a
tray of 2X SSC, air dried and fixed by UV cross-linking. Replica membranes were made for each plate. The membranes were hybridised with a gene-specific probe, as for Southern blots, and exposed to X-ray film. Plaques which hybridised with the gene-specific probe were removed from the original plate, using a cut P-200 yellow-tip and stored overnight, at 4°C, in 1ml SM, to allow phage to diffuse out of the agar plug. Phage from the plaque were then re-plated at low density to allow selection of well separated plaques. Again, Hybond N membrane lifts were taken and hybridised with a gene-specific probe. Hybridising plaques were selected and stored in 1ml SM, 50ml chloroform, at 4°C.

2.2.3.3 Preparation of bacteriophage lambda DNA.

Liquid lysates were used to prepare phage lambda DNA. The phage from which DNA was to be prepared were plated out at low density and a single well-separated plaque removed into 300ul adsorption buffer. The plaque was left to diffuse overnight at 4°C. A single colony of host cells was inoculated into L-broth, supplemented with 0.4% maltose, and the culture grown to an OD_{600} of 0.6-1.0. 200ul of these Log-phase cells were added to the 300ul of adsorption buffer containing the diffused plaque. The cells/phage mixture was incubated for 10min at room temperature and then for 30min at 37°C. The adsorbed phage were then added to 50ml L-broth (supplemented with 0.1% glucose, 10mM Magnesium Choride) and incubated at 37°C overnight, with shaking. If lysis was incomplete, chloroform was added to 1ul/ml. Cell debris was removed by centrifugation at 8000xg for 10min at 4°C. The supernatant was removed to a fresh tube and 100ul of Qiagen L1 solution added. The lysate was incubated at 37°C for 30min. 10ml of ice-cold Qiagen L2 was added, and the phage precipitated for at least 60min on ice. The precipitated phage were harvested by centrifugation at >10,000xg for 10min and resuspended in 3ml of Qiagen L3. 3ml of Qiagen L4 was added and the phage coats denatured at 70°C for 20min. The phage solution was cooled on ice, 3ml of Qiagen L5 added and precipitated protein removed by centrifugation at >15,000xg for 30min, at 4°C. The supernatant was added onto a Qiagen Tip-100 column (previously equilibrated with
3ml QBT buffer) and allowed to enter completely. The column was washed with 20ml QC buffer and the phage DNA eluted with 5ml QF buffer. DNA was precipitated by addition of 3.5ml isopropanol and pelleted in a centrifuge at 15,000xg for 30min, at 4°C. The pellet was washed in 80% (v/v) ethanol, dried in vacuo and resuspended in TE. Typically, this method yielded 60-100ug of phage DNA. Phage were either stored as plaques in SM/chloroform, or 1ml samples of cell debris-free liquid lysate was supplemented with 50ul chloroform and stored at 4°C.

2.2.4 IN SITU HYBRIDISATION TO WHOLE-MOUNT TISSUES

2.2.4.1 Fixation of tissues.

Tissues were dissected in paraformaldehyde/PIPES (PP) solution (50ml; 40ml DEPC dH₂O, 2g paraformaldehyde, 30ul 10M Sodium Hydroxide, 5ml 0.5M PIPES, 100ul 0.5M EGTA [ethylene-glycol-bis{2-aminoethylether}-N,N-tetra acetic acid], 100ul 1M Magnesium Sulphate, pH 6.8). The tissues were transferred to 1ml of Fix solution (1:1 heptane:PP) and allowed to fix for 1-1.5hr, with shaking. The Fix was removed and the tissues treated as follows. The tissues were washed three times for 1min in 9:1 methanol:0.5M EGTA; once for 5min in 7:3 methanol:EGTA; once for 5min in 5:5 methanol:EGTA; once for 5min in 3:7 methanol:EGTA; once for 5min in PP and once for 20min in PP. At this stage, tissues could be dehydrated by serial dehydration and stored in ethanol at 20°C. They must, however, be re-hydrated before proceeding with further steps.

2.2.4.2 Dehydration and re-hydration of tissues.

Dehydration was carried out as follows. Tissues were washed twice for 5min in 1X PBS, once for 5min in 30% (v/v) ethanol; once for 5min in 50% (v/v) ethanol and once for 5min in 70% (v/v) ethanol. The tissues were then stored at -20°C in ethanol. Re-hydration was carried out as follows. Tissues were washed once for 5min in 70%
(v/v) ethanol; once for 5min in 50% (v/v) ethanol; once for 5min in 30% (v/v) ethanol; once for 5min in 1X PBT and once for 30min in 1X PBT.

2.2.4.3 Tissue pre-treatment and hybridisation.

The following steps were all carried out in 1.5ml Eppendorf tubes, at room temperature, using a revolving wheel, unless otherwise stated. The tissues were rinsed three times, for 5min each time, in PBT (1X PBS + 0.1% [v/v] Tween-20) and then treated for 3.5min with 50ug/ml proteinase K (in PBT). Proteinase digestion was stopped by washing twice, for 1min each time, in 2mg/ml glycine. The tissues were washed twice, for 5min each time, in PBT and once, for 20min, in PP. This was followed by six, 5min, PBT rinses. Finally, the tissues were rinsed in 200ul 1:1 PBT:Hybrix (50% [v/v] formamide, 5X SSC, 50ug/ml heparin, 0.1% [v/v] Tween-20, 100ug/ml sonicated, denatured, salmon sperm DNA) for 10min and pre-hybridised in 100ul Hybrix for 60min at 45°C, without shaking. DIG-Labelled riboprobe was ethanol precipitated, pelleted, resuspended in Hybrix and added to the pre-hybridised tissues. Hybridisation was carried out overnight, at 45-48°C, without shaking.

2.2.4.4 Post-hybridisation washes and anti-DIG-AP antibody adsorption.

Following hybridisation, the tissues were washed, for 20min, in 500ul Hybrix at 45°C, then twice, for 20min, in 500ul 1:1 PBT:Hybrix at 45°C. This was followed by a 20min PBT rinse at 45°C and five, 5min, PBT rinses at room temperature. Anti-DIG-AP antibody (@ 75mU/ml) was pre-adsorbed to a second sample of the relevant tissues for 1.5hr (to reduce non-specific background signal) and then added to the probe-hybridised tissues and incubated for 1.5hr at room temperature. The tissues were washed once, for 5min, and then three times, for 20min, in PBT, to remove excess antibody.
2.2.4.5 Detection of anti-DIG-AP antibody.

Tissues were washed three times, for 5min, in NMTT (100mM Sodium Chloride, 50mM Magnesium Chloride, 100mM Tris-HCl [pH 9.5], 1mM levamisole [a potent lysosomal phosphatase inhibitor]). The tissues were then incubated in the dark, for 1-10hr, in 1ml NMTT, containing 3.5vol 50mg/ml X-phosphate, 4.5vol 75mg/ml NBT. Tissues were checked periodically, for colour development and the colour reaction stopped by 3 PBT washes.

2.2.5 POLYMERASE CHAIN REACTION

2.2.5.1 Reaction conditions and optimisation.

Polymerase chain reactions (PCRs) were carried out using Promega Taq (*Thermus aquaticus*) thermo-stable DNA polymerase and 10X PCR reaction buffer (500mM Potassium Chloride, 100mM Tris-HCl [pH 8.3], 0.01% [w/v] gelatin). Reactions were performed in a final volume of 50ul and were overlaid with 50ul of mineral oil. As a starting point, reactions were set up as shown below;

- 3ul 25mM Magnesium Chloride
- 1ul 10mM dNTP's (dATP, dTTP, dCTP, dGTP)
- 5ul 10X PCR Reaction buffer
- 0.5ul Each oligonucleotide primer (50pmol/ul)
- 2ul Template
- 2ul Taq Polymerase (1.25 U/ul)
- 36ul dH2O

50ul Total

Cycling conditions are essentially divided up into 3 phases; denaturation (@ 94-96°C), annealing (@ 4 degrees below the primer annealing temperature [Tm])
and extension (@ 72°C). The approximate Tm of any given primer was calculated using the equation; Tm = 2 x (Number of A & T nucleotides) + 4 x (number of G & C nucleotides). For Promega Taq polymerase, the extension phase was estimated to proceed at the rate of 1-2kb per min. Thus, for amplification of a 1kb product, using primers of Tm = 60°C, the following cycling conditions might be used:

Stage 1 (1 cycle): Hot Start. Reactions (excluding Taq) placed in thermal cycler and heated to 95°C for 2min. Temperature held at 80°C, while Taq is added.

Stage 2 (25 cycles): denaturation @ 94°C for 15sec; annealing @ 56°C for 15sec; extension @ 72°C for 1min.

Stage 3 (1 cycle): Final extension @ 72°C for 10min.

The hot-start method was used to prevent non-specific priming during the initial climb to 94°C, in stage 2.

These conditions represent a starting point and each set of primers and template was optimised to minimise non-specific products and maximise yield. In general, all phases were kept as short as possible to maximise enzyme life, and denaturation was done at 94°C, rather than 95°C-96°C, for the same reason. The production of non-specific products was minimised in the following ways; reduction of annealing/extension times, reduction of cycle number, increasing annealing temperature, reduction of Magnesium Chloride concentration, reduction of dNTP concentration and reduction of Taq concentration. PCR products were analysed on agarose gels.

2.2.5.2 PCR of reverse transcription products.

10ul of Oligo dT (1mg/ml) was added to 10ul of total cellular RNA (1mg/ml) and incubated at 70°C for 10min, then snap-cooled on dry ice. The following components
were added; 40 l 5X 1st strand cDNA synthesis buffer, 20 l 0.1M DTT, 10 l 10mM dNTPs, 10 l DEPC dH₂O, 0.30 l RNAguard™ (Pharmacia), 10 l Superscript II™ reverse transcriptase (200U/μl). The reverse transcription reaction was incubated at 37°C for 60min, made up to 1000 l with DEPC dH₂O and phenol/chloroform extracted once. 50 l of reverse transcriptase reaction was used in each PCR reaction, which were carried out as described above.

2.2.5.3 Rapid amplification of 3' cDNA ends (3' RACE).

Single-stranded cDNA was synthesised as described above. PCRs were carried out using an oligodT primer and a gene-specific primer (polymerising towards 3' end of transcript). In this way it is possible to amplify the 3' end of a cDNA, rapidly, from total cellular RNA.

2.2.5.4 Rapid amplification of 5' cDNA ends (5' RACE).

Single-stranded cDNA was synthesised, as described above. However, instead of phenol/chloroform extraction, the reaction was stopped via inactivation of the reverse transcriptase by heating to 55°C for 5min. 10 l of RNAase H (2U/μl) was added and the reaction was incubated at 55°C for 10min, then placed on ice. cDNA was purified away from primers and dNTPs using Qiaex DNA affinity matrix, as described above. 130 l of the total 400 l of cDNA, was removed to a fresh tube, denatured at 70°C for 10min and snap-cooled on ice. A PolyA tail was added to the 5' ends of the single-stranded cDNA molecules, as follows. The following components were added to the cDNA on ice; 40 l 5X Terminal transferase (TdT) reaction buffer, 20 l 2mM ATP, 10 l TdT (10U/μl). The terminal transferase reaction was incubated at 37°C for 10min and stopped at 70°C for 10min. 50 l of the stopped terminal transferase reaction was used in PCRs with oligodT and a gene specific primer (polymerising towards 5' end of transcript) to amplify the 5' end of transcripts represented in the single-strand cDNA pool.
2.2.6 PROTEIN ANALYSIS METHODS

2.2.6.1 Analysis of proteins by polyacrylamide gel electrophoresis.

Polyacrylamide gels for protein analysis were 10% acrylamide and were made up as follows; 13.3ml 30% (w/v) acrylamide, 10ml 1.5M Tris-HCl (pH 8.8), 0.2ml 20% (w/v) SDS, 16.1ml dH₂O. 400μl 10% (w/v) APS and 16μl TEMED were added and the gel was poured between glass plates. 1ml of H₂O-saturated butanol was used to overlay the gel, until polymerisation was complete. A 5% stacking gel was made up as follows; 1.7ml 30% (w/v) acrylamide, 1.25ml 1M Tris-HCl (pH 6.8), 50μl 20% (w/v) SDS, 6.85ml dH₂O. 100μl 10% (w/v) APS and 10μl TEMED were added and the gel was poured so as to overlay the 10% resolving gel. A comb was inserted and the gel was left to polymerise. Protein samples were prepared for loading by addition of an equal volume of 2X protein loading buffer and boiling for 3min. Gels were run in Tris/glycine buffer (25mM Tris, 250mM glycine, 0.1% [w/v] SDS, pH 8.3), at voltages recommended by the gel apparatus manufacturer. Following electrophoresis, gels were stained using the ISS Pro-Blue system (Integrated Separation Systems). Gels were fixed in 12% TCA for at least 60min. The TCA was then removed and the gel soaked in pretreatment solution (20ml Reagent A, 80ml dH₂O) for 60min. The gel was then soaked in staining solution (16ml Reagent A, 1.6ml Reagent B, 20ml methanol, 64ml dH₂O) until protein bands were apparent (>1hr). Excess stain was removed in 25% (v/v) methanol. The gel was photographed and stored in 20% (v/v) ethanol at room temperature.

2.2.6.2 Purification of GST-fusion proteins.

The open reading frame of interest was cloned into an expression vector of the pGEX series of glutathione-S-transferase (GST) fusion protein expression vectors. This construct encoded a protein containing a region of GST which binds to its substrate, glutathione. Thus, the protein can be purified using glutathione immobilised upon an agarose matrix. A single pGEX-transformed colony was
inoculated into 100ml L-broth, supplemented with 100µg/ml carbenicillin and grown at 37°C, with shaking, until an OD$_{600}$ of 0.6-1.0 was reached. Fusion protein expression was induced by addition of IPTG to 0.1mM and the culture grown for a further 4 hours. Cells were harvested by centrifugation at 5000xg for 10min, at 4°C, and resuspended in 20ml ice-cold 1X PBS. The cells were lysed by brief sonication on ice and proteins solubilised by addition of Triton X-100 to 1% (v/v). Insoluble protein and cell debris was removed by centrifugation at 10,000xg for 5min, at 4°C. 1ml of 50% (v/v) glutathione-agarose slurry (Sigma) was added and gently mixed for 30min. The slurry was collected by a 10sec centrifugation at 500xg, washed three times in 50ml ice-cold 1X PBS and resuspended in 1ml ice-cold 1X PBS. GST-fusion protein was eluted in 1ml 50mM Tris-HCl (pH 8.0), 5mM glutathione, by mixing gently for 2min. Glutathione-agarose slurry was pelleted by a 10sec centrifugation at 500xg and the eluate removed to a fresh tube. Three further elutions were performed and the eluates stored at -70°C.

2.2.6.3 Western blot analysis of protein SDS polyacrylamide gels

Following electrophoresis of protein samples (2.2.6.1) the gel was electroblotted to Hybond C nitrocellulose membrane. A sheet of Hybond C was cut to the size of the gel and placed over it. The gel and membrane was sandwiched between 4 sheets of blotting paper and the whole assembly placed between two sheets of sponge. This was then inserted into a Biorad electroblotting apparatus, with the Hybond membrane facing the positive electrode. The tank was filled with Tris/glycine/methanol buffer (1 litre: 9g Tris, 432g glycine, 3g SDS, 600ml methanol) and blotted overnight at 15-20mA (~6v). The membrane was removed from the gel and blocked in TBS (10mM Tris-HCl [pH 8.0], 100mM Sodium Chloride, 0.5% [v/v] Tween-20) + 5% (w/v) fat-free milk powder, 0.02% (w/v) Sodium Azide, for 1-2hr at room temperature. Primary antibody (e.g. Sigma anti-GST rabbit IgG) was added to the membrane, at 1-10µg/ml in TBS + 5% (w/v) fat-free milk powder, 0.02% (w/v) Sodium Azide and incubated for 2hr at 4°C. The membrane was washed five times, for 5min each time, in TBS and secondary
antibody (e.g. Promega anti-rabbit IgG Horseradish peroxidase [HRP] conjugate) was added (@ 1μg/ml in TBS + 5% [w/v] milk powder) and incubated for 1-2hrs at room temperature. Excess antibody was removed by five, 5 min, washes in TBS. Anti-rabbit IgG HRP conjugate was detected using the insoluble peroxidase substrates Diaminobenzidine tetrahydrochloride (DAB) and Urea Peroxide. One pre-prepared tablet of each chemical (Sigma) was suspended in 1-20ml of TBS and poured over the membrane. The membrane was incubated at room temperature, with shaking, until colour precipitate was formed (typically within 15min).

2.2.6.4 Kinase assays of crude bacterial extracts.

A pGEX expression construct encoding the protein to assayed for kinase function was transformed into XL-1 blue cells. A single colony of these cells was inoculated into 50ml 2XYT medium, supplemented with 100μg/ml carbenicillin. The culture was grown at 37°C, with shaking, until an OD_{600} of 0.6-1.0 was reached. Protein expression was induced by addition of IPTG to 1mM, and the culture was grown for a further 3-4hr at 30°C, with shaking. A 1ml sample was taken, and the cells harvested by brief centrifugation in a microcentrifuge. The cells were washed in 1ml 1X PBS and resuspended in 100μl 1X PBS. Lysis was performed by freeze-thawing four times (liquid Nitrogen <-> 42°C water bath) and brief sonication on ice. A 25μl aliquot was removed to a fresh tube, containing a 1.5μl cocktail of proteinase inhibitors (1:1:1 ratio of 100μg/ml leupeptins, 100μg/ml pepstatin A, 100mM PMSF). 0.3μl of $^{32}$P-γ-ATP (3μCi) was added to the crude protein extract and the kinase assay reaction incubated for 60min at 25°C. Reactions were analysed using SDS PAGE and autoradiography. Control reactions were carried out in parallel with extracts from cells expressing pGEX vector alone and with un-induced cells. In addition, samples were taken, to which $^{32}$P-γ-ATP was not added, but which were otherwise treated exactly as the assay samples. These samples were analysed by SDS PAGE and the gels were stained, to examine expression efficiency and integrity of expressed proteins used in the assay.
CHAPTER 3

CLONING OF A GENE ENCODING SEX-SPECIFIC TRANSCRIPTS IN DROSOPHILA MELANOGASTER
3.1 INTRODUCTION

The phenotypic effects of the sex determination hierarchy genes; \textit{Sxl, tra, tra-2} and \textit{dsx} indicates that the majority of genes controlling the sexual identity of the fly are under control of this hierarchy. It is clear that the \textit{yp}'s in the fat body remain under continuous control of \textit{dsx} throughout development, while, in the ovary, the control of these genes appears to be handed over to ovary-specific factors early in development. The mode of action of \textit{Sxl, tra, tra-2} and \textit{dsx} is to cause their target genes to produce sex-specific transcripts. Since it is clear that the majority of genes responsible for sex-specific development are under control of the sex determination hierarchy, it is reasonable to suppose that a significant proportion of these, as yet uncharacterised, sex differentiation genes are likely to give rise to male-specific or female-specific transcripts.

As discussed earlier, previous differential screens have only isolated genes which express gonad-specific transcripts. The high abundance of \textit{yolk protein} gene transcripts means that these genes are often re-isolated in differential screens. In an effort to deal with these problems, a differential screen was carried out in this lab using only flies which had first had their gonads dissected out. Thus, much of the \textit{yolk protein} and \textit{chorion} gene transcripts were eliminated from the screen. To remove those \textit{yolk protein} transcripts which are also produced in the female fat body, polyA\(^+\) RNA prepared from this tissue was firstly hybridised to cloned \textit{yolk protein} DNA and passed through hydroxyapatite, which selectively binds double stranded nucleic acid molecules. These precautions successfully eliminated \textit{yolk protein} transcripts from the differential screen, since no \textit{yp}-derived recombinants were isolated. Radiolabelled cDNA was made from mRNA prepared from male and female carcasses. This was used to differentially screen a Charon 4 bacteriophage lambda ('phage \(\lambda\)) genomic library, as shown in figure 3.1. From this screen, eight putative female-specific and seven putative male-specific recombinant 'phage were isolated. In this chapter, I describe the initial analysis of one of the putative female-specific 'phage, \(\lambda\text{fs}(1)\). This recombinant was isolated on the basis of
preferential hybridisation to female cDNA, although some hybridisation to male cDNA was also seen.

Figure 3.1
Diagram showing strategy of differential genomic library screen. Only non-gonadal tissue was used for the screen. PolyA⁺ RNA was extracted from male and female OrR carcass tissue. Female RNA was hybridised with yp cDNA and then passed over a hydroxyapatite affinity column to remove double stranded molecules. Radiolabelled cDNA was synthesised from the yp-free female RNA and male RNA. This cDNA was used as a probe to screen replica filters lifted from a plated bacteriophage λ genomic DNA library. Recombinants which hybridised with cDNA from one sex only were isolated for further analysis.
Initial characterisation of \( \lambda \text{fs}(1) \) (together with four other putative female-specific phage) was the subject of a Ph.D. project carried out by O. Claire Gilman (thesis submitted 1991) and is summarised briefly below.

After checking that \( \lambda \text{fs}(1) \) did not hybridise to cloned \( yp \) DNA and was therefore not a reisolate of these genes, the phage was restriction mapped to reveal the pattern of EcoRI sites shown in figure 3.2.

![EcoRI restriction map of bacteriophage \( \lambda \) recombinant fs(1) [taken from Gilman, 1991]](image)

To determine the cytogenetic location of any genes contained in \( \lambda \text{fs}(1) \), in situ hybridisation to third instar larval salivary gland polytene chromosomes was carried out (Robert Saunders, University of Dundee). The 4.3kb EcoRI restriction fragment of \( \lambda \text{fs}(1) \) was subcloned into the EcoRI site of pGemini-1 (pGEM-1), and this plasmid was used to synthesise a probe for the in situ hybridisation. The cytological position detected by this probe was determined to be 061C1-3, lying at the distal end of the left arm of chromosome 3. The Flybase database was searched for existing mutations in this region which result in sex-specific phenotypes, but none were isolated.

Reverse Northern were then carried out with radiolabelled cDNA from male and female flies as probes. Of the four EcoRI restriction fragments, the 3.5kb fragment hybridised to both male and female cDNA and the 4.3kb fragment hybridised only to female cDNA. This indicates that there is at least one gene present in \( \lambda \text{fs}(1) \), giving rise to both female-specific and non sex-specific or male-specific transcripts.
However, the cDNA used as a probe for these reverse Northern blots was derived from whole flies and so does not rule out the possibility that the sex-specific transcripts are of gonadal origin. Northern analysis using the same 4.3kb EcoRI fragment in pGEM-1 suggested the presence of a female-specific transcript, as would be expected from the reverse Northern data, but was not particularly clear (data not shown).

3.1.2 PRELIMINARY WORK - M. TODMAN

Further analysis of the 4.3kb EcoRI λfs(1) fragment was carried out by Martin Todman (Research Assistant, 1991-1992). The fragment was recloned into the EcoRI site of the sequencing vector pBluescript-SK\(^+\) (This subclone henceforth referred to as λfs(1)\(^{\lambda}\)) and was used as a probe to screen a λgt11 phage ÷ adult body cDNA library. Three recombinant phage which hybridised with λfs(1)\(^{\lambda}\) were isolated (cDNAa, cDNAb and cDNAc). These cDNAs contained no internal EcoRI sites and so could be separated from their lambda arms by EcoRI digestion. Thus, the entire cDNAs were cloned into the EcoRI site of pBluescript-SK\(^+\). Sequence analysis of the three subcloned cDNAs revealed that two of them were either truncated (cDNAb) or contained artefacts (cDNAc). One 3.0kb cDNA (cDNAa) was apparently full length, having both a polyA tail and a long open reading frame. The pBluescript subclone containing cDNAa (cDNAa\(^{\lambda}\)) was used to probe a Northern blot of male and female total RNA. Hybridisation was seen to a large female carcass-specific transcript of around 4.5kb (figure 3.3). An ovary-specific transcript of around 3.5kb and a very faint male-specific 3.0kb band were also seen but no non sex-specific transcripts were revealed by this analysis. Thus, the gene which produces the transcript represented by cDNAa fulfils one of the criteria for a candidate non-gonadal sex differentiation gene. Namely, the production of non-gonadal sex-specific transcripts. The second criterion; that these transcripts be under the control of the sex determination hierarchy genes, remained to be satisfied.
Figure 3.3
Autoradiograph of a Northern blot prepared from hand-dissected OrR tissues. Approximately 20μg of total RNA was loaded per lane. The filter was hybridised with a probe synthesised from cDNAaaSK. A 4.5kb female carcass-specific transcript and a 3.5kb ovary-specific transcript are apparent. There also appears to be a very faint 3.0kb male-specific transcript. This filter was not available for a loading check to be carried out.

The relation of cDNAaa to λfs(1)4.3 as revealed by restriction mapping is shown in figure 3.4.

Figure 3.4
Partial restriction maps of cDNAaaSK and fs(1)4.3. The orientation of these pBluescript SK+ inserts is shown with reference to the T3 and T7 promoters present at either end of the vector multiple cloning site. Partial sequence data indicates that around 1.5kb at the extreme 5' end of fs(1)4.3 consists of intron sequence (hatched area), although the boundaries of this intron were not precisely delimited (indicated by semi-hatched area).
Although not complete at the time, partial sequence data showed that around 1.5kb at the 5' end of \( \lambda \)fs(1)\(^{4.3} \) was not homologous to cDNAa, indicating the presence of a large intron somewhere in the middle of the cDNA. To try to isolate the remainder of the gene, cDNAa\(^{SK+} \) was used as a probe to screen a 'phage \( \lambda \) "FixII" genomic library (Stratagene-see Materials & Methods). Five recombinant 'phage were isolated (designated Fix1, Fix2, Fix4, Fix5 & Fix7) which hybridised with cDNAa\(^{SK+} \). These were not analysed further at the time, but will be discussed later in this chapter.
3.2 RESULTS

3.2.1 TRANSCRIPTS PRODUCED FROM STK6I

To confirm the Northern results obtained by Martin Todman, more Northern blots were carried out. To produce a probe of as high a specificity as possible, it was preferred that the cDNAa insert be purified away from its pBluescript-SK vector prior to labelling. However, when cDNAa<sup(SK)</sup> is digested with EcoRI, both the cDNAa insert and the pBluescript-SK<sup+</sup> vector are almost exactly 3.0kb, making separation on an agarose gel impossible. To circumvent this problem, a pBluescript-SK<sup-</sup> subclone of one of the truncated cDNA isolates, cDNA<sub>b</sub> was used. This truncation was isolated in the cDNA screen performed by Martin Todman, and is an EcoRI fragment consisting of the 3' 1557bp of cDNAa. This 1.5kb EcoRI fragment was gel purified, radiolabelled, and used to probe a Northern blot of total RNA from whole males, female carcasses (whole flies with gonads removed) and ovaries (figure 3.5).

![Figure 3.5](image)

**Figure 3.5**

Autoradiograph of a Northern blot prepared from hand dissected OrR tissues. Approximately 20ug of total RNA was loaded per lane. The filter was hybridised with a probe synthesised from cDNA<sub>b</sub>. A 4.5kb female carcass-specific transcript, a 3.5kb ovary-specific transcript and a 3.0kb male-specific transcript are apparent.
The Northern shown in figure 3.5, confirms the presence of a 4.5kb female carcass-specific transcript and a 3.5kb ovary-specific transcript. A male-specific transcript is also clearly visible and appears to be of around 3.0kb in size. The earlier reverse-Northern data suggested the presence of a common transcript as well as a female-specific transcript. Thus the observation of a male-specific transcript on Northern blots is not unexpected. The fact that the male transcript appears to be of around the same size as cDNAa, suggests that cDNAa may represent this transcript. To investigate this, and to further characterise the regulation of the gene, an attempt was made to isolate more cDNAs representing transcripts from this gene.

3.2.2 ISOLATION OF FURTHER cDNAs

A NM1149 'phage λ adult body cDNA library (courtesy Prof. M. Ashburner-see Materials and Methods) was screened using the same cDNAb probe that was used for the Northern in figure 3.5. This probe comprises the 3' 1.5kb of cDNAa and hybridised to 9 'phage λ recombinants. These 'phage were isolated and DNA prepared from them. The NM1149 library used for this screen was constructed by addition of linkers containing restriction sites to the double stranded cDNA. The sequence of these linkers was such that a EcoRI site is added to the 5' end of the cDNA and a HindIII site to the polyadenylated 3' end. The cDNA's were then directionally cloned into the EcoRI and HindIII sites of the NM1149 polylinker. Thus, EcoRI/HindIII double digestion of the cDNA library isolates liberates the cDNA from the vector arms and reveals any internal EcoRI and HindIII sites. DNA from the 9 isolates was digested with EcoRI and HindIII, and the resulting pattern of restriction fragments showed that only 2 of these isolates were both unique and large enough to represent the transcripts observed on Northern blots (data not shown). Hereafter, these cDNA's are referred to as cDNA1 and cDNA11. Figure 3.6(a) shows an EcoRI/HindIII digest of cDNA's 1 & 11. cDNA 11 consists of 1 EcoRI/HindIII fragment of 1.9kb and 2 HindIII fragments of 1.0kb and 1.4kb. cDNA1 consists of 1 EcoRI/HindIII fragment of 0.9kb and 2 HindIII fragments of 1.0kb and 1.4kb. The
HindIII fragments and EcoRI/HindIII fragments were differentiated using EcoRI and HindIII single digests (data not shown).

To confirm that cDNAs 1 and 11 represented transcripts from the same gene as cDNAa, EcoRI/HindIII digests of the cDNAs were run out on an agarose gel, Southern blotted to nylon membrane and probed with a random-primed probe made using cDNA\textsubscript{b}. The results of this are shown in figure 3.6(b). The 1.0kb and 1.9kb fragments of cDNA\textsubscript{11} hybridise to this probe, as do the 0.9kb and 1.0kb fragments of cDNA\textsubscript{1}. Thus, cDNAs 1 and 11 do represent transcripts from the same gene as cDNA\textsubscript{a}. The fact that the 1.4kb HindIII fragments of cDNA\textsubscript{1} and cDNA\textsubscript{11} did not hybridise with cDNA\textsubscript{b} suggests that cDNA\textsubscript{11} and cDNA\textsubscript{1} may be derived from a differentially processed variant of the transcript represented by cDNA\textsubscript{a}.

**Figure 3.6**
(a) 1.0% Agarose gel showing EcoRI/HindIII restriction digests of cDNAs 1 and 11. cDNA\textsubscript{11} contains one EcoRI/HindIII fragment of 1.9kb and two HindIII fragments of 1.4kb and 1.0kb. cDNA\textsubscript{1} contains one EcoRI/HindIII fragment of 0.9kb and two HindIII fragments of 1.4kb and 1.0kb. (b) Autoradiograph of a Southern blot filter made from the gel shown in figure 3.6(a). The Filter was hybridised with a probe synthesised from cDNA\textsubscript{b}. The 1.9kb and 1.0kb fragments of cDNA\textsubscript{11} hybridise with this probe, as do the 0.9kb and 1.0kb fragments of cDNA\textsubscript{1}.

121
cDNAs 1 & 11 were subcloned into pBluescript-SK+ using a shotgun cloning method. The phage containing the cDNA's were digested either with HindIII alone, or double digested with HindIII/EcoRI. pBluescript-SK+ was digested with the corresponding enzymes and then treated with Shrimp Alkaline Phosphatase to prevent religation of the vector to itself. Cut/phosphatased vector was then added to the phage digests and ligated. An aliquot of these ligations was then used to transform XL-1 Blue cells. Transformed cells were plated on X-Gal/IPTG-containing plates to allow isolation of cells containing recombinant pBluescript via blue/white colour selection. DNA was prepared from white colonies and analysed by restriction digestion with HindIII and HindIII/EcoRI. In this way, constructs containing all 3 fragments of both cDNA's were generated.

3.2.3 SEQUENCE ANALYSIS OF cDNAs

The insert DNA in these constructs was completely sequenced using chain termination dideoxy sequencing (Materials and Methods). cDNAa SK+ was also re-sequence (partially sequenced by M. Todman.) to clear up ambiguities and false frame shifts in the long open reading frame of the cDNA. The relationship between cDNA a,1 and 11, as revealed by this data, is shown in figure 3.7.

![Restriction Map Diagram](image)

**Figure 3.7**
Diagram showing partial restriction maps of cDNAs a, 1 & 11. All EcoRI, HindIII and EcoRI/HindIII fragments were subcloned into pBluescript SK+, as described in the text. The extreme 5' 0.6kb of cDNA1 consists of the 3' end of a known Drosophila ribosomal protein transcript, including its polyA tail. H3=HindIII, RI=EcoRI, Xh=XhoI.
The 1.0kb and 1.4kb fragments of cDNAs I and 11 were shown to be identical in both cDNAs. A 0.2kb HindIII fragment was discovered as a coligation in several subclones from both cDNAs I and 11 which was identical in sequence to the 0.2kb HindIII fragment of cDNAa. This fragment had not been previously seen clearly on agarose gels due to its small size. It was also seen that cDNA 1 is in fact a truncated version of cDNA 11, containing 0.6kb of the 3' end of a transcript from a known Drosophila ribosomal protein gene at its 5' end. Therefore, cDNA 1 was not analysed any further.

The complete sequence of cDNA 11 is shown in figure 3.8. The sequence of cDNA a is entirely contained within the sequence of cDNA 11 and the 5' and 3' ends of cDNA a are indicated on the cDNA 11 sequence. Also indicated is the long open reading frame and the position of all primers used for sequencing and PCR analysis. cDNAs a and 11 differ in length at both their 5' and 3' ends. The addition of a 7-methylguanosine cap to the 5' end of mRNA means that full length cDNAs have a guanine nucleotide at their 5' ends which is not encoded in the transcription unit. By this criterion, neither cDNA a or cDNA 11 appear to be entirely full length at their 5' ends. It may be that the full length transcripts represented by cDNA a and cDNA 11 share a common 5' end. However, the possibility of alternate promoters cannot be ruled out at this stage. The 3' ends of both cDNAs are complete, having polyA tails and the correct upstream polyadenylation consensus sequence, as shown in figure 3.8. Thus, the transcripts represented by cDNA a and cDNA 11 utilise alternate polyadenylation sites.
Figure 3.8

DNA sequence of cDNA11. The sequence of cDNAa and cDNA11 is identical apart from the 5' and 3' ends, which are shorter in cDNAa than in cDNA11. The 5' and 3' ends of cDNAa are indicated on the cDNA11 sequence. The positions of all oligonucleotide primers used for sequencing and PCR analysis are shown. Primers which have the same 5' -> 3' direction as the cDNA11 sequence have been given letter designations (A-G). Those primers which face in the 3' -> 5' direction, relative to cDNA11, are designated by numbers (1-7) and have the complementary sequence to that shown for cDNA11. Also indicated are the positions of intron/exon boundaries and any relevant consensus sequences.

<table>
<thead>
<tr>
<th>Primer 1</th>
<th>Primer A</th>
<th>Primer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  AGTGCTGTAG</td>
<td>TGGTGCATT</td>
<td>AAATAATACT</td>
</tr>
<tr>
<td>51  TACATPAAAAT</td>
<td>CTAATATCGGA</td>
<td>ATAAATAAAG</td>
</tr>
<tr>
<td>101  GTGTCTCAAAG</td>
<td>ATATTACGGC</td>
<td>GTGTTAAGAA</td>
</tr>
<tr>
<td>151  CAAAGATGAG</td>
<td>GTGTAGGCCC</td>
<td>GTCGTGGAGG</td>
</tr>
<tr>
<td>201  ATAACAGACA</td>
<td>ATACAGCGGT</td>
<td>TACTGGTCAG</td>
</tr>
<tr>
<td>251  GGAATTATTTG</td>
<td>TCTTTGATTAT</td>
<td>GTGTTAAGG</td>
</tr>
<tr>
<td>301  TTCAAGAAAC</td>
<td>ATACTGCGGC</td>
<td>TTGTGCCAAT</td>
</tr>
<tr>
<td>351  ACAATACAC</td>
<td>ATATACACAG</td>
<td>GCACCGCGAC</td>
</tr>
<tr>
<td>401  GCAACTACAC</td>
<td>ATATACTACAG</td>
<td>GCACCGCGAC</td>
</tr>
<tr>
<td>451  TGGCTCTCTGA</td>
<td>CAGTGTGTGC</td>
<td>TGCTTCGGTC</td>
</tr>
<tr>
<td>501  TCTGCTGTGC</td>
<td>CATTAAAGAT</td>
<td>ATGATTTT</td>
</tr>
<tr>
<td>551  AAACAAAGAG</td>
<td>CCGCGCTCTG</td>
<td>GCCGACAGC</td>
</tr>
<tr>
<td>601  CAACAGCCTC</td>
<td>AATTGGACAG</td>
<td>GTTCAAGCA</td>
</tr>
<tr>
<td>651  GTAAATCACGG</td>
<td>TAGCATCGGA</td>
<td>AACTCGACAG</td>
</tr>
<tr>
<td>701  CGAGCATCAC</td>
<td>CAGTGCTCAT</td>
<td>CACGACACAG</td>
</tr>
<tr>
<td>751  CAGAGGCAAC</td>
<td>ATACAGCGGT</td>
<td>ATGGCAGACT</td>
</tr>
<tr>
<td>801  TCCCTGGGCG</td>
<td>AGTCGATTTT</td>
<td>CAGGATATC</td>
</tr>
<tr>
<td>851  GGTGTCGAA</td>
<td>GGCGCATTG</td>
<td>GCAGAAGA</td>
</tr>
<tr>
<td>901  GCGGAGCGGT</td>
<td>ATGTGTTACT</td>
<td>GAGGACAG</td>
</tr>
<tr>
<td>951  AGCAATGCAC</td>
<td>TGACATACCG</td>
<td>CCTTGGGCG</td>
</tr>
<tr>
<td>1001  TAGCAACACGT</td>
<td>TGCTGGCTTG</td>
<td>GCATGGCCA</td>
</tr>
</tbody>
</table>

(continued...)
1051 ACTGCAGTCT GCAGCAGTAC CAGAATGACA TAAGGCAGCA GACGGAGATA
1101 TTGGACACGT TGCCGACAGA GCATCACAGC GGGTACCAGT CGCAGCAACA
1151 GCAACAACAG CCGCAGACGC AGCAAGAACA AGAAGACAGC CAGGAGCAAT
1201 CGCAGCAGCA GCAACAGCTA CAGAATCCTG CGCCCAGGAG GTCTCCGAAT
1251 GATTTCAATT TGCGTCTGTTA CATAGGCCAG GGCAGCTATA GCAATGTTTA
1301 TCTGGCCCGTG GATATACACT CTGCGCCGCA GTAAGCAATT AAAGTAGTGC
1351 AGAAGCGGCT GATGCTTGCGC GAAGCAAGGAAG AGGACTACAT CAAGGCTGA
1401 CGCCGAGGTTG TGCACCAGAT GACCAACGTG CCGCTTTTCG TAAAACGTGC
1451 GTGCACCTTC CAGGACCAGC GTTCTTCTTTA CTTGTGTGAG ACATGACCAC
1501 GAAAAGGCGGA CATGTTGCCA TACATCAACC GCCGCTACCG TTTTGACGTC
1551 GCCCTGCAGGC GCCACTACGC TGGCTCAGGC TAAACCTGTC
1601 GCACCAGCCGT GATGTGCGCG ACCGGGACCT CAAGCGTGAA GCAGCAGGTT
1651 TCGACGAGGA CATGCACACG CTAATCGCCG ACTTGGCTTC CGCCAAGGTG
1701 ATGACAGCCC AGCAAGGCGC TGCTGGCCACG GAGCACTGTT CGGAGCAAGC
1751 GCAGCAGCACG TCCGATGAAG AGCATGAAGA CAGATGACCG CAGTACGGC
1801 AAGACGGAAGA CTGCTATGGTG GCCGATTCGG AGGAGTTGGA CGACCACTAC
1851 GACCGAACGCG AGCAAGGAGA GATGGACTCC CCACGCCATC GTCAAAGGCG
1901 TTACACCAGT CACCGAAAGG CAAAGCTTGTG GGGCCTTGGC CAGTAGGCTG
1951 CACCAGGAAGT GCTCCAATAA GGACCTATAA CCGGGCGCGC GGACCTTGTC
2001 GCACTGGGAT GATATGGTTA TCAGATGATC GCCGCGTACC GCGATCTCGG

(continued...)
(continued...)
3051 TGTAATTTA AITTAGTGG ACCAAGCGG ATGGCGTAGA
3101 AGAGGGCGG GAAATAGGTA AGCTTAGTTG AGAATGTAGA
3151 GCCAAATAAT CACGTTAAGG CAACTAGTA TAAACTTATC
3201 CGATPGCTCG AGCACTTAGT AAATTACCTG CAGCAACTAT
3251 TACATCTATA ACGACTAGGT AAATTCAATT TAAACTTATC
3301 CACGTTAAGG CAACTAGTA TAAACTTATC TCACTTAGTC
3351 GCCAAATAAT CACGTTAAGG CAACTAGTA TAAACTTATC
3401 TAAAGTTGAA TAAATTCTTG AACAGCAACT TAAACTTATC
3451 TTAATTCTG GAAACCTAA TAAACTTATC TCACTTAGTC
3501 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3551 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3601 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3651 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3701 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3751 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3801 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3851 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3901 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3951 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4001 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4051 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4101 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4151 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4201 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4251 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4301 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4351 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4401 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4451 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
4501 ATTTGCTGGA ATGAACTTTT TTTTCTTTTA CCACTGACT
3.2.4 RELATIONSHIP BETWEEN cDNAs AND TRANSCRIPTS

The sizes of the cDNAs a and 11 suggest they may represent the male-specific and female carcass-specific transcripts seen on the Northern in figure 3.5. To investigate this possibility, probes were generated to the extended 5' and 3' ends specific to cDNA11. This was carried out via PCR analysis, using primers A and 1 at the 5' end, and primers F and 7 at the 3' end (figure 3.8 for primer positions). The PCR reactions were run on an agarose gel and the single bands produced were excised from the gel and purified. This DNA was radiolabelled by random priming and used to probe Northern blots. The results of this analysis are shown in figure 3.9.

Figure 3.9
Autoradiographs from Northern blots prepared using hand-dissected OrR tissues. Approximately 5μg of polyA+ RNA was loaded per lane. The same filter was used for both (a) and (b). The loading of RNA on this filter was checked by hybridisation with a probe synthesised from an 18s small subunit ribosomal protein cDNA. The female carcass lane appears to have a slightly higher loading relative to the other two lanes. (a) Autoradiograph of the Northern filter hybridised with a probe synthesised from cDNAa. The male-specific, female carcass-specific, ovary-specific and common transcripts all hybridise with this probe. (b) Autoradiograph of the Northern filter hybridised with a probe synthesised from a PCR product containing only cDNA11-specific 3' UTR sequence. Only the female carcass-specific and common transcripts hybridise with this probe.
Interpretation of these results is complicated by the presence of a common transcript on these blots. This band runs either just above or directly on top of the female carcass-specific transcript. The size difference between the common transcript and the female carcass-specific transcript can be clearly seen on the Northern blot shown in figure 3.10. This common band was not seen on earlier Northern blots (figure 3.5) but is apparent on later blots, especially those using polyA' RNA rather than total RNA, probably due to increased sensitivity. It is possible that this common band represents a partially processed version of the female carcass-specific transcript, and will be discussed in detail in chapter 4. On some gels, the 4.5kb female carcass-specific and common bands run at the same position, possibly due to differences in length of polyA tails, or to small amounts of degradation of the RNA. This would make any size difference between them so small as to be indistinguishable on agarose gel systems. We have been unable to develop a gel system capable of consistently differentiating between these two transcripts.

Figure 3.10
Autoradiograph of a Northern blot prepared from OrR tissues. Approximately 20µg of total RNA was loaded per lane. The filter was hybridised with a probe synthesised from cDNA. The Male-specific, Female carcass-specific and common transcripts all hybridise with this probe. The size difference between the female carcass-specific and common transcripts can be clearly seen. These two transcripts appear to differ in size by 200-300bp. As a loading check, the filter was hybridised with a probe synthesised from an 18s small subunit ribosomal protein cDNA.
Although the female carcass-specific transcript cannot be clearly differentiated in figure 3.9, an increase in intensity is seen. Since the loading check bands have been quantitated and are of virtually identical intensity, it is reasonable to assume that the increase in intensity in the female carcass track is due to the presence of the female carcass-specific transcript.

The 5' Untranslated Region (5' UTR) probe produced no signal on Northern blots (data not shown). Since this probe was much smaller that the 3' UTR probe, this may be explained by small amounts of degradation at the 5' ends of the mRNAs which could reduce the target area for the probe to levels which mean that the signal is too weak to be visualised by this method.

It is quite clear that the PCR probe homologous to the cDNA11-specific 3' UTR hybridises to the female carcass-specific and common transcripts, but not to the male-specific or ovary-specific transcripts. Sequence evidence presented in chapter 4, together with the observed size of the transcripts, makes it unlikely that cDNAa represents the ovary-specific transcript. The fact that the cDNA11-specific 3' UTR probe only hybridises with the female carcass-specific and common transcripts, together with their observed size, shows that this cDNA must represent one of these two transcripts. Again, sequence evidence presented in chapter 4 suggests that cDNA11 represents the female carcass-specific transcript. Thus, we can be fairly certain that cDNAa and cDNA11 represent the male-specific and female carcass-specific transcripts, respectively.

### 3.2.5 ISOLATION AND MAPPING OF FURTHER GENOMIC DNA

It is clear from figures 3.4 and 3.7 that the existing genomic DNA included in \( \lambda fs(1)^{+3} \) does not contain either the extreme 5' or 3' ends of the isolated cDNAs. In order to determine the structure of the gene it was necessary to isolate genomic DNA representing the entire transcription unit. To this end, restriction mapping of the previously isolated genomic DNA 'phage recombinants, Fix1-Fix7, was carried out.
The recombinant phage Fix1-Fix7 were digested with a variety of restriction enzymes which release insert DNA from the phage λ arms, and the enzyme XbaI was found to produce the most conveniently sized fragments. XbaI digests of all 5 'phage are shown in figure 3.11.

![Image of agarose gel showing XbaI restriction digests of recombinant genomic DNA bacteriophage Fix1-Fix7. Fragment sizes are indicated in kilobases. See text for details.]

All 5 recombinant 'phage contain large genomic DNA inserts, ranging in size from 9kb to 15.8kb. The 'phage λ "FixII" library vector contains a T7 RNA polymerase promoter adjacent to its multiple cloning site which enables the synthesis of RNA across the vector/insert boundary and into the insert itself. This system was applied to generate digoxygenin-labelled RNA probes using the Fix1 recombinant as a template for the transcription reaction. By allowing the reaction to proceed for different times, probes of different lengths were produced and used to probe a Southern blot of a Fix1 XbaI digest. Following hybridisation, the probe was visualised using the Lumigen-PPD system (Materials & Methods). The results of this
experiment are shown in figure 3.12. The shortest RNA probe hybridises only with the 2.8kb fragment of Fix1, showing that this fragment must be the one closest to the T7 promoter, at one extreme end of the insert.

![Image of Southern blots](image)

**Figure 3.12**

Autofluorographs of Southern blots prepared from Xhol digests of genomic bacteriophage λ recombinant Fix1. A single 1.0% agarose gel was loaded with three identical Fix1 Xhol digests and blotted to a single filter. This filter was then cut into three sections and each section hybridised with a separate digoxigenin-labelled RNA probe. RNA probes were synthesised using Fix1 as a template for the transcription. The transcription reactions were stopped at either 2mins, 5mins and 15 mins to produce RNA probes extending different lengths into the Fix1 insert. Hybridised probe was visualised using the Lumigen-PPD fluorescent system.

Only the 2.8kb fragment of Fix1 hybridises with the 2 min-riboprobe, showing that this fragment must lie at one end of the Fix1 insert. Unfortunately, the autofluorograph from the filter hybridised with the 5 min-riboprobe has a strong background signal obscuring the 1.7kb fragment. However, the result from the 15min-riboprobe shows that the 1.7kb fragment is the least intense of the five bands which suggests that the 1.7kb fragment lies at the opposite end of the Fix1 insert from the 2.8kb fragment.

The gel shown in figure 3.11 was Southern blotted and probed with various segments of DNA to further elucidate the orientation of the fragments in Fix1-7. The following DNA was used to make probes; the Fix2 4.8kb Xbal fragment, a cDNAa EcoRV
fragment consisting of the extreme 5' 480bp of the cDNA, the truncated cDNAb
comprising the extreme 3' 1.5kb of cDNAa, λfs(1)^4,3 and the Fix5 2.9kb XbaI
fragment. The results of this are shown in figure 3.13, and table 3.1 shows a
summary of which probes hybridised to which XbaI fragments of Fix1-7. This data
enables the construction of the restriction maps of the recombinants Fix1-7 as shown
in figure 3.14.

Figure 3.13
Autoradiographs of a Southern blot filter from the agarose gel of Fix1-7 XbaI restriction digests,
shown in figure 3.11. The sizes of fragments which hybridised with the various probes are indicated
in kilobases. These results are summarised in table 3.1. (a) Probe synthesised from the 4.8kb
fragment of Fix2. The 4.8kb fragments of all five Fix recombinants hybridise with this probe. (b)
Probe synthesised from the 480bp EcoRV restriction fragment of cDNAa, representing the extreme
5' end of the cDNA. The 3.4kb and 3.1kb fragments of Fix1, 4 & 5 hybridise with this probe, as do
the 3.4kb and 2.8kb fragments of Fix7 and the 3.0kb fragment of Fix2. (c) Probe synthesised from
the truncated cDNAb, representing the extreme 3' 1558bp of cDNAa. This probe hybridises with
the 2.8kb fragment of Fix1, the 2.4kb fragment of Fix4 and the 1.5kb fragment of Fix5. (d) Probe
synthesised from λfs(1)^4,3. This probe hybridises with the 2.8kb fragment of Fix1, the 2.4kb
fragment of Fix4 and the 1.5kb fragment of Fix5. (e) Probe synthesised from the 2.9kb fragment of
Fix5. Despite the high background signal seen here, significant hybridisation is seen to the 1.7kb
fragment of Fix1, the 1.9kb fragment of Fix4 and the 2.9kb fragment of Fix5.
Table 3.1
Summary of results shown in figure 3.13. Fragments of genomic DNA bacteriophage λ recombinants, Fix1-7, which hybridise with various probes are shown.

<table>
<thead>
<tr>
<th>PHAGE:</th>
<th>Fix1</th>
<th>Fix2</th>
<th>Fix4</th>
<th>Fix5</th>
<th>Fix7</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROBE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fix2 4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
</tr>
<tr>
<td>cDNAa 5’ 480bp</td>
<td>3.1kb</td>
<td>3.0kb</td>
<td>3.1kb</td>
<td>3.1kb</td>
<td>2.8kb</td>
</tr>
<tr>
<td>eDNAa 3’ 1.5kb</td>
<td>2.8kb</td>
<td>NONE</td>
<td>2.4kb</td>
<td>1.5kb</td>
<td>NONE</td>
</tr>
<tr>
<td>λfs(1)43</td>
<td>2.8kb</td>
<td>NONE</td>
<td>2.4kb</td>
<td>1.5kb</td>
<td>NONE</td>
</tr>
<tr>
<td>Fix5 2.9kb</td>
<td>1.7kb</td>
<td>NONE</td>
<td>1.9kb</td>
<td>2.9kb</td>
<td>NONE</td>
</tr>
</tbody>
</table>

Figure 3.14
Diagram showing restriction maps of bacteriophage λ FixII genomic DNA recombinants, Fix1-Fix7. XbaI sites and restriction fragment sizes are indicated.
The following reasoning was used to construct these maps. It is clear that the 4.8kb fragment of all 5 recombinants is in fact the same fragment. We can also say that the 3.1kb fragment of Fix1, 4 and 5 is the same fragment, as is the 3.4kb fragment of Fix1, 4, 5 and 7. Considering Fix1, 4 and 5 initially; the data shows that the 3.1kb, 3.4kb and 4.8kb fragments are identical in all 3 recombinants. It follows that the remaining two fragments of each recombinant must be at the extreme ends of the inserts, as they are of different sizes in each recombinant. It has already been shown that the 2.8kb fragment of Fix1 lies at one end of the insert. The fact that the cDNAa 3' probe hybridises with the 2.8kb fragment of Fix1, the 2.4kb fragment of Fix4 and the 1.5kb fragment of Fix5 confirms that all these fragments lie at one end of their inserts. It follows that the 1.7kb, 1.9kb and 2.9kb fragments of Fix1, Fix4 and Fix5 respectively should all lie at the other end of their inserts. This is confirmed by the fact that all 3 fragments hybridise with the 2.9kb fragment of Fix5. With respect to Fix2, we can see that the 3.0kb fragment must lie at the 5' end of the 4.8kb fragment, for if it lay at the 3' end it would have hybridised to the cDNAa 3' probe. The fragment which does lie at the 3' end of the 4.8kb fragment of Fix2 (1.2kb) is apparently too small to reach the sequence covered by the cDNAa 3' probe. This leaves the orientation of the two remaining fragments (3.1kb and 3.4kb of Fix1, 4 and 5; 2.8kb and 3.4kb of Fix7) to be determined. All of these fragments hybridise to the cDNAa 5' probe. This means that the 3.4kb fragment of Fix7 must be the same as the 3.4kb fragments of Fix1, 4 and 5. The 2.8kb fragment of Fix7 must be a truncation of the 3.1kb fragment seen in the other 3 recombinants and as such must lie at the end of the insert. This means that the 3.1kb and 3.4kb fragments of Fix 1, 4, and 5 must lie in the order 3.1kb/3.4kb with the 3.1kb fragment being the most 5'.

The map indicates that the Fix1-7 genomic DNA is highly likely to contain the 5' end of the transcription unit, as shown by the hybridisation pattern of the cDNAa 5' 480bp probe. However, since the probe representing the 3' 1.5kb of cDNAa hybridises with the 1.5kb fragment of Fix5 but not with the 1.2kb fragment of Fix2, the 5' end of the DNA used to make the probe must lie between the 3' ends of these two Fix XbaI fragments. Since we know that cDNA11 extends a further 2.7kb 3' of
this point, the Fix1-7 recombinants do not contain enough 3' genomic DNA to cover the 3' end of the transcription unit.

To recover genomic DNA including the 3' end of the transcription unit, the 'phage λ FixII genomic library was screened using a probe made from the PCR product representing the cDNA11-specific 3' UTR. Four recombinant 'phage were isolated which hybridised with this probe. These were designated Genλ.1-Genλ.4. DNA was prepared from these 'phage, digested with XbaI and run out on an agarose gel, as shown in figure 3.15.

---

**Figure 3.15**

Two representations of the same 1.0% agarose gel showing XbaI restrictions digests of the genomic DNA bacteriophage recombinants Genλ.1-Genλ.4. The 1.4kb and 0.54kb fragments are not clearly seen here but were apparent on the original gel and on subsequent gels (data not shown). Genλ.1 contains four XbaI insert fragments of sizes 6.3kb, 4.8kb, 1.6kb and 0.54kb. Genλ.2 and Genλ.3 contain five XbaI insert fragments of sizes 6.3kb, 4.8kb, 2.7kb, 1.6kb and 1.4kb. Genλ.4 contains five XbaI insert fragments of sizes 4.8kb (two fragments), 3.4kb, 3.1kb and 3.0kb.
All four phage contain large inserts, varying in size from 13.24kb to 19.1kb. Using the same techniques as described for phage Fix1-7, the phage Genλ1-4 were restriction mapped. Figure 3.16 shows the restriction maps of all 9 recombinant phage inserts, and a summary of which probe DNA hybridised to which XbaI restriction fragments of all 9 recombinant phage is shown in table 3.2.
Table 3.2
Summary of Southern blot results, showing fragments of genomic DNA bacteriophage λ recombinants, which hybridise with various probes.

<table>
<thead>
<tr>
<th>PHAGE:</th>
<th>Fix1</th>
<th>Fix2</th>
<th>Fix4</th>
<th>Fix5</th>
<th>Fix7</th>
<th>Genλ 1</th>
<th>Genλ 2</th>
<th>Genλ 3</th>
<th>Genλ 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fix2</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td>4.8kb</td>
<td></td>
</tr>
<tr>
<td>cDNAa 5' 480bp</td>
<td>3.1kb</td>
<td>3.0kb</td>
<td>3.1kb</td>
<td>3.1kb</td>
<td>2.8kb</td>
<td>1.6kb</td>
<td>2.7kb</td>
<td>2.7kb</td>
<td>3.1kb</td>
</tr>
<tr>
<td>cDNAa 3' 1.5kb</td>
<td>2.8kb</td>
<td>None</td>
<td>2.4kb</td>
<td>1.5kb</td>
<td>None</td>
<td>6.3kb</td>
<td>6.3kb</td>
<td>6.3kb</td>
<td>4.8kb</td>
</tr>
<tr>
<td>£x(1)£3</td>
<td>2.8kb</td>
<td>None</td>
<td>2.4kb</td>
<td>1.5kb</td>
<td>None</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Fix5</td>
<td>1.7kb</td>
<td>None</td>
<td>1.9kb</td>
<td>2.9kb</td>
<td>None</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The Southern hybridisation data makes it highly likely that the entire transcription unit of the gene is contained within the genomic DNA present in the 'phage recombinants Fix1-Genλ4. This cannot be confirmed, however, until the intron/exon structure of the gene is known. The fact that the cDNAa 5' 480bp probe hybridises to the 3.1kb and 3.4kb XbaI fragments of the Fix insert DNA (figure 3.13; table 3.1) tells us there must be an intron in this region. This is because there is no XbaI site present in the cDNAa fragment from which the probe was made.

3.2.6 ANALYSIS OF GENE STRUCTURE

To determine the precise intron/exon structure of the gene, a differential PCR analysis of cDNA and genomic DNA was carried out. For these PCRs, DNA preps from the 'phage recombinants Fix5 and Genλ2 were used as genomic DNA templates and a DNA prep from the NM1149 'phage recombinant containing the whole of cDNA11 as an insert was used as the cDNA template. PCR reactions were optimised
and carried out as described in Materials and Methods. Loading buffer was added and 1/10th of the entire reaction was run on 1.0% agarose gels. The results of this analysis are shown in figure 3.17, and a summary of the product sizes produced from each reaction is shown in table 3.3.

![Figure 3.17](image)

Figure 3.17
1.0% agarose gels showing the PCR products produced from recombinant λ bacteriophage containing either cDNA or genomic DNA inserts. The primers and template used in each reaction is indicated, as are the sizes of products produced. A summary of these results is presented in table 3.3. Three primer sets revealed the presence of intron sequences: 3.0kb between primers A & 3, 5.6kb between primers B & 4 and 0.3kb between primers D & 5. See text for details.

This data suggests the presence of 3 areas from which introns have been spliced out of the transcripts represented by cDNAa and cDNA11. The most 5' region lies between nucleotides +420 and +733 (a distance of 313 nucleotides) of cDNA11, and contains intron sequences of around 3kb in size. This confirms the presence of the putative intron which was suggested by the hybridisation pattern of the 5' 480bp probe shown in figure 3.13. The second lies between nucleotides +787 and +1608 (821 nucleotides) of cDNA11 and contains about 5.6kb of intron sequence. Finally, between nucleotides +2274 and +3010 (736 nucleotides) of cDNA11 there is a further 0.3kb of intron sequence.
### Table 3.3
Summary of results of differential PCRs shown in figure 3.17. The positions of the primers used are shown in figure 3.8. Three sets of primers revealed intron sequences; A & 3 (~3kb); B & 4 (~5.6kb); D & 5 (~300bp).

The differential PCR data does not discriminate between single and multiple introns between any two PCR primers. To determine the precise nature of the intron/exon structure of the gene, the genomic DNA-derived PCR products were gel purified and sequenced using dye-labelled cycle sequencing (ABI Prism - see Materials and Methods). The sequence reaction products were analysed on an automated DNA sequencer. This method yielded sequence covering both the 5' and 3' splice site junctions for the introns representing the 3kb and 0.3kb intron sequences, but not for the 5.6kb intron sequences.
The PCR product containing the 5.6kb intron sequences was around 6.5kb in length and it proved difficult to produce enough of this product to yield reliable sequence. However, standard sequencing of the λfs(1)13 construct reveals the presence of an intron/exon boundary at position +1338/+1339 in cDNA11. The boundary lies between PCR primers B and 4. Thus, if the 5.6kb of intron sequence between these primers is in fact a single intron, then the identified boundary is the 3' end of this intron. We have already seen that the probe made from the 3' 1557bp of cDNAa hybridises to the 1.5kb XbaI fragment of Fix5, but not to the 1.2kb fragment of Fix2 (figure 3.13; table 3.1). This places nucleotide position +1713 of cDNA11 between the 3' ends of the 1.5kb and 1.2kb fragments of Fix5 and Fix2, respectively. If we assume that this position lies as far upstream on the genomic DNA as possible (i.e. precisely at the 3' end of the 1.2kb fragment of Fix2), the identified intron/exon boundary lies 375bp upstream of this. Since Southern analysis shows that none of the sequence in cDNA11 hybridises to the 4.8kb XbaI restriction fragments of the Fix genomic DNA 'phage recombinants (data not shown), there must be a single intron of at least 4.8kb+[1.2kb-0.375kb]=5.625kb. Thus, it is highly likely that the central 5.6kb of intron sequences revealed by the differential PCR is, in fact, a single intron of at least 5.6kb in size (designated intron 2).

The sequence of the PCR product containing the 3kb of intron sequence shows that this is, in fact, a single intron (designated intron 1), with its intron/exon boundary at position +585/586 in cDNA11. Sequence of the PCR product containing the 0.3kb of intron sequence revealed two introns in this region. The first (intron 3) has its boundary at position +2448/2449 in cDNA11, and is 64bp in size. The final intron (intron 4) is 298bp in size and has its intron/exon boundary at position +2723/2724 in cDNA11. All four introns have donor and acceptor sites which match the consensus to a greater or lesser degree. The relevance of these sites will be discussed in chapter 4.

The relationship between cDNAa, cDNA11 and their corresponding genomic sequence is shown in figure 3.18. This shows that we have now isolated genomic
DNA containing the entire transcription unit which produces the transcripts represented by cDNAa and cDNA11.

**Figure 3.18**
Diagram showing relation of cDNAs a & 11 to cloned genomic DNA. The open reading frames of cDNAa and cDNA11 are represented by shaded areas. Actual sizes of all introns and exons are shown. The 5' end of exon 3 must lie at least 0.8kb downstream of the XbaI site at the 3' end of the 4.8kb genomic fragment, as indicated.
3.3 DISCUSSION

In this chapter I described the cloning of a gene from *Drosophila melanogaster* which generates three sex-specific transcripts; a 3.0 kb male-specific transcript, a 3.5 kb ovary-specific transcript and a 4.5 kb female carcass-specific transcript. This gene was cloned via a differential screen designed to isolate genes which give rise to sex-specific carcass transcripts. Therefore, this approach has clearly been successful. The production of sex-specific transcripts implies a sex-specific function and makes this gene a good candidate for a sex differentiation gene. We have provisionally named this gene *serine/threonine kinase 61 (stk61)*, based upon the cytogenetic location and sequence homology of the gene (see Chapter 5), until a name can be assigned upon the basis of phenotype.

Evidence presented in chapter 4 will show that the extended 3' UTR found only in cDNA11, is not present in either the 3.0 kb male-specific or 3.5 kb ovary-specific transcripts. As well as indicating that cDNA11 does in fact represent the 4.5 kb female carcass-specific, this suggests that the 4.5 kb female carcass-specific transcript may be subject to 3' UTR-mediated post-transcriptional regulation. The fact that both cDNAs a and 11 (and therefore the male and female transcripts) have exactly the same protein-coding potential, supports the possibility of a role for the extended 3' UTR of cDNA11 (and therefore the female carcass-specific transcript) in modulation of translation or localisation.

In *Drosophila*, the part played by the 3' UTR sequences of maternally-supplied oogenic transcripts in spatial localisation has been well documented. The specific localisation of certain mRNAs within the oocyte is required to provide correct protein expression domains, which lead to the determination of embryonic polarity (St Johnston & Nüsslein-Volhard, 1992). Furthermore, 3' UTR sequences have been shown to be sufficient for the correct localisation of *bicoid, fs(l)K10, nanos, oskar* and *orb* maternal mRNAs (Ding & Lipshitz, 1993; Decker & Parker, 1995). Although the system employed by these genes is oocyte-specific, and therefore
unlikely to be directly applicable to the female carcass-specific stk6l transcript, it
does demonstrates how 3' UTR sequences can modulate mRNA localisation, leading
to specific protein expression patterns. What may be of more relevance to stk6l is
the observation that the zygotically expressed pair-rule genes; fushi-tarazu, hairy and
even-skipped are specifically localised, in a 3' UTR-dependent manner, in the
cytoplasm of the cells in which they are expressed (Davis & Ish-Horowicz, 1991).
This shows that specific localisation of mRNAs is not limited to the germline in
Drosophila. It has been proposed that specific localisation of mRNAs to components
of the cytoskeleton may result in an increase in translation from those transcripts
(Decker & Parker, 1995). This is based on observations that both mRNAs and certain
components of the translation machinery are seen to be associated with filamentous
components of the cytoskeleton. Observations of mRNA localisation in sea urchin
eggs have indicated that mRNAs may dissociate from the cytoskeleton when inactive
(Moon et al., 1983). Thus, it may be that the 3' UTR in the female carcass-specific
transcript of stk6l is involved in up-regulation of translation by localising this
transcript to cytoskeletal components, thereby increasing the local concentration of
translation machinery components. A direct demonstration of 3' UTR-mediated
translational up-regulation comes from the gene for the human brain-specific
amyloid protein (De Sauvage et al., 1992). Two equally stable mRNA species are
produced from this gene which are differentially polyadenylated such that one
transcript contains an extended 3' UTR. The mRNA with the extended 3' UTR
produces more protein than the shorter message and the extended 3' UTR was shown
to induce increased translation from a reporter gene-3' UTR fusion. Conversely,
specific elements have been identified in the 3' UTRs of the maternally supplied
hunchback and bicoid transcripts which down-regulate translation under the action
of Nanos protein (Wharton & Struhl, 1991). These elements have been termed Nanos
response elements or NRE's.

Translational repression via the 3' UTR has been demonstrated for the mouse
protamine 1 (mP1) gene (Braun et al., 1989). The mP1 gene is transcribed in the
early stages of spermiogenesis but translation does not begin until around 1 week
later, when the spermatids are elongating. Construction of transgenic mice carrying a reporter gene fused to the \textit{mPl} 3' UTR has shown that the 3' UTR of \textit{mPl} directs this translational repression. A testis-specific 18 kDa protein has been shown to be directly responsible for the translational repression of transcripts from another protamine gene, \textit{mP2} (Kwon & Hecht, 1993). This protein is regulated by phosphorylation, being inactivated by dephosphorylation. The mechanism for this repression is not known as yet. However, it is interesting to note that when the \textit{mPl} 3' UTR is replaced by reporter gene sequences in the reporter gene-3' UTR fusion described above, mislocalisation of the transcript was concomitant with onset of translation (Braun et al., 1989). This suggests that the 3'UTRs of protamine genes may repress translation by sequestering the mRNA away from areas in the cell which have a high concentration of translation machinery components.

The stability of mRNAs can also be regulated by 3' UTR sequences. For example, the mammalian Transferrin receptor mRNA 3' UTR contains iron response elements (IRE's), acting as target sequences for a \textit{trans}-acting factor which protects the message from endolytic cleavage (Owen & Kühn, 1987; Müllner & Kühn, 1988; Koeller et al., 1989). A 3' UTR-mediated mRNA-stabilisation has also been proposed for the transcripts of the \textit{exuperantia} gene (Crowley & Hazelrigg, 1995—see chapter 1 for details). However, since we see no consistent difference in the levels of the three \textit{stk6l} transcripts, it seems unlikely that the extended 3' UTR of the female carcass-specific transcript acts to either stabilise or destabilise this transcript.

It is possible that the extended 3' UTR of the \textit{stk6l} female carcass-specific transcript acts to either promote or repress translation of the message. This may or may not involve specific cytoplasmic relocalisation. In general, translational regulation via 3' UTR sequences appears to be utilised early in development, when little transcription is occurring. Thus, if the function of the extended 3' UTR in the \textit{stk6l} female carcass-specific transcript is to modulate translation of STK61 protein, we might expect the female carcass-specific function of \textit{stk6l} to include early embryonic components. Preliminary whole mount \textit{in situ} RNA hybridisation evidence (D.
Clyde, pers. comm.) indicates that stk6l transcript is not expressed in a spatially restricted pattern in 3rd instar larva, with staining seen in the brain and all imaginal discs studied. However, this does not rule out the possibility of intracellular localisation of mRNA, or of specific STK61 protein localisation. It is also possible that hybridisation to the common stk6l transcript is masking specific localisation of the female carcass-specific transcript. The sex of the larvae used in these RNA in situ hybridisation experiments was not known. Further experiments, using sexed larvae, may reveal quantitative differences in signal, in specific tissues, which would indicate the presence of female carcass-specific transcript.

In the next chapter I will discuss evidence concerning further regulation of the stk6l transcripts, with a view to elucidating the possible role for stk6l in sex differentiation.
CHAPTER 4

REGULATION OF SEX-SPECIFIC TRANSCRIPTS PRODUCED FROM A NOVEL GENE IN DROSOPHILA MELANOGASTER
4.1 INTRODUCTION

Examination of Drosophila which carry mutations in genes of the sex determination hierarchy makes it clear that the genes which are responsible for differentiation of the vast majority of sex-specific features are under hierarchy control at some level. In the case of the yp genes, this control is direct, with Dsx protein acting as a transcriptional regulator of the genes. No other genes have yet been isolated which are under direct Dsx control. It is also clear that dsx is not epistatic to all sex-specific differentiation genes. Certain traits lie downstream of tra and tra-2, but not dsx, such as a pole cell sex-determining somatic signal, certain aspects of behaviour and developmental repression of the male-specific Muscle of Lawrence. Similarly, the overall size of the fly appears to be dependent upon Sxl, but not tra-2 or dsx, since XX/tra-2 flies, while being morphologically male, are of female size. Other than the yp genes, no non-gonadal sex differentiation genes have been cloned.

In the previous chapter we described the cloning of a gene which produces a male-specific transcript of 3.0kb and a female carcass-specific transcript of 4.5kb. We showed that this size difference was likely to be brought about by differential polyadenylation resulting in a female-specific 3'UTR. The production of sex-specific transcripts implies a sex-specific function for this gene, making it a good candidate for a sex differentiation gene. If this gene does indeed have a role in sex differentiation, we would expect it to be under the control of at least one of the genes of the sex determination hierarchy. In this chapter, I describe experiments using flies mutant for sex determination hierarchy genes, designed to test this hypothesis.
4.2 RESULTS

4.2.1 THE MALE-SPECIFIC TRANSCRIPT IS TESTIS-SPECIFIC

The production of a male-specific transcript from *stk6l* does present something of a paradox, as the gene was originally identified on the basis of preferential hybridisation to female cDNA. However, the sex-specific cDNA used for the initial screen contained no gonad-derived cDNA. If the male-specific transcript were, in fact, gonad-specific, it would account for the fact that the genomic recombinants screened did not show strong hybridisation to male non-gonadal cDNA. To test this hypothesis, gonads were dissected from male flies and PolyA⁺ RNA prepared from both gonad and carcass tissue. This RNA was used in Northern blot analysis, with a probe synthesised from cDNAa, as shown in figure 4.1.

![Figure 4.1](image)

Autoradiograph of a Northern blot prepared from hand-dissected OrR tissues. (~5µg of polyA⁺ RNA loaded per lane). The filter was hybridised with a probe synthesised from cDNAa. The 3.0kb male-specific transcript is only seen in RNA from male gonad tissue. The filter was stripped and re-probed with a probe synthesised from an 18s small subunit ribosomal protein cDNA. The male gonad lane appears to be underloaded relative to the other lanes. However, a good signal is produced from this lane using the cDNAa probe. Thus, the low levels of ribosomal protein transcript is probably a consequence of generally low levels of ribosomal protein gene expression in testis tissue.
It is clear from this that the male-specific transcript is entirely gonad-specific. To determine whether the transcript is expressed in testis or paragonial tissue, a similar Northern blot was carried out, but this time using total RNA prepared from accessory gland and testis tissue, as shown in figure 4.2.

![Northern blot image](image)

**Figure 4.2**

Autoradiographs of Northern blots prepared from hand-dissected male adult Oregon R tissue. Approximately 20μg of total RNA was loaded per lane. A probe synthesised from an 18s small subunit ribosomal protein cDNA was used as a loading check. The testis track appears to be underloaded from this. However, ribosomal proteins are generally of low abundance in the testis, and re-probing the same filter with a PD-10a cDNA (courtesy M. Wolfner, Cornell University, NY) representing an accessory gland-specific transcript, (a), shows that accessory gland RNA is present in sufficient quantity and quality. When this filter was probed with a probe synthesised from cDNAa, (b), male-specific transcript was only detected in RNA from testis tissue. No signal was seen in RNA from accessory gland tissue.

Male-specific stk6 transcript was only detected in testis RNA. The accessory gland lane appears to be underloaded compared to the testis lane. However, when the same filter was hybridised with a probe made from a cDNA, PD-10a, representing an accessory gland-specific transcript (courtesy M. Wolfner, Cornell University, NY), the correct band was clearly seen (figure 4.2a). The fact that the 3.0kb male gonad-specific transcript is entirely testis-specific has been confirmed by whole...
mount in situ hybridisation which exhibits signal in testis tissue, but not accessory
gland (D. Harbison, pers. comm.). It is clear, therefore, that the 3.0kb male-specific
transcript is, in fact, testis-specific. However, this does not show whether the 3.0kb
transcript is present in somatic or germline cells. In either case, its production is
highly unlikely to be governed by genes of the sex determination hierarchy. Studies
on yp expression in tra-2 temperature sensitive mutants show that, although yp
expression requires tra-2 in the fat body, tra-2 is not needed for yp expression in the
follicle cells of the ovary. Similar results have been obtained for transcripts
expressed specifically in the somatic component of the male gonad (see chapter 1 for
details). In the case of germline cells, a hierarchy gene-dependent signal is required
from the soma to determine a pole cell as female. However, once a pole cell has
been determined as male or female, it becomes hierarchy-independent. In the testis,
tra-2 is required for correct spermatogenesis to occur. This spermatogenic function
of tra-2 is independent of tra, and is due, at least in part, to a role in the regulation of
exuperantia (exu) transcripts. exu is required for correct gametogenesis to occur in
both sexes. A set of ovary-specific and testis-specific exu transcripts are produced,
the ovary transcripts having an intron spliced out of the 3' UTR which remains
present in the testis transcripts. It has been shown that tra-2 has a role in the
production of the testis exu transcripts, since they are produced less efficiently in
tra-2 mutants. Several 13 nucleotide repeats (13-nt repeats), similar to the repeats
present downstream of the dsx transcript sex-specific splice site, have been found in
and around the intron which is spliced from the exu ovary transcripts. In dsx, these
repeats are thought to act as binding sites for Tra and Tra-2 proteins which then
stabilise the splicing apparatus at the female-specific splice acceptor site (see chapter
1 for details). In exu, it is thought that Tra-2 may associate with the 13-nt repeats,
interfering with the splicing process and inhibiting splicing of the female-specific
intron. However, it is not yet known which of the identified 13-nt repeats, if any,
play a role in exu regulation in the testis. Tra-2 protein is also required in the testis to
inhibit the splicing out of the M1 intron from the tra-2 transcript. dsx-like 13-nt
repeats have also been identified in tra-2 transcripts and may play a similar role to
that which has been postulated for the repeats in exu.
For these reasons, we would not expect the 3.0kb transcript to be under the continual control of genes of the sex determination hierarchy, although there may be some level of Tra-2-mediated regulation, as has been suggested in the cases of the exu and tra-2 transcripts.

4.2.2 TRA-2 REGULATION OF THE TESTIS-SPECIFIC TRANSCRIPT

Figure 4.3 shows a Northern blot of RNA prepared from tra-2ts flies, hybridised with a probe made from cDNAa. In these mutants, tra-2 is functional at 16°C and produces no active product at 29°C. In wild type flies, the 3.0kb transcript is only present in XY (i.e. male=tra-2 OFF) flies. Figure 4.3 shows that it is also present in XX/tra-2ts flies which have been raised at 29°C (tra-2 OFF), but not present in XX/tra-2ts flies raised at 16°C (tra-2 ON). Although, at first glance, this seems to indicate tra-2 regulation of the transcript, the appearance of the transcript in XX/tra-2ts flies raised at 29°C is more likely to be due to the development of testes in these flies. Indeed, when we look at the gonads of XX/tra-2ts flies raised at 29°C (figure 4.3b), we see that testes do develop in these mutants.

XX/tra-2ts flies raised at 29°C develop as pseudomales; which is to say that they are male in all respects apart from their size and the fact that they are sterile, due to lack of the tra-2-mediated spermatogenic functions. Since tra-2 is required for correct germline development in the testis, the fact that the 3.0kb transcript from stk61 is present in testes from XX and XY tra-2ts homozygotes raised at 29°C (both of which lack active Tra-2) argues against, rather than for, any role for tra-2 in the production of this transcript. If there is no tra-2 regulation of this transcript, it suggests that the transcript is either involved in some tra-2-independent gametogenic function, or is not involved in gametogenesis at all but is actually expressed in the somatic component of the testis. However, since testis-specific exu transcripts are also found in testes lacking Tra-2, albeit at reduced levels, it is clear that whatever control tra-2 exerts over exu transcripts, it is not as absolute as the control which tra and tra-2
exert over the \( dsx \) transcript in the female fat body. Thus, we cannot rule out a role for \( tra-2 \) in the regulation of 3.0kb \( stk6l \) testis transcript.

![Diagram of Northern blot and photographs of gonads](image)

**Figure 4.3**

(a) Autoradiograph of a Northern blot prepared using hand-dissected tissues from flies homozygous for the \( tra-2^{ts} \) temperature-sensitive allele. The filter was hybridised with a probe synthesised from cDNA. Both male gonad-specific and female carcass-specific transcripts are present in RNA from \( XX/tra-2^{ts} \) flies. A probe made from an 18s small subunit ribosomal protein cDNA was used as a loading check. (b) Photographs of gonads from male OrR and \( XX/tra-2^{ts} \) flies. Magnification=×4. The gonads from both \( XY \) and \( XX/tra-2^{ts} \) flies contain apparently wild type accessory glands. There are also recognisable testes, although their structure is somewhat perturbed.

### 4.2.3 TRA REGULATION OF THE TESTIS-SPECIFIC TRANSCRIPT

The evidence from the \( tra2^{ts} \) flies makes it reasonable to suppose that any fly which develops testes will have the 3.0kb \( stk6l \) transcript present. Northern blots using \( tra^{FEM} \) mutants, which constitutively express the \( tra \) female-specific cDNA, and \( dsx \) null mutants show this to be the case.
Figure 4.4 shows a Northern blot analysis of RNA from \( trd^{\text{FEM}} \) mutant flies.

(a) Autoradiograph of a Northern blot prepared using hand-dissected tissues from OrR and \( trd^{\text{FEM}} \) homozygous flies. Approximately 5ug of polyA\(^+\) RNA was loaded per lane. The filter was hybridised with a probe synthesised using cDNA\(^{\text{5K}}\). A probe synthesised from an 18s small subunit ribosomal protein cDNA was used as a loading check.

(b) Photographs of gonads from OrR and \( XY/trd^{\text{FEM}} \) flies. Magnified 4X. This tissue is extremely perturbed in the \( trd^{\text{FEM}} \) mutants. Some accessory gland tissue is apparent, but no testis-like structures are seen.
Again, the probe was made using cDNAa as a template. Both male and female wild type flies produce active Tra-2. Thus, _XY/tra^FM_ flies develop as pseudofemales due to the presence of active _tra_ gene product. They appear female in morphology, although they are of wild type male size. Examination of their gonads (figure 4.4b) reveals that they do not develop recognisable testes, while there does appear to be fairly well developed accessory glands. As predicted, the 3.0kb _stk6l_ testis transcript is not present in these flies.

### 4.2.4 _DSX_ REGULATION OF THE TESTIS-SPECIFIC TRANSCRIPT

The 3.0kb _stk6l_ testis transcript is not present in _XY/dsx_ null mutants, as shown by the Northern blot in figure 4.5. Again, the gonads from these flies do not contain recognisable testes (figure 4.5b). Both _XY/dsx_ and _XX/dsx_ null mutant flies are intersexual, exhibiting a mixture of both male and female characteristics (Laugé, 1980). The flies are of wild-type size, but sex combs bristles are partially developed in both _XY/dsx_ and _XX/dsx_ flies (Hannah-Alava & Stern, 1957). The genitalia and analia are composed of both male and female features and male-like abdominal pigmentation is seen in both _XY_ and _XX dsx_ null mutants. The gonads of both _XX_ and _XY dsx_ flies contain male-like ducts and accessory glands, although the level of differentiation of these structures can vary from individual to individual. Well developed ovaries can develop in _XX/dsx_ flies, but ovarian structures are rarely seen in _XY/dsx_ individuals and, when present, are poorly differentiated (Laugé, 1980). The gonads from the _XY/dsx_ flies used in the Northern blot experiment shown in figure 4.5 contained no testes, as shown in figure 4.5(b). The 3.0kb _stk6l_ testis transcript was not seen in the RNA from flies of either _XX/dsx_ or _XY/dsx_ genotype.
Figure 4.5

(a) Autoradiograph of a Northern blot prepared using hand-dissected tissues from OrR and dsex mutant flies. Approximately 5ug of polyA+ RNA was loaded per lane. The filter was hybridised with a probe synthesised using cDNAa. No male gonad-specific transcript is seen in the dsex mutant flies. A probe synthesised from an 18s small subunit ribosomal protein cDNA was used as a loading check.

(b) Photographs of gonads from male OrR and XYdsex flies. Magnified 4X. The gonads from the XYdsex flies contain some recognisable accessory gland tissue, but no testis tissue is seen.
4.2.5 THE TESTIS-SPECIFIC TRANSCRIPT CONTAINS TCE ELEMENTS

Examination of the sequence of the cDNA thought to represent the testis transcript (cDNAa) provides some interesting clues as to the possible regulation of this transcript. In the 5' UTR of cDNAa there are two 12 nucleotide repeats which have similarity to an element known to play a role in translational control of testis-specific transcripts. This is known as the translational control element, or TCE (consensus=ACATCNAAAATT; see figure 4.6). In order to understand the relevance of this, it is necessary to discuss some aspects of spermatogenesis which were not covered in Chapter 1.

<table>
<thead>
<tr>
<th>Element</th>
<th>Sequence</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>ACATCNAAATT</td>
<td>12/12</td>
</tr>
<tr>
<td>TCE 1</td>
<td>AAATCGAAATTA</td>
<td>10/12</td>
</tr>
<tr>
<td>TCE 2</td>
<td>ACTTCAACATTG</td>
<td>9/12</td>
</tr>
<tr>
<td>TCE 3</td>
<td>ACTTCAAAATTT</td>
<td>11/12</td>
</tr>
</tbody>
</table>

**Figure 4.6**
Sequence of the 5' end of cDNAa, showing positions of TCE-like elements; TCE 1 & TCE 2. TCE 3 is present between introns 3 & 4 (figure 4.7). The homology of TCE 1, 2 & 3, to the consensus sequence is shown. Nucleotides which are perfect matches to the consensus are highlighted.

Radiolabelled uridine experiments show that virtually no transcription occurs after the primary spermatocyte stage of spermatogenesis (Olivieri & Olivieri, 1965). Thus, the transcripts required for protein synthesis during development of the haploid spermatids into functional spermatozoa (spermiogenesis) are present before
spermiogenesis begins. The Mst(3)CGP gene family comprises a set of genes which encode structural sperm tail proteins. These genes are transcribed in the primary spermatocyte (Kuhn et al., 1988). However, antibody in situ hybridisation to preparations from testis tissue, shows that translation of transcripts from members of this gene family does not occur until very late in spermiogenesis, when the spermatids are fully elongated (Schäfer et al., 1993). The Mst(3)CGP transcripts contain the 12-nt TCE element at an invariant position within the 5' UTR, 28bp downstream from the 5' end. Transgenic flies have been made using fusion constructs consisting of a LacZ reporter gene, incorporating 5' UTR sequence from a member of the Mst(3)CGP gene family at its 5' end (Kuhn et al., 1988). LacZ staining of gonads from male 3rd instar larvae (which do not contain spermatogenic cells beyond the primary spermatocyte stage) showed that no translation of LacZ was occurring. When the TCE was deleted from the fusion gene, however, LacZ staining could be seen at these stages. Thus, the TCE element is clearly regulating the onset of translation from the LacZ fusion transcripts. When transversion mutations of bases 5 and 7 (A<—>C) were present in the TCE, similar levels of LacZ expression were seen in 3rd instar gonads as when the TCE is completely deleted. This indicates that the integrity of nucleotides 5 and 7 in the TCE is vital to its function in regulating translation. Transversion of bases 9 and 10 (AT—>GC) also caused some loss of regulation but the effect was not as marked as with bases 5 and 7.

Gel shift experiments using in vitro transcript containing a TCE, revealed three bands from which three differently sized proteins could be purified (Kempe et al., 1993). These bands were only produced when the in vitro transcript was incubated with extracts from testis tissue. Extracts from any other tissue produced no bands. Only one protein could be purified following UV crosslinking, however, suggesting that a single protein may bind to the TCE which, when bound, attracts other proteins to it. The TCE element is clearly vital to the production of the gel-shifted bands, as only the largest gel shift band was seen when the in vitro transcript used contained the TCE carrying the base 5 and 7 transversions. This band was also relatively unstable, as shown by cold competitor controls. Since this mutant TCE also resulted
in loss of translational regulation of the LacZ fusion transcript, the protein binding revealed by the gel shift experiments is likely to be functionally relevant. Disruption of protein interactions with the in vitro transcript was also seen when transversions of bases 9 and 10, and base 12 (T→G) were present in the TCE, but to a lesser degree.

The two TCE-like elements present in the 5' UTR of the 3.0kb stk6l testis-specific transcript, are good matches to the consensus sequence. The upstream element (TCE 1) matches at 10/12 nucleotides and the downstream element (TCE 2) at 9/12. The mutational analysis of the Mst(3)CGP genes showed that nucleotides 5 and 7 were vital for TCE function. Both of the elements present in stk6l have these two nucleotides conserved. Nucleotides 9, 10 and 12 were also shown to affect TCE function when mutant, although less dramatically than positions 5 and 7. Both of the TCE-like elements in stk6l match the consensus at positions 9 and 10 but differ at position 12, with the upstream element having an A rather then a T and the downstream element exhibiting a G rather than a T. Although transversion at base 12 was shown to affect protein complex formation in the gel shift experiments described above (middle gel shift band did not form) the complexes which did form were quite stable, indicating that this base does not play as vital a role in TCE function as bases 5 and 7.

Another factor which is vital for the correct functioning of the Mst(3)CGP TCE elements is the position of the TCE element relative to the 5' end of the transcript. In all of the transcripts studied from this gene family, the TCE is invariably positioned 28bp from the 5' end. If the element is repositioned 54bp downstream of the 5' end, all TCE-dependent translational control is lost. In cDNAa, the TCE like elements lie 173bp and 359bp from the 5' end of the cDNA. A similar arrangement to this exists in another gene, Mst59D (Huang & Nöthiger, pers. comm. in Schäfer et al., 1995). The 5' UTR of the Mst59D transcript contains two TCE-like elements, also matching the consensus at 9/12 and 10/12 positions. This "TCE box" of Mst59D is located 60bp from the 5' end of the transcript, but the position of the box does not appear to
be vital to its translational control function, provided the overall integrity of the TCE elements within it is maintained. It seems likely, therefore, that the TCE elements present in the 5' UTR of the 3.0kb stk6l testis transcript represent a "TCE box" of the kind seen in Mst59D, regulating the translation of this transcript as has been described for the Mst(3)CGP gene family. It may be that the position-independent TCE box, containing two TCE elements, represents a separate class of TCE regulation from that seen in transcripts containing single, position-dependent, TCE elements. One model to account for this could be that a single TCE is unable to prevent translation once it has initiated at the 5' end of the transcript. Thus, it must be placed close to the 5' end to ensure that translation cannot initiate. However, if two TCE elements are present in tandem, they may be capable of "derailing" the translational apparatus as it proceeds down the transcript in search of the translational start site. This would account for the fact that the TCE box present in Mst59D is still operative even when repositioned, while the single TCE elements found in transcripts from the Mst(3)CGP gene family must remain at the extreme 5' end of the transcript to exert translational control.

The onset of translation of the transcripts from the Mst(3)CGP genes is associated with a secondary polyadenylation event which lengthens their polyA tails from 140 to 380 nucleotides (Schäfer et al., 1990). This polyadenylation event is also controlled by the TCE sequence. In the mutational analysis described above, this secondary polyadenylation was not seen either when the TCE was deleted, or when nucleotides 5 and 6 were mutated. This is interesting with regard to the 3.0kb testis transcript from stk6l. The cDNA thought to represent this transcript, cDNAa, is polyadenylated 1.2kb upstream from the site used to yield cDNA11, the cDNA thought to represent the 4.5kb female carcass-specific transcript. The secondary polyadenylation event observed for the Mst(3)CGP gene transcripts is an extension of the existing tail, rather than the utilisation of a second polyadenylation site. However, it may be that the utilisation of the upstream polyadenylation site in cDNAa occurs as part of a secondary polyadenylation under the control of the TCE sequences in the transcript. This could explain why the upstream polyadenylation
site is not used in the female carcass-specific transcript as shown by sequence analysis of cDNA11. The observation that gel shifts of TCE-containing transcripts only occur with testis extracts, indicates that the trans-acting factors which act at the TCE are testis-specific. Thus, if polyadenylation at the upstream site, as in cDNAa, is part of a TCE-mediated process, we would not expect to see the upstream site used in carcass-specific transcripts.

These considerations make it highly likely that the 3.0kb stk6 l testis transcript is expressed in the germline, playing some part in spermiogenesis.

4.2.6 REGULATION OF THE FEMALE CARCASS-SPECIFIC TRANSCRIPT

The initial objective of the differential screen, by which stk6l was isolated, was to identify putative sex differentiation genes. We might expect any non-gonadal sex-specific transcripts from such genes to be under the continual control of the genes of the sex determination hierarchy, as is the case for the yp genes. To investigate whether stk6l is under such control, we used Northern blots to examine the presence of the non-gonadal sex-specific 4.5kb transcript.

This analysis proved problematic due to difficulties connected with visualising the 4.5kb transcript on Northern blots. Firstly, this transcript is of very low abundance, necessitating the production of blots which were both low in background signal and of very high sensitivity. This was complicated by the low viability of some of the mutant strains used, making large scale isolation of polyA+ RNA impractical. However, the biggest problem was the presence of the common transcript which runs just above the 4.5kb transcript (see figure 3.10), and is present in RNA isolated from carcass and gonads of both sexes. On the blot shown in figure 3.10, a size difference of several hundred base pairs is seen between the 4.5kb transcript and the common transcript. However, in most cases, the common transcript was found to run so close to the 4.5kb transcript as to be inseparable on agarose gels.
The female carcass-specific transcript contains \( dsx \)-like 13-nt repeats.

The presence of the \( stk6l \) common transcript presents theoretical as well as practical problems. If cDNA11, which contains full protein-coding potential, represents the common transcript, then the sex-specific \( stk6l \) transcripts would apparently be redundant. For this reason, it is more probable that cDNA11 represents the female carcass-specific transcript and that the common transcript is, in fact, a partially processed RNA. Where size differences between the 4.5kb and common transcripts have been observed, the common transcript appears to be large enough to contain one or both of the smaller introns, intron 3 (64bp) and intron 4 (298bp), which are spliced out of the protein coding region of cDNAs a and 11. Thus, the common transcript may be identical to the transcript represented by cDNA11 but still containing intron 3 and/or intron 4 in the protein coding sequence, which would prevent active protein from being translated. Transcripts are not seen which are large enough to contain introns 1 and 2. However, these species would be far too large to be visualised using standard RNA preparation and Northern blotting methods.

Since work on the \( yp \) genes indicates that sex-specific gene expression in non-gonadal tissue is most likely to be under continual hierarchy gene control, we were particularly interested in looking at the regulation of the carcass-specific transcripts of \( stk6l \). Figure 3.10 shows that male carcass tissue contains only the common transcript, while female carcass tissue contains the 4.5kb female carcass-specific transcript and the common transcript. As we have suggested, the female carcass transcript may in fact be a fully processed form of the common transcript, differing only in the splicing of intron 3 and/or intron 4. The question which arises is; why would these introns not be spliced out of the common transcript in male carcass tissue?

This situation is similar to that discussed in chapter 1 for the \( dsx \) transcript. In the case of \( dsx \), a suboptimal splice acceptor site at the 3' end of the sex-specific intron is stabilised by Tra and Tra-2 proteins so that this site is only used in females. Tra and
Tra-2 act at a series of six tridecamer repeats (13-nt repeats) downstream of the splice acceptor site. Examination of the *stk6l* sequence surrounding introns 3 and 4 provides some interesting clues as to what may be happening to regulate the expression of the *stk6l* female carcass-specific transcript.

Two sites with homology to the *dsx* 13-nt repeats are present downstream of *stk6l* intron 4, lying at positions +2895 (*dsx* 13-nt rep 1) and +2942 (*dsx* 13-nt rep 2), as shown in figure 4.7. Also shown in this figure is a third TCE-like element, the relevance of which will be discussed later in this chapter.

The upstream *dsx*-like 13-nt repeat matches the *dsx* consensus at 9/13 positions and the downstream repeat at 8/13 positions. The *dsx*-like 13-nt repeats found in the *exu* gene exhibit similar levels of homology to the *dsx* 13-nt repeats. What is also interesting, is the similarity which the two *stk6l* repeats bear to each other, with the first 9 nucleotides of both repeats being identical. This strongly suggests that these two repeats are, in fact, functionally relevant. The consensus sequence for a splice acceptor site has been identified as Y,N,CAG, where Y is a pyrimidine base and N is T, C, G or A (Ohshima & Gotoh, 1987). In *dsx*, the female-specific splice acceptor
site contains only 6/11 pyrimidine bases in the Y{sub}11, or polypyrimidine (polyY), tract (Burtis & Baker, 1989). The other, non Tra/Tra-2-regulated, acceptor sites have better matches to the consensus, containing 8-10 Y's in their polyY runs. Introduction of an 18nt polypyrimidine tract into the dsx female-specific splice acceptor site enables this site to be efficiently utilised, even in the absence of Tra and Tra-2 (Hoshijima et al., 1991). Therefore, in dsx, the reason why the female-specific splice site is not used in the absence of Tra and Tra-2 is that it lacks a good polyY tract, making it a poor substrate for the splicing machinery. The sequences of the intron donor and acceptor sites within stk6l are shown in table 4.1.

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Donor</th>
<th>Acceptor</th>
<th>Y-tract Ycontent</th>
</tr>
</thead>
<tbody>
<tr>
<td>(C/A)AG/GTAAGT</td>
<td>YYYYYYYYYYY N CAG/(G/A)</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Intron 1</td>
<td>AAC/GTAAGT</td>
<td>CATATATTTTT C CAG/A</td>
<td>8</td>
</tr>
<tr>
<td>Intron 2</td>
<td>XXX/XXXXXX</td>
<td>ATCTTCTCCTT G CAG/T</td>
<td>10</td>
</tr>
<tr>
<td>Intron 3</td>
<td>AAG/GTAAGT</td>
<td>TTTGTGTTCTT T CAG/A</td>
<td>10</td>
</tr>
<tr>
<td>Intron 4</td>
<td>AGG/GTAAGT</td>
<td>TTGGTTCTCATG G CAG/C</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 4.1
Donor and acceptor splice sites from introns 1-4 of cDNAs a & 11. Nucleotides which are perfect matches to the consensus are highlighted. The number of pyrimidine nucleotides within the polypyrimidine tract of the splice acceptor sites are shown. Y=pyrimidine; N=T, C, G or A; X=nucleotide at this position not known.

The acceptor site which is the poorest match with the consensus sequence is that of the most downstream intron, intron 4. As well as having the lowest overall number of Y's in the polyY run, it lacks a consecutive run of Y's longer than 4. As we can see from figure 4.7, the two dsx-like 13-nt repeats lie 172bp downstream of this intron. In dsx, the most upstream of the 13-nt repeats lies 295bp downstream of the regulated splice site. Thus, the two stk6l dsx-like 13-nt repeats are ideally positioned to be involved in the regulation of the intron 4 acceptor site. The intron 3 acceptor site is a very good match to the consensus, having 10/11 Y's in the polyY run and so is unlikely to require any additional regulation to be efficiently utilised. Thus, if the common transcript is a partially spliced form of the female carcass-specific transcript, it is most likely to contain only intron 4.
The Northern in figure 3.10 shows that the common transcript from stkl is still present, even in female carcass tissue. If the common transcript is indeed a partially spliced form of the female carcass-specific transcript, then we can see that splicing rarely occurs to completion. The Northern in figure 4.8 has been hybridised with a probe made from a dsx cDNA. This probe hybridises to both male and female-specific dsx transcripts. We can see here that, in the female, none of the transcript remains in the unspliced (i.e. male) form. The smaller 2.9kb male-specific band is a minor product of the dsx gene in males, and is thought to be generated by use of a second polyadenylation site located upstream of that used in the 3.9kb male-specific transcript (Burtis & Baker, 1989). Hence, this minor product is not relevant to this discussion.

Figure 4.8
Autoradiograph of a Northern blot prepared using hand-dissected tissues from OrR and dsx mutant flies. Approximately 5ug of polyA+ RNA was loaded per lane. The filter was hybridised with a probe synthesised using a dsx cDNA. Both mutant and wild type flies produce the expected male-specific and female-specific transcripts. After a longer exposure, both transcripts can also be seen at greatly reduced levels in the gonads. In the female, splicing at the female-specific acceptor site occurs to completion since only the 3.5kb transcript is seen in RNA from these flies. A probe synthesised from an 18s small subunit ribosomal protein cDNA was used as a loading check.
If Tra and Tra-2 are acting on transcripts from \(stk6\), as they do in \(dsx\), then why is it that common transcript (putatively unspliced form) is still present in female tissue? It is interesting to note that if all but two of the 13-nt repeats are deleted from the \(dsx\) transcript, a significant drop in the utilisation of the the female-specific splice site is observed (Hoshijima et al., 1991). This might explain why the \(stk6\) common transcript is still seen on Northern blots, in female carcass tissue. Another possibility is that competition between the acceptor sites of introns 3 and 4 for components of the splicing apparatus reduces the efficiency of intron 4 splicing. The two introns are only separated by 374bp, and the intron 3 acceptor site is a vastly more optimal site than that of intron 4 (figure 4.7; table 4.1). Interestingly, in \(dsx\), when the female-specific acceptor site is deleted, several downstream cryptic acceptor sites start to become utilised (Ryner & Baker, 1991). In transcripts from the wild type gene, these cryptic sites are never used. This is presumably due to the fact that the female-specific site is a better match to the consensus sequence and efficiently competes with the cryptic sites for splicing apparatus components (in the presence of Tra and Tra-2). Thus, a model where the efficacy of \(stk6\) intron 4 splicing is hindered by the superior efficiency of intron 3 splicing, is clearly a strong possibility. It may be that the combination of a sub-optimal polyY run and a the competition with intron 3, makes the splicing of intron 4 a very inefficient process, such that the aid of Tra and Tra-2, acting at two downstream 13-nt repeats, is required to enable any fully spliced transcript to be produced.

4.2.6.2 \(tra-2\) regulation of the female carcass-specific transcript.

In an attempt to determine whether \(tra-2\) is involved in production of the \(stk6\) 4.5kb female carcass-specific transcript, a Northern blot was performed using RNA extracted from \(tra-2^{\text{m}}\) mutant flies (figure 4.3). If, as we suggest, the common transcript is spliced under the control of Tra/Tra-2 proteins, to produce the 4.5kb female carcass-specific transcript, we would not expect the 4.5kb transcript to be produced in the absence of Tra-2. However, we do see a very small amount of 4.5kb transcript in RNA prepared from carcasses of \(XX/tra-2^{\text{m}}\) flies raised at 29°C (\(tra-2^{\text{m}}\) flies raised at 29°C
inactive). This result does not rule out the possibility of \textit{tra-2} involvement, since temperature sensitive mutations are more likely to be hypomorphic rather than completely amorphic. Thus, there may be enough active \textit{tra-2} transcript present at the restrictive temperature to enable some splicing to occur. If this is the case, we might expect to be able to detect \textit{yp} transcripts in \textit{XX/tra-2°} female carcass tissue at the restrictive temperature. However, when the same filter was re-probed with a \textit{yp} probe, no \textit{yp} transcript was detected from the carcass tissue of \textit{29°C XX/tra-2°} flies (figure 4.9).

The fact that no \textit{yp} transcripts are detected in \textit{XX/tra-2°} flies raised at 29°C does not mean that there is no active \textit{tra-2} transcript present in these flies, however. When \textit{XX/tra-2°} flies are reared at 18°C and then shifted to 29°C for several days, no \textit{YP

Figure 4.9

Autoradiograph of a Northern blot prepared using hand-dissected tissues from \textit{tra-2°} mutant flies. Approximately 20\textmu g of total RNA was loaded per lane. The filter was hybridised with a probe synthesised using a cDNA representing the \textit{yp} transcript. No \textit{yp} transcript is detected in RNA from tissues of \textit{XX/tra-2°} flies raised at the restrictive temperature of 29°C. As a loading check, the filter was stripped and re-hybridised with a probe synthesised using an 18s small subunit ribosomal protein cDNA.
protein is detected (Bownes et al., 1987). However, female-specific \( dsx \) transcript can still be detected in these flies (Nagoshi et al., 1988). Thus, there is clearly enough active \( tra-2 \) transcript present in these flies to enable a small amount of splicing to occur at the female-specific acceptor site of \( dsx \). However, sufficient levels of \( dsx \) transcript are not produced to enable de-repression of the \( yp \) genes. Thus, we cannot rule out a role for \( tra-2 \) in the splicing of \( stk6l \), since there may be sufficient levels of active \( tra-2 \) transcript present to account for the levels of female carcass-specific transcript present in \( XX/tra-2ts \) flies raised at 29°C. We can, however, say that the production of the female carcass-specific transcript is not likely to be under direct \( dsx \) control, as the regulation of this transcript is clearly different from that of the \( yp \)'s.

### 4.2.6.3 \( tra \) regulation of the female carcass-specific transcript

A Northern blot was prepared using RNA from \( trd^{FEM} \) mutant flies (figure 4.4). Flies of the genotype \( XY/trd^{FEM} \) ectopically express the female form of \( tra \) and so develop as pseudofemales. If the production of the \( stk6l \) female carcass-specific transcript is dependent upon \( Tra/Tra-2 \), we would expect to see this transcript in the carcass tissue of \( XY/trd^{FEM} \) flies. Figure 4.10 shows that \( XY/trd^{FEM} \) flies are clearly genetically female, expressing high levels of \( yp \) transcripts. In fact, higher levels of \( yp \) expression are seen in \( XY/trd^{FEM} \) pseudofemales than in wild type females.
Figure 4.10

Autoradiograph of a Northern blot prepared using hand-dissected tissues from OrR and $trd^{FEM}$ mutant flies. Approximately 5ug of polyA+ RNA was loaded per lane. The filter was hybridised with a probe synthesised from a $yp2$ cDNA. A high level of $yp2$ transcript was detected in RNA from $XY/trd^{FEM}$ flies. No $yp2$ transcript was detected in RNA from OrR male flies. A probe synthesised from an 18s small subunit ribosomal protein cDNA was used as a loading check.

This is in agreement with studies of YP protein expression levels using extracts from $trd^{FEM}$ mutants (M. Bownes, pers. comm.). Again, interpretation of the result shown in figure 4.4 is hampered by the fact that the $stk61$ common and female carcass-specific transcripts are poorly separated. The experiment was repeated several times, but good separation between these two transcripts was never observed. However, there is a very noticeable increase in band intensity in the female carcass as opposed to the male lanes (figure 4.4, lanes 1 & 2). The loading check shows that all lanes are equally loaded, and so this increase in band intensity is presumably due
to the presence of the female carcass-specific transcript. Comparing wild type male tissue with carcass tissue from $XY/trd^{FEM}$ flies (figure 4.4, lanes 1 & 4), we see the same increase in intensity. In fact, there is an even greater increase in intensity in $XY/trd^{FEM}$ flies than with wild type females. These bands have been quantitated, which allows us to visualise these increases in band intensity more clearly (figure 4.11). The loading check bands were also quantified, and these results were used to correct the figures depicted in figure 4.11 for small variations in loading.

Application of the $\chi^2$ test shows a probability of 50-70% that these increases are significant, where a probability of >5% is generally taken as indicative of significance. Although the nature of these results means they cannot be taken as conclusive proof, they are strongly indicative of a role for $tra$ and $tra-2$ in the production of the $stk61$ female carcass-specific transcript.
4.3 DISCUSSION

Taken together, the above evidence enables us to propose models for the regulation of the 3.0kb male-specific and 4.5kb female carcass-specific transcripts produced from *stk6l*.

Northern analysis suggests a role for the sex determination genes *tra* and *tra-2* in the production of the *stk6l* female carcass-specific transcript. cDNA sequence data indicates that this control is direct, with Tra and Tra-2 proteins acting at two *dsx*-like 13-nt repeats to enable the removal of intron 4. RT-PCR analysis was performed using primers which flank intron 4, in an attempt to confirm this hypothesis (data not shown). However, the results from this were inconclusive and a more reliable technique is required to answer this question. One such technique may be RNAase protection, which has not yet been attempted. A riboprobe could be made to a cDNA region including the intron 4 intron/exon boundary. When hybridised to extracted RNA and digested with RNAase, single stranded RNA would be destroyed and species diagnostic of the fully spliced transcript would be produced. If this *tra/tra-2* regulation hypothesis is correct, these diagnostic bands would be seen using RNA from female carcass but not from male carcass.

An RNAase protection-based assay for the presence of the *stk6l* female carcass-specific transcript would also enable the *tra/tra-2* regulation of the gene to be demonstrated more conclusively. Northern analysis is not a convenient method to determine the presence of this transcript, due to very low levels of expression and the complication of the common transcript. The rarity of the *stk6l* transcripts can be clearly seen by comparison of figures 4.4(a) and 4.9. To produce the bands seen in figure 4.4(a), the filter was hybridised with a gene-specific probe and exposed to film for two weeks. Figure 4.9 shows the same filter, which was stripped and hybridised with a *yp* probe. An exposure of only 2-3 hours was required to produce the *yp* bands shown here. This demonstrates the extreme rarity of the transcript. Other probes, such as *dsx*, were also used (figure 4.8) and no significant problems were
encountered in detecting transcripts from this gene. Thus, Northern blots had to be produced which were sufficiently free of background signal to enable long periods of exposure, as well as being of high sensitivity. Numerous blots were carried out, only a fraction of which yielded useful data. This was further complicated by the inviability of several of the mutants fly stocks. For the \textit{dsx} and \textit{tra-2"} mutants, trans-heterozygotes had to be constructed, using two separate alleles, to enable sufficient numbers of mutant flies to be obtained. This means that continual repetition of Northern blots becomes a very time consuming process. Thus, it is clear that a more amenable assay for the presence of the female carcass-specific transcript must be developed. RNAase protection should provide a sensitive and reliable assay which would make the characterisation of the \textit{stk61} female carcass-specific transcript less problematic.

The \textit{stk61} male-specific transcript has been shown to be entirely testis-specific. The presence of TCE elements in the 5' UTR make it highly likely that this transcript is germline-specific and is translated late in spermiogenesis. The onset of translation may be coupled to a secondary polyadenylation event which effectively removes the longer 3' UTR seen in cDNA11.

A model whereby Tra and Tra-2 are required for the splicing of \textit{stk61} intron 4 presents difficulties with regard to the testis transcript. The cDNA thought to represent the testis transcript, cDNAa, does not contain intron 4. Tra-2 alone has been shown to be able to promote splicing from the \textit{dsx} female-specific acceptor site (Hoshijima \textit{et al.}, 1991; Ryner & Baker, 1991). Thus, production of the testis transcript in wild type testes, which express \textit{tra-2}, is not unexpected. However, this transcript is also produced at similar levels in the testes of both \textit{XX} and \textit{XY/tra-2"} mutant flies raised at the restrictive temperature. In this regard, it is interesting to note the presence of a third TCE-like element present in cDNAs a and 11. This element lies at position +2698 in cDNA11, between introns 3 and 4 as shown in figure 4.7, and shows very high homology to the consensus, matching at 11/12 positions. All of the nucleotides which have been identified as important to TCE
function (nt's 5, 7, 9, 10 and 12) match the consensus in this element. We have already considered the possibility that competition between the optimal acceptor site of intron 3 and the sub-optimal site of intron 4 is one reason why utilisation of the intron 4 acceptor site might be an inefficient process. The gel shift and UV-crosslinking experiments previously discussed show that certain testis-specific factors associate with the TCE sequence. Since the third TCE element lies 250bp downstream of the intron 3 acceptor site, it is possible that, in the testis, association of TCE-specific proteins with the element, interferes with the efficiency of the intron 3 acceptor site. This might reduce competition between the sites of introns 3 and 4 to a level where the intron 4 acceptor site is utilised in the absence of tra or tra-2 function. Alternatively, the splicing of the testis-specific stk6l transcript may be under the control of testis-specific factors. This type of tissue-specific RNA processing is seen in the transposase-encoding transcript from the Drosophila P-element, where the intron between open reading frames 2 and 3 is only removed from the primary transcript in the germline (Laski et al., 1986). Thus, active transposase is only produced in this tissue.

In summary, both the testis-specific and female carcass-specific stk6l transcripts appear to be regulated in a tissue-specific and sex-specific manner. It is likely that both transcripts give rise to active protein in their respective tissues which play differential roles in spermiogenesis and female somatic sex differentiation. Precisely what functions these proteins may be performing will be considered in the next chapter.
CHAPTER 5

SEQUENCE AND BACTERIAL EXPRESSION ANALYSIS
OF THE PREDICTED PROTEIN FROM A DROSOPHILA MELANOGASTER
GENE ENCODING SEX-SPECIFIC TRANSCRIPTS
5.1 INTRODUCTION

In the previous two chapters we have described the cloning of a *Drosophila melanogaster* gene which encodes three different sex-specific transcripts; a 3.0kb testis-specific transcript, a 3.5kb ovary-specific transcript and a 4.5 female carcass-specific transcript. We have discussed evidence which suggests that the testis transcript encodes a protein which is translated late in spermiogenesis and is likely to be involved in the latter stages of sperm differentiation. We have also provided evidence which supports a model whereby the production of the fully-processed female carcass-specific transcript is under direct control of the sex determination genes *tra* and *tra-2*. As yet, we have no direct evidence as to the regulation mechanism of the ovary transcript. These findings suggest that the proteins encoded by the different sex-specific transcripts are functionally active. In this chapter, the possible roles of these proteins in sex differentiation are discussed, including analysis of the predicted protein sequence and bacterial expression studies.
5.2 RESULTS

5.2.1 PREDICTED PROTEIN SEQUENCE ANALYSIS.

All computer sequence analysis was carried out using the University of Wisconsin Genetics Computer Group, Sequence Analysis Software Package, Version 7 (Devereux et al., 1984). The predicted protein primary sequence encoded by cDNAs a and 11 is shown in figure 5.1. In an attempt to assign a possible function to this protein, this sequence was used to perform exhaustive computer searches of the Genbank gene database. Initially, no significant homologies to known gene products were identified. Since it is possible that small areas of homology could be missed when using the entire sequence as a basis for comparison, the computer searches were repeated using smaller sections of the sequence. This method revealed that certain short runs of amino acids within the sequence were highly homologous to parts of protein kinase catalytic domains from a number of different genes. Using a mixture of computer searching and alignment by eye, it was found that the predicted protein from cDNAs a and 11 contained a complete kinase catalytic domain.

The eleven subdomains (I-XI) which make up this catalytic domain, are indicated in figure 5.1. Thus, the transcripts produced from \( \text{stk61} \) encode a novel protein kinase. In a comparative study of 38 serine/threonine-specific and 27 tyrosine-specific protein kinases, it was found that these eleven regions were present in all 65 primary protein sequences (Hanks et al., 1988). One of the most interesting regions of homology lies in subdomain VI, where the consensus sequence D-L-A-A-R-N is diagnostic of tyrosine-specific protein kinases, while the consensus D-L-K-P-E-N indicates serine/threonine-specificity. In figure 5.1, we can see that subdomain VI contains the sequence D-L-K-P-E-N indicating that this protein is a serine/threonine-specific protein kinase (S/T kinase). Within the catalytic domain, 19 residues were found to be invariant or nearly invariant in all of the S/T kinases analysed. All of these amino acids are conserved in the predicted protein from \( \text{stk61} \) and have been boxed in figure 5.1. A further 18 positions within the domain
exhibited amino acids of similar chemical nature in all 38 S/T kinases. Again, the nature of these residues is conserved within the primary sequence of STK61 protein and their positions have been underlined in figure 5.1.

**Figure 5.1**
Predicted sequence of STK61 protein. Serine/threonine-specific protein kinase catalytic subdomains I-XI are indicated. Residues which were found to be invariantly conserved in a study of 38 S/T kinases (see main text for details) are boxed. Underlining indicates positions where amino acids of similar chemical character are found.
These conserved residues are thought to play a pivotal role in the functioning of the catalytic domain (Hanks et al., 1988). The conserved glycine and valine residues in subdomain I (boxed in figure 5.1) have been proposed to be involved in the binding of ATP in the tyrosine-specific kinase (Y kinase) Src. Mutation of the conserved lysine in subdomain II abolishes the kinase activity of Src and is believed to be involved in the transfer of the phosphor group from ATP. The conserved aspartate (D) and asparagine (N) residues in subdomain VI, and the conserved "DFG" triplet in subdomain VII are also thought to be involved in ATP binding. In Src, the three residues representing the conserved "APE" triplet in subdomain VIII (corresponding to "SPE" in figure 5.1) are all required for kinase function.

It is difficult to propose what function the protein shown in figure 5.1 may be carrying out on the basis of homology to other protein kinases. This is because the homology within the kinase domain is highly localised and no significant homology is seen outwith the catalytic domains. The subdomains of the kinase catalytic domain which show the greatest levels of conservation are subdomains I, II, VI, VII, and VIII. Since some of the conserved residues within these subdomains have been shown to be required for kinase activity, it is likely that these domains are responsible for the regulation and catalytic activity of the active site, while the other, less conserved, regions from less critical structural components of the protein. Thus, in an attempt to categorise the predicted STK61 protein, the sequences of subdomains I, II, VI, VII and VIII were compared with those from a variety of different kinases know to undergo distinct forms of regulation and perform disparate functions (figure 5.2).
Predicted protein sequence comparison of STK6I catalytic kinase domains with those from 21 other S/T kinases. Positions of identity to STK6I are highlighted. For each sequence, the number and percentage of identities to STK6I are shown. cAPK-α/β=bovine cAMP-dependent kinase subunits, SRA3=Yeast suppressor of Ras cAMP-dependent kinase, PKC-α/β/γ=Bovine protein kinase C subunits, CaMII-α/β=Rat Ca \(^{2+}\)/calmodulin-dependent kinase subunits, PhK-γ=Rabbit phosphorylase kinase subunit, SNF1=Yeast "sucrose nonfermenting" kinase, Nim1=Yeast "new inducer of mitosis" kinase, KIN1=Putative yeast protein kinase, CDC28=Yeast cell-division-cycle 28 kinase, CDC2=Yeast cell-division-cycle 2 kinase, CDC2Hs=Human homolog of yeast CDC2, Raf=Human Raf kinase, A-Raf=Human Raf-related oncogenic kinase, PKS=Human Raf-related cellular kinase, DCKII=\textit{Drosophila} casein kinase II α subunit, CDC7=Yeast cell-division-cycle 7 kinase, STE7=Yeast "Sterile" 7 kinase.

<table>
<thead>
<tr>
<th>Subdomain</th>
<th>Identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>cAPK-α</td>
<td>DQFERIKTLGTSFGRVMLVKHME</td>
</tr>
<tr>
<td>cAPK-β</td>
<td>GDFERKKTGTSFGRVMLVKHKA</td>
</tr>
<tr>
<td>SRA3</td>
<td>KNFQILRTLGTSFGRVLHIRSRH</td>
</tr>
<tr>
<td>PKC-α</td>
<td>TDFNFLMVLKGSFGRMLADRKG</td>
</tr>
<tr>
<td>PKC-β</td>
<td>TDFNFLMVLKGSFGRMLSERKG</td>
</tr>
<tr>
<td>PKC-γ</td>
<td>SDFSFLMVLKGSFGRMLAERRG</td>
</tr>
<tr>
<td>CaMII-α</td>
<td>EEYQIFEELGKAFSVRRVCVKL</td>
</tr>
<tr>
<td>CaMII-β</td>
<td>DEYQLYEDIGKAFSVRRVCVKL</td>
</tr>
<tr>
<td>PhK-γ</td>
<td>ENYEPKELGRRGSSVRRCICHK</td>
</tr>
<tr>
<td>SNF1</td>
<td>GNYQIVKLGEVSFGKVKLAYHTT</td>
</tr>
<tr>
<td>Nim1</td>
<td>GVWRLKLTGLTGSTSVRLAKHAK</td>
</tr>
<tr>
<td>KIN1</td>
<td>GDWEBVETVAGSMGKVKLAKHRY</td>
</tr>
<tr>
<td>CDC28</td>
<td>ANYKRLEKVGETYGVYKALDLR</td>
</tr>
<tr>
<td>CDC2</td>
<td>ENYQKVEKIGETYGVYKARHKL</td>
</tr>
<tr>
<td>CDC2Hs</td>
<td>EDYTKEIKIGETYGVYKGRHKT</td>
</tr>
<tr>
<td>Raf</td>
<td>SEVMLSTRIGGSFGTYYKGRWHG</td>
</tr>
<tr>
<td>A-Raf</td>
<td>SEVQLKRICTGSFGRVFRGWHG</td>
</tr>
<tr>
<td>PKS</td>
<td>SEVQLKRICTGSFGRVFRGWHG</td>
</tr>
<tr>
<td>DCKII</td>
<td>DDYQLVKLRGKYESVFRAINIT</td>
</tr>
<tr>
<td>CDC7</td>
<td>NEYKLIDKIGETFSSVYKAKIT</td>
</tr>
<tr>
<td>STE7</td>
<td>QDLVQLGKAGNSGTVVKALHVP</td>
</tr>
<tr>
<td>STK61</td>
<td>NDFIFGTYEGSIVLYAVDIH</td>
</tr>
</tbody>
</table>

(continued...)
Figure 5.2(b)

<table>
<thead>
<tr>
<th>Subdomain II</th>
<th>Identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>cAPK-α</td>
<td>TGNHYMKILDKQKV 4 25%</td>
</tr>
<tr>
<td>cAPK-β</td>
<td>TEQYYMKILDKQKV 4 25%</td>
</tr>
<tr>
<td>SRA3</td>
<td>NGRYYMKVKKEIVV 6 38%</td>
</tr>
<tr>
<td>PKC-α</td>
<td>TEELYAAILKMDVVI 5 31%</td>
</tr>
<tr>
<td>PKC-β</td>
<td>TDELYAVKILKMDVVI 4 25%</td>
</tr>
<tr>
<td>PKC-γ</td>
<td>SDELYAAILKMDVIV 7 44%</td>
</tr>
<tr>
<td>CaMII-α</td>
<td>AGQEYAAKIIINTKKLS 4 25%</td>
</tr>
<tr>
<td>CaMII-β</td>
<td>TGHEYAAKIIINTKKLS 4 25%</td>
</tr>
<tr>
<td>PhK-γ</td>
<td>TCKEYAVKIIVTVGGG 4 25%</td>
</tr>
<tr>
<td>SNF1</td>
<td>TGQKVALKIIINKVALA 3 19%</td>
</tr>
<tr>
<td>Nim1</td>
<td>TGDIAAIKIIPIR---4 25%</td>
</tr>
<tr>
<td>KIN1</td>
<td>TNEVCAVKIVNRTDKA 2 13%</td>
</tr>
<tr>
<td>CDC28</td>
<td>GQRVVALKIRLESED 3 19%</td>
</tr>
<tr>
<td>CDC2</td>
<td>SGRIVAMKIRLEDES 4 25%</td>
</tr>
<tr>
<td>CDC2Hs</td>
<td>TGQVVAMKIRLESEE 2 13%</td>
</tr>
<tr>
<td>Raf</td>
<td>---DVAVKILKVDPFT 2 13%</td>
</tr>
<tr>
<td>A-Raf</td>
<td>---DVAVKVLVSPQPT 3 19%</td>
</tr>
<tr>
<td>PKS</td>
<td>---DVAVKLVSPQPT 3 19%</td>
</tr>
<tr>
<td>DCKII</td>
<td>TTEKCVVKILPKVKK 1 6%</td>
</tr>
<tr>
<td>CDC7</td>
<td>GSNYVALKIVYVTS---2 13%</td>
</tr>
<tr>
<td>STE7</td>
<td>DSKIVAKTIPVEQNN 2 13%</td>
</tr>
<tr>
<td>STK61</td>
<td>SRREYAIKVCEKRLIL 16 100%</td>
</tr>
</tbody>
</table>

(continued...)
Figure 5.2(c)

<table>
<thead>
<tr>
<th>Subdomain VI</th>
<th>Identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>cAPK-α</td>
<td>FSEPFRFAYAAQIVLTFEVLHSL-DLITRDLKPNELL1DDQG 17 40%</td>
</tr>
<tr>
<td>cAPK-β</td>
<td>FSEPFRFAYAAQIVLTFEVLHSL-DLIYRDLKPNELL1DHQG 17 40%</td>
</tr>
<tr>
<td>SRA3</td>
<td>FSEPFRFAYAAQIVLTFEVLHSL-DLIYRDLKPNELL1DDQG 17 40%</td>
</tr>
<tr>
<td>PKC-α</td>
<td>FSEPFRFAYAAQIVLTFEVLHSL-DLIYRDLKPNELL1DDQG 17 40%</td>
</tr>
<tr>
<td>PKC-β</td>
<td>FSEPFRFAYAAQIVLTFEVLHSL-DLIYRDLKPNELL1DDQG 17 40%</td>
</tr>
<tr>
<td>PKC-γ</td>
<td>FSEPFRFAYAAQIVLTFEVLHSL-DLIYRDLKPNELL1DDQG 17 40%</td>
</tr>
<tr>
<td>CaMII-α</td>
<td>YSEADASHCIQQILEAVLHCHQM-GVTHRDLKPNELL1DASKL 18 43%</td>
</tr>
<tr>
<td>CaMII-β</td>
<td>YSEADASHCIQQILEAVLHCHQM-GVTHRDLKPNELL1DASKL 18 43%</td>
</tr>
<tr>
<td>PhK-γ</td>
<td>YSEADASHCIQQILEAVLHCHQM-GVTHRDLKPNELL1DASKL 18 43%</td>
</tr>
<tr>
<td>SNF1</td>
<td>MSEQEEARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>Nim1</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>KIN1</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>CDC28</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>CDC2</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>CDC2Hs</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>Raf</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>A-Raf</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>PKS</td>
<td>LSEERAARFFQIQIISAVEYCHR-RIVHRDLKPNELL1DEHL 18 43%</td>
</tr>
<tr>
<td>DCKII</td>
<td>LTDYERFLLKALDYCHSM-GIMMRDVKPHNMIDHEN 13 31%</td>
</tr>
<tr>
<td>CDC7</td>
<td>LTDYERFLLKALDYCHSM-GIMMRDVKPHNMIDHEN 13 31%</td>
</tr>
<tr>
<td>STE7</td>
<td>LTDYERFLLKALDYCHSM-GIMMRDVKPHNMIDHEN 13 31%</td>
</tr>
<tr>
<td>STK61</td>
<td>LTDYERFLLKALDYCHSM-GIMMRDVKPHNMIDHEN 13 31%</td>
</tr>
<tr>
<td>(continued...)</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 5.2(d)**

<table>
<thead>
<tr>
<th>Subdomain VII</th>
<th>Identities</th>
</tr>
</thead>
<tbody>
<tr>
<td>cAPK-α</td>
<td>YIQYTDFGFAKRVKGRT-- 5 26%</td>
</tr>
<tr>
<td>cAPK-β</td>
<td>YIQYTDFGFAKRVKGRT-- 5 26%</td>
</tr>
<tr>
<td>SRA3</td>
<td>HIKITDFGFAKYPVPDVI-- 7 37%</td>
</tr>
<tr>
<td>PKC-α</td>
<td>HIKIADFGMCKEHMDGVTT 7 37%</td>
</tr>
<tr>
<td>PKC-β</td>
<td>HIKIADFGMCKENMDGVTT 7 37%</td>
</tr>
<tr>
<td>PKC-γ</td>
<td>HIKITDFGCMCKENVFPGST 6 32%</td>
</tr>
<tr>
<td>CaMII-α</td>
<td>AVKLADFGALIEVEEQQQA 6 32%</td>
</tr>
<tr>
<td>CaMII-β</td>
<td>AVKLADFGALIEVQGQQQA 6 32%</td>
</tr>
<tr>
<td>PhK-γ</td>
<td>NIKLTDHGSQPLDEKEL 4 21%</td>
</tr>
<tr>
<td>SNF1</td>
<td>NVKIADFGSNIMTDGNFL 7 37%</td>
</tr>
<tr>
<td>Nim1</td>
<td>QIKIADFMATVEPNDSCL 7 37%</td>
</tr>
<tr>
<td>KIN1</td>
<td>EIKIADFGSNNYDSKQL 4 21%</td>
</tr>
<tr>
<td>CDC28</td>
<td>NLKLADFGALAFGVRPLRA 6 32%</td>
</tr>
<tr>
<td>CDC2</td>
<td>NLKLADFGALSFGVPLRN 6 32%</td>
</tr>
<tr>
<td>CDC2Hs</td>
<td>TIKLADFGARAGPIPRV 6 32%</td>
</tr>
<tr>
<td>Raf</td>
<td>TVKIGDFGLATKSRWSGS 6 32%</td>
</tr>
<tr>
<td>A-Raf</td>
<td>TVKIGDFGLATKTRWSGA 8 42%</td>
</tr>
<tr>
<td>PKS</td>
<td>TVKIGDFGLATKTRWSGA 8 42%</td>
</tr>
<tr>
<td>DCKII</td>
<td>KLRLEIYGGLAEFYHPQEQY 3 16%</td>
</tr>
<tr>
<td>CDC7</td>
<td>RGVLDVFGLAQAQMDYKSM 4 21%</td>
</tr>
<tr>
<td>STE7</td>
<td>QIKLADFGSKKLINSIA- 4 21%</td>
</tr>
<tr>
<td>STK61</td>
<td>HTLIAFDGSAKVMTAHERA 19 100%</td>
</tr>
</tbody>
</table>

(continued...)

182
<table>
<thead>
<tr>
<th>Subdomain VIII</th>
<th>Identities</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>cAPK-α</td>
<td>WTLCGTPYLAPELV</td>
<td>32%</td>
</tr>
<tr>
<td>cAPK-β</td>
<td>WTLCGTPYLAPELV</td>
<td>33%</td>
</tr>
<tr>
<td>SRA3</td>
<td>YTLCGTPYIAPELV</td>
<td>40%</td>
</tr>
<tr>
<td>PKC-α</td>
<td>RTFCGTPYIAPELV</td>
<td>35%</td>
</tr>
<tr>
<td>PKC-β</td>
<td>KTTCGTPYIAPELV</td>
<td>31%</td>
</tr>
<tr>
<td>PKC-γ</td>
<td>RTFCGTPYIAPELV</td>
<td>35%</td>
</tr>
<tr>
<td>CaMII-α</td>
<td>FGFAGTPGLSPEVL</td>
<td>35%</td>
</tr>
<tr>
<td>CaMII-β</td>
<td>FGFAGTPGLSPEVL</td>
<td>36%</td>
</tr>
<tr>
<td>PhK-γ</td>
<td>REVCGTPSYLAPELV</td>
<td>32%</td>
</tr>
<tr>
<td>SNF1</td>
<td>KTSCGSPNAAPEVL</td>
<td>34%</td>
</tr>
<tr>
<td>Nima</td>
<td>ENYCGSLHPEIV</td>
<td>33%</td>
</tr>
<tr>
<td>KIN1</td>
<td>HTFCGSLYFAPELV</td>
<td>28%</td>
</tr>
<tr>
<td>CDC28</td>
<td>THEIVTLWYRAPEVL</td>
<td>28%</td>
</tr>
<tr>
<td>CDC2</td>
<td>THEIVTLWYRAPEVL</td>
<td>31%</td>
</tr>
<tr>
<td>CDC2Hs</td>
<td>THEMIVLWYRSPEVL</td>
<td>33%</td>
</tr>
<tr>
<td>Raf</td>
<td>EQPTGSVLWMAPEVL</td>
<td>24%</td>
</tr>
<tr>
<td>A-Raf</td>
<td>EQPSGSLWMAAEVI</td>
<td>28%</td>
</tr>
<tr>
<td>PKS</td>
<td>EQPSGVLWMAAEVI</td>
<td>28%</td>
</tr>
<tr>
<td>DCKII</td>
<td>NVRVASRYFKGPELL</td>
<td>23%</td>
</tr>
<tr>
<td>CDC7</td>
<td>ANRAGTRGFAPEVL</td>
<td>29%</td>
</tr>
<tr>
<td>STE7</td>
<td>DTFTVGTSTYMSPEVI</td>
<td>27%</td>
</tr>
<tr>
<td>STK61</td>
<td>ASFVGTAQVSRPEVL</td>
<td>100%</td>
</tr>
</tbody>
</table>
The invariant or nearly invariant residues indicated in figure 5.1 are highly conserved throughout all of the kinases shown in figure 5.2. These residues are likely to be essential for the catalytic domain to perform its kinase function. It is clear from the comparison in figure 5.2 that residues other than these invariant ones are much less well conserved between members of different families of kinases. Thus, it is reasonable to assume that significant homology in residues other than the invariant ones is an indicator of kinase function. The levels of homology which the kinases shown in figure 5.2 exhibit to STK61 varies from subdomain to subdomain. For example, in subdomain I, CDC7 has 46% identity to STK61. This is a very high level, relative to the other proteins examined. However, CDC7 shows average to low levels of homology throughout the rest of the subdomains. It may be that subdomain I is required for some shared regulatory mechanism in both CDC7 and STK61. Generally, the highest levels of homology to STK61 are seen by members of the cAMP-dependent and Ca\(^{2+}\)/calmodulin-dependent protein kinase families. Subdomain VI exhibits the highest number of conserved residues of all the subdomains (figure 5.2(c)). In this subdomain, the Ca\(^{2+}\)/calmodulin-dependent phosphorylase kinase \(\gamma\) subunit (PhK-\(\gamma\)) shows 50% identity with STK61, and the cAMP-dependent Ras suppressor kinase (SRA3) shows 45% identity. The lowest level of homology shown with this subdomain is to the STE7 kinase, at 24%.

Between the N-terminal end of STK61 and the start of the kinase catalytic domain there are several runs of the acidic amino acid glutamine (Q). These Q-repeats have been termed OPA repeats and are found in a number of developmentally important proteins, including kinases. However, no OPA repeat-containing kinases have been identified from Drosophila or indeed from any dipteran insect. Initially discovered in the Notch gene (Wharton et al., 1985; Grabowski et al., 1991), OPA-encoding repeats have been seen in a number of different genes which have functions in development, including twist, hunchback and dorsal. Although little is known about what function OPA repeats may have, the proteins in which they are found are developmentally and/or spatially regulated, being involved in tissue-specific functions. Thus, we might expect stk61 to be expressed in a tissue-specific manner.
However, *in situ* RNA hybridisations performed using probes synthesised from cDNAa to whole mount carcass tissue and imaginal discs show expression in a global non tissue-specific manner (data not shown). This is probably due to the presence of the non tissue/sex-specific common transcript and it will only be when antibody *in situ* are done that the true expression pattern of stk6l is revealed.

Since no significant homology is seen to between STK61 and other known protein kinases, other than short stretches of globally-conserved residues within the catalytic domain, together with the fact that no other *Drosophila* OPA repeat-containing kinases are known, it is reasonable to suppose that the stk6l gene encodes a member of a novel family of OPA repeat-containing S/T kinases.

The predicted STK61 protein is of a fairly neutral pH, having a pKa of 6.59 (figure 5.3). The secondary structure of STK61 was predicted using Chou/Fasman and Kyle/Doolittle algorithms as shown in figures 5.4 and 5.5. This analysis indicates that the STK61 protein is cytosolic since no hydrophobic transmembrane repeats are seen. This is unsurprising, since the vast majority of S/T kinases are cytosolic. In figure 5.5, the hydrophilic OPA repeats are clearly seen covering amino acids 119-151. The catalytic domain begins 8 amino acids down from this hydrophilic stretch in the C-terminal direction and extends over amino acids 160-490. The catalytic domain is preceded by a short break of \( \alpha \)-helix, after which subdomains I-VII consist predominantly of \( \beta \)-sheet. There is a large spacer region between subdomains VII and VIII covering amino-acids 318-383. In figure 5.5 it is clear that this spacer region is extremely hydrophilic and contains a large quantity of \( \alpha \)-helix. Subdomain VIII begins after this hydrophilic \( \alpha \)-helix spacer and the remainder of the catalytic domain consists of 4 stretches of \( \beta \)-sheet separated by short \( \alpha \)-helical regions. After the end of the catalytic domain, the protein C-terminus is largely \( \alpha \)-helix and exhibits no striking areas of high hydrophobicity or hydrophilicity. Similarly, the N-terminus of the protein, prior to the OPA repeats, is mainly \( \alpha \)-helical. It appears, therefore, that the STK61 protein is divided up into areas by its secondary structure which seem to demarcate the catalytic areas of the protein.
Figure 5.3
Isoelectric plot of the STK61 predicted protein sequence.
Figure 5.4
Computer analysis of STK61 predicted protein sequence. Plots indicate hydrophobic and hydrophilic regions, as well as areas of \(\alpha\)-helix and \(\beta\)-sheet. A pictorial representation of this data is shown in figure 5.5.
Figure 5.5
Pictorial representation of data presented in figure 5.4. Hydrophobic and hydrophilic regions are indicated. Areas which are predicted to consist of α-helix or β-sheet are also shown.
The region containing the kinase catalytic domain and OPA repeats is bounded by α-helical C and N-termini. The catalytic domain itself consists mainly of β-sheet, with a large hydrophilic α-helix spacer between subdomains VII and VIII. It seems likely that the catalytic domain, hydrophilic OPA repeats and hydrophilic spacer region present on the surface of the protein and fold together to form the active site, with the less hydrophilic α-helical termini forming the internal structure of the protein.

5.2.2 BACTERIAL EXPRESSION STUDIES.

5.2.2.1 Codon preference analysis of the stk61 open reading frame.

The above analysis is based on the predicted protein sequence encoded by cDNAs a and 11. However, it is not known whether this protein is actually produced in vivo. Figure 5.6 (overleaf) shows a codon preference analysis of cDNA11. This graph indicates the probability that the codons used in the cDNA11 (and therefore cDNAa) ORF are typical of codons generally used in Drosophila genes. It is clear from this that the ORF of cDNAa and cDNA11 contains typical Drosophila codons and so is highly likely to be translated in vivo.
Figure 5.6
Codon preference computer analysis of cDNA11. The lower panel shows the correct reading frame for translation of the STK61 protein. The long STK61-encoding open reading frame is shown as an open box at the bottom of this panel. The lower trace in each panel indicates the frequency with which a particular codon is used in *Drosophila*, with values above the lower line indicating typical codons. The lower panel shows that the cDNA11 open reading frame utilizes typical *Drosophila* codons. A similar result is seen for G/C bias (upper line and upper trace in each panel).
5.2.2.2 Construction of pGEX<sub>STK61</sub>.

To determine whether the cloned cDNAs are in fact capable of producing the predicted protein, the ORF from cDNA<sub>a</sub> was cloned into a bacterial expression vector. Figure 5.7 shows the cloning strategy that was used and the composition of the final construct, pGEX<sub>STK61</sub>. The pGEX series of GST-fusion vectors were used for this analysis due to the ease of detection and purification of GST-fusion proteins via anti-GST immunofluorescence and GST-affinity chromatography.

**Figure 5.7**
Construction of pGEX<sub>STK61</sub> expression construct. cDNA<sub>a</sub> open reading frame is shaded. Relevant restriction sites are shown. GST=Glutathione-S-transferase, pB/S<sub>SK</sub>=pBluescript SK<sup>+</sup> plasmid vector, Ptac=IPTG-inducible promoter. cDNA<sub>a</sub><sub>SK</sub> was digested with EcoRI and partially digested with EcoRV, to produce the fragment shown, which was ligated into pGEX<sup>IX</sup> to create pGEX<sub>STK61</sub>. The cDNA<sub>a</sub> ORF in pGEX<sup>STK61</sup> lacks the codons encoding the 19 amino acids at the extreme amino-terminus of the predicted STK61 protein. The resulting pGEX<sup>STK61</sup> construct encodes STK61 protein with GST sequence fused to its amino terminus.
5.2.2.3 Bacterial expression of STK61/GST fusion protein.

The results of initial expression analysis are shown in figure 5.8. The cDNA ORF which was cloned into the expression vector encodes a predicted STK61 protein of 83.4kDa. Therefore, when fused to the 26kDa GST protein, the full length protein encoded by pGEX\textsuperscript{STK61} has a predicted molecular weight of 109.4 kDa.

![SDS PAGE of protein extracts from bacterial cells transformed with pGEX\textsuperscript{STK61} expression construct. MWM=Molecular weight markers. Samples were taken from the expression culture prior to IPTG-induction (T=0) and at hourly intervals after induction (T=1, 2, 3, 4, 5). The optical density of each sample was determined and used to correct for the increasing density of the culture through time. Thus, each lane shown above represents protein extracted from an approximately equal number of cells. The band representing induced STK61/GST fusion protein is indicated. This band appears larger than expected (109.4 kDa). This may be due, in part, to differences in salt concentrations between the marker and bacterial extract lanes. However, some post-translational modification (such as glycosylation) may also have occurred, increasing the observed molecular weight of the fusion protein.

One hour after IPTG-induction an induced band of >100kDa is clearly visible, indicating that the full length STK61 protein is being synthesised. Smaller bands are also induced, indicating that some proteolytic degradation of the expressed protein is occurring.
5.2.2.4 Western blot analysis of STK61/GST fusion protein.

Anti-GST antibody was used to perform Western analysis to confirm that the induced band was being produced from the pGEX\textsuperscript{STK61} construct (figure 5.9). The 26kDa GST protein produced from pGEX\textsuperscript{3X} is well expressed and is detected very specifically by anti-GST as shown in figure 5.9.

![Figure 5.9](image)

(a) SDS PAGE of protein extracts from pGEX\textsuperscript{3X} and pGEX\textsuperscript{STK61} transformed cells before (T=0) and after (T=2) IPTG-induction. Both pGEX\textsuperscript{3X}-derived and pGEX\textsuperscript{STK61}-derived proteins can be seen in the induced extracts. (b) Anti-GST Western blot of protein gel shown in (a). pGEX\textsuperscript{3X}-derived GST protein is only detected in the extract from induced cells. Similarly, no STK61/GST fusion protein is detected in un-induced pGEX\textsuperscript{STK61}-transformed cells. Extracts from pGEX\textsuperscript{STK61}-transformed induced cells clearly contain STK61/GST fusion protein. Only a very small amount of this protein is full-length. Nine GST-containing STK61/GST proteolytic degradation products are detected. MWM=Molecular Weight Markers.

Expression of the fusion protein from pGEX\textsuperscript{STK61} is at quite a low level, however, although it is still detected by anti-GST. Nine smaller bands are also detected by anti-GST, indicating that the STK61/GST fusion is undergoing significant levels of proteolysis. It is most likely that this proteolysis is occurring while the protein is being expressed in culture, as cells were frozen in liquid Nitrogen immediately after harvesting and anti-proteolytic chemicals were used in all subsequent steps.
Different host strains were also used but showed similar levels of proteolysis (data not shown). Similarly, attempts to reduce proteolysis by culturing cells at low temperature (25°C-30°C) proved ineffective.

Levels of pGEXSTK61 expression are much lower than levels of expression of the 26kDa GST protein from pGEX3X. This could partly be explained by the much larger size of the pGEXSTK61-encoded protein but could also be due to deleterious effects of the STK61 kinase function on the bacterial cells. To maximise levels of fusion protein expression, freshly-transformed cells were used for each expression experiment. Cultures were grown from single-cell inoculations to Log-phase when they were induced with IPTG. This strategy was used in preference to inoculation from overnight cultures to prevent the cultures ever reaching turbidity, since the competition this would cause between cells may increase the incidence of pGEXSTK61 plasmid loss. To prevent uninduced expression from the pGEXSTK61 construct, BL21 DE3 LysS cells were also used as hosts. This strain contains the LyS plasmid which prevents expression from the vector promoter in the absence of IPTG induction (Materials and Methods). However, even lower levels of pGEXSTK61 expression were seen with these cells and so XL-1 cells were used in preference.

5.2.2.5 Kinase assays of crude protein extracts containing STK61/GST protein.

In an attempt to ascertain whether the protein expressed from pGEXSTK61 was functionally active, preliminary kinase assays were carried out. Essentially, this was done by preparing crude bacterial extracts from IPTG-induced or IPTG-uninduced cells containing either pGEXSTK61, or pGEX3X vector. γ-radiolabelled ATP was then added to the extracts which were incubated for 1 hour at 25°C and then analysed by SDS-PAGE and autoradiography (Materials & Methods). A typical result is shown in figure 5.10. Although it is clear that general kinase activity is occurring in these extracts, no increase in kinase activity is seen in extracts from cells which contain pGEXSTK61, either induced or uninduced. Figure 5.6 shows that the ORF which was cloned into pGEX3X does not encode the first 19 amino acids of the predicted STK61...
protein. However, the predicted secondary structure of the protein, as discussed above, suggests that these residues are unlikely to be critical to the protein's function. Similarly, the presence of the GST-fusion is unlikely to be interfering with the STK61 protein function, since the GST protein has generally not been seen to interfere with protein function in fusion proteins.

![Diagram](image)

**Figure 5.10**

Typical result from preliminary kinase assay experiments. (a) SDS PAGE of protein extracts from pGEX<sup>STK61</sup> and pGEX<sup>3X</sup> transformed cells, either un-induced (T=0) or induced (T=2). (b) An aliquot of the protein extracts shown in (a) was incubated with γ-radiolabelled ATP and analysed via SDS PAGE and autoradiography. Bands representing phosphorylated proteins are indicated. No increase in kinase activity is seen in the pGEX<sup>STK61</sup>-transformed cells following IPTG induction. MWM=Molecular Weight Markers.
5.2.2.6 Solubility of STK61/GST protein in bacterial cells.

Figure 5.11 shows the result of a pGEX$^{STK61}$ expression experiment, where soluble and insoluble protein has been separated.

![Figure 5.11](image)

**Figure 5.11**
SDS PAGE of protein extracts from pGEX$^{STK61}$-transformed cells, either un-induced (T=0) or induced (T=2) with IPTG. Soluble and insoluble protein fractions were separated and loaded separately. Most STK61/GST fusion protein is insoluble. MWM=Molecular Weight Markers.

It is clear from this that most of the pGEX$^{STK61}$-expressed protein is present in the insoluble fraction of the bacterial extracts and so is likely to be sequestered in inclusion bodies. Thus, the lack of kinase activity seen from induced pGEX$^{STK61}$ cell extracts does not necessarily indicate that the produced protein is non-functional.
5.2.2.3 Purification of STK6I/GST and GST proteins.

Since the STK6I/GST fusion protein is largely insoluble in bacterial cells, the use of crude bacterial extracts for functional assays is clearly not adequate. A purification was attempted using glutathione-agarose affinity chromatography to concentrate any fusion protein which may still be in the soluble fraction of the pGEX\textsuperscript{STK6I}-containing induced cells (figure 5.12).

![Figure 5.12](image)

**Figure 5.12**
SDS PAGE of various samples taken throughout a GST affinity chromatography protein purification experiment. Glutathione-agarose matrix (GAM) is added to cell lysates. GST protein associates with the glutathione-agarose and is eluted in the presence of reduced glutathione. Sample 1; pGEX\textsuperscript{SS}-transformed induced cell lysate, soluble fraction, after binding and removal of GAM. Sample 2; pGEX\textsuperscript{SS}-transformed induced cell lysate, soluble fraction. Sample 3; pGEX\textsuperscript{SS}-transformed induced cell lysate, after addition of lysozyme and detergent. Sample 4; pGEX\textsuperscript{SS}-transformed induced cell lysate, crude extract. Sample 5; pGEX\textsuperscript{SS}-transformed un-induced cell lysate, crude extract. Sample 6; pGEX\textsuperscript{STK6I}-transformed induced cell lysate, soluble fraction, after binding and removal of GAM. Sample 7; pGEX\textsuperscript{STK6I}-transformed induced cell lysate, soluble fraction. Sample 8; pGEX\textsuperscript{STK6I}-transformed induced cell lysate, after addition of lysozyme and detergent. Sample 9; pGEX\textsuperscript{STK6I}-transformed induced cell lysate, crude extract. Sample 10; pGEX\textsuperscript{STK6I}-transformed un-induced cell lysate, crude extract. Sample 11; Eluted protein from pGEX\textsuperscript{SS}-transformed cells (0.2% of total eluate loaded). Sample 12; Eluted protein from pGEX\textsuperscript{SS}-transformed cells (0.02% of total eluate loaded). Sample 13; Eluted protein from pGEX\textsuperscript{STK6I}-transformed cells (15% of total eluate loaded). Some bands are visible in lane 13 which may represent purified STK61/GST protein, but such a large proportion of the total eluate was loaded that these bands may represent carry-over of endogenous bacterial proteins. MWM=Molecular Weight Markers.
Very efficient purification of GST protein from the pGEX\textsuperscript{3X}-transformed induced cell control is seen (figure 5.12; lanes 11 & 12). However, very poor recovery is seen from pGEX\textsuperscript{STK61}-containing induced cells (figure 5.12; lane 13). This is especially striking considering that around 800 times more glutathione-agarose eluate was loaded in lane 13 than in lane 12. Clearly, virtually no STK61/GST fusion protein remains in the soluble fraction of the pGEX\textsuperscript{STK61}-containing induced cells. A lengthy inclusion body purification protocol was not attempted, since this involves denaturation/renaturation of the purified protein and could be deleterious to the protein's function.

In summary, cDNAs a and 11 encode an OPA repeat-containing serine/threonine protein kinase. This protein is produced from the ORF \textit{in vivo}. This protein may be regulated in a tissue-specific or developmentally-regulated manner.
5.3 DISCUSSION

Over recent years a great deal of work has been done on protein kinases which has revealed their importance in developmentally significant processes such as intercellular signal transduction and cell cycle regulation. To further understand what roles a sex-specific kinase might play, it is necessary to review the known functions of kinases in *Drosophila*.

Protein kinases can be divided into two groups depending on which amino acids they catalyse the transfer of phosphate to. These are the tyrosine kinases and the serine/threonine kinases. Tyrosine kinases can be further subdivided into those proteins which are membrane localised receptors (receptor tyrosine kinases, or RTK's) and those which are free in the cytoplasm. The receptor tyrosine kinases can be further subdivided according to the type of intracellular and extracellular domains they possess. The intracellular kinase domain may or may not be split by a kinase insert (KI) region. Extracellularly, the protein may exhibit a number of ligand binding domains such as cysteine repeats, immunoglobulin C2 domains, acidic domains, fibronectin type III repeats and the *sevenless*-like YWTD amino acid repeats. Serine/threonine kinases on the other hand tend to be purely cytoplasmic (reviewed in Pawson & Bernstein, 1990; Seifried *et al.*, 1990).

5.3.1 PROTEIN KINASES IN SIGNAL TRANSDUCTION

Perhaps the most widespread function of protein kinases is the transduction of intercellular signals into intracellular changes in gene expression. In all cases, this involves an extracellular ligand making itself known to some component of the plasma membrane which then causes intracellular changes to take effect. This can occur in several ways, all of which involve the action of intracellular protein kinases (reviewed in Watson *et al.*, 1987). Some ligands act via the membrane-bound enzymes phospholipase C and adenylate cyclase. When active, phospholipase C catalyses the breakdown of the plasma membrane phospholipid phosphatidyl inositol.
diphosphate (PIP$_2$) into the second messengers inositol-3-phosphate (IP$_3$) and diacylglycerol (DAG). Adenylate cyclase catalyses the formation of the second messenger cyclicAMP (cAMP) from ATP. Increases in the intracellular concentration of these second messengers activates the cAMP-dependent, Ca$^{2+}$/calmodulin-dependent and phospholipid/calmodulin-dependent protein kinases which then go on to modulate gene expression. How these kinases exert their function is, as yet, poorly understood. The activity of the enzymes phospholipase C and adenylate cyclase is regulated by the binding of extracellular peptide growth hormones to receptor tyrosine kinases (RTK's). It has been known for some time that tripartite GTP-binding protein (G-protein) complexes (consisting of α, β and γ subunits) are involved in the transduction of extracellular hormone signals into changes in intracellular concentration of second messengers (Gilman, 1984). In the absence of hormone stimulation, GDP is bound to the G-protein complex which is inactive. When the G-protein complex encounters a hormone-bound RTK, the GDP dissociates and is replaced by GTP which activates the complex. The GTP-bound α-subunit then dissociates from the complex and either stimulates or represses the second messenger-producing enzymes. Figure 5.13 shows a summary of these transduction mechanisms.

**Figure 5.13**
Summary of the major pathways of extracellular signal transduction. See text for details.
One particular α-subunit G-protein encoded by the ras oncogene has recently provided a great deal of information concerning the role of protein kinases in signal transduction. This gene was originally found in a mutated form, causing malignant transformation of mammalian culture cells (Satoh et al., 1992). Like other α-subunit G-proteins, active Ras binds GTP which it gradually hydrolysates to GDP via its own GTPase activity. Furthermore, Ras deactivator proteins, GAP proteins (GTPase Stimulating Proteins), exist which can associate with Ras and stimulate its GTPase activity. Injection of anti-Ras antibody into fibroblast cells suppresses those cells' normal responses to peptide hormones such as PDGF and EGF (Mulcahy et al., 1985), indicating that Ras has a role in transducing extracellular signals via RTK's.

When extracellular ligands bind to RTK's on the surface of the cell, this brings about dimerisation of the receptors which become autophosphorylated on certain tyrosine residues of their intracellular domains (Shilo, 1992). Recently, the mechanisms by which RTK's bring about changes in Ras activity has been further elucidated by studies on a conserved 100 amino acid domain which appears to mediate specific binding to phosphorylated tyrosine residues present in the intracellular domains of RTK's. This domain has been termed the src homology-2 (SH2) domain and is found in many signaling proteins including Src, Abl, Phopholipase C and Phosphatidyl inositol (3') kinase (Pawson & Gish, 1992). It has been shown that SH2 domains directly recognise and bind to the phosphotyrosine residues present in the intracellular domain of activated RTK's and that the affinity of these interactions is dependent upon the amino acids immediately surrounding the phosphorylated group. SH2 domain-containing proteins also often contain SH3 domains which have been implicated in binding to proline-rich target sequences in other proteins (Cicchetti et al., 1992). For these reasons, proteins which contain both SH2 and SH3 domains have been postulated to act as adaptor proteins, linking activated RTK's to other components of the transduction machinery (Pawson & Gish, 1992).
Direct confirmation of the function of SH2/SH3-containing proteins as RTK adaptors has come from the cloning and analysis of the mouse grb2 gene (Suen et al., 1993). The Grb2 protein contains no catalytic domains, consisting of only one SH2 and two SH3 domains. Treatment of culture cells with peptide hormones results in a SH2-dependent association of Grb2 with the cognate receptor tyrosine kinase. The SH3 domains of Grb2 were shown to bind to a different subset of proteins from those bound by the SH2 domain, supporting a function for Grb2 as an adaptor between activated RTK's and other signal transduction mediators.

In a similar way to Ras, injection of anti-Grb2 antibody into mammalian culture cells suppresses these cells responses to PDGF and EGF (Matuoka et al., 1993). This strongly suggests that the Grb2 SH2/SH3 adaptor acts in some way to link activated RTK's to the activation of Ras protein.

The missing link between the Grb2 adaptor and Ras has been revealed by studies on eye development in Drosophila. In Drosophila, the compound eye is made up of around 800 ommatidia, each consisting of 8 photoreceptor cells (R1-R8) surrounded by 4 non-neural cone cells and 8 accessory cells. The R7 photoreceptor is particularly distinctive, and has been the object of much genetic study (reviewed in Rubin, 1991). Mosaic analysis has shown that while a receptor tyrosine kinase encoded by the sevenless (sev) gene is required in the R7 cell for correct development, the putative transmembrane protein product of the gene bride of sevenless (boss) is required in cell R8. Antibody in situ evidence shows that the Boss protein is taken into the R7 cell and that this is dependent upon Sev function. This suggests that the Boss protein is in fact the ligand which activates Sev. Genes downstream of the Sev RTK have been identified via a sev hypermorphic constitutive allele. These genes include Son of sevenless (Sos; Rogge et al., 1991), gap1 (Gaul et al., 1992) and ras1, the Drosophila homolog of mammalian ras (Simon et al., 1991). The cloning and characterisation of another gene important in R7 specification, downstream of receptor kinases (drk), has enabled the construction of a model for ras activation (Olivier et al., 1993). The drk gene is highly
homologous to mammalian grb2, encoding a protein consisting of one SH2 domain and two SH3 domains. As expected, Drk protein is seen to associate with the plasma membrane in a SH2-dependent manner. In vitro, Drk protein binds to an area of the Sos protein known to contain proline rich repeats. Since the Sos gene encodes a putative guanine nucleotide-releasing factor, this suggests a model for ras activation whereby SH2/SH3 adaptor proteins create a link between activated RTK's and guanine nucleotide-releasing factors which act to release GDP from inactive Ras protein, allowing it to bind GTP and become active. This model is summarised in figure 5.14.

Figure 5.14
Diagram showing model for Ras activation. Extracellular ligands bind to receptor tyrosine kinases, which dimerise and autophosphorylate in their intracellular domains. The SH2 domain of the Drk/Grb2 adaptor protein binds to the phosphotyrosine in the intercellular domain of the receptors. The Guanine nucleotide releasing factor, Sos, contains proline-rich motifs which bind to the SH2 domains of Drk. Sos catalyses the removal of GDP from Ras, enabling Ras to bind GTP and become active. KD=Kinase Domain.

In an attempt to identify the downstream targets of ras, active GTP-bound Ras protein was added to mammalian cell culture lysates and then immunoprecipitated using anti-Ras antibody (Koide et al., 1993). The S/T kinase Raf was co-precipitated with Ras, indicating that Raf is a downstream target of Ras with which it physically
interacts. Indeed, direct protein-protein interaction between mammalian Ras and Raf has been demonstrated using the yeast dihybrid screening system (Vojtek et al., 1994). It is clear that Ras activation leads to the subsequent activation of a number of downstream protein kinases, including Raf, which are themselves regulated by phosphorylation (Blenis, 1993). Of these proteins, the mitogen-activated protein kinases (MAP kinases) and the cell-cycle regulated S6 protein kinase (RSK kinase) are particularly interesting. The RSK kinase is a cell cycle-regulated protein which brings about phosphorylation of ribosomal protein, presumably playing some role in cell cycle control. RSK kinase is activated by phosphorylation under control of the S/T-specific MAP kinase. Activation of MAP kinases has been shown to result in the phosphorylation of the Myc and Jun transcription factors. MAP kinase is also activated by phosphorylation, this time under control of the S/T/Y-specific kinase MEK. MEK kinase is also activated by phosphorylation. In vitro, Raf kinase is capable of reactivating phosphatase-treated MEK kinase (Howe et al., 1992). Raf kinase also appears to be upstream of MAP kinase (Howe et al., 1992; Samuels et al., 1993). These results suggest a pathway whereby the G-protein Ras activates the S/T kinase Raf, which goes on to initiate a cascade of kinase activity which terminates in phosphorylation of ribosomal proteins and nuclear factors such as transcription factors.

Another level of regulation of Ras-mediated transduction involves the phosphatase, protein phosphatase 2A (PP2A). In Drosophila, use of transgenic flies which constitutively express Ras or Raf has revealed that PP2A has a role in regulating the activity of these proteins (Wassarman et al., 1996). Constitutive expression of ras results in a number of ommatidial cone cells developing as R7 photoreceptors. If only one functional copy of PP2A is present, an increase in these transformations is seen, indicating that PP2A represses members of the Ras-induced pathway. However, a reduction in transformed cells is seen in flies constitutively expressing raf, when only one active copy of PP2A is present, suggesting that PP2A activates members of the Raf-induced pathway. Thus, it seems clear that PP2A acts on different members of the Sev signaling pathway, as both an activator and a
deactivator. The substrates of PP2A in the Sev pathway are unclear as yet. It may be that multiple forms of PP2A are involved and that some PP2A substrates are yet to be discovered. In a screen for dominant suppressors/enhancers of eye phenotype caused by a hypermorphic ras allele, three genes were identified which are involved in the Sev signaling pathway (Wassarman et al., 1996). These are kinase suppressor of ras (ksr), the product of which acts either upstream of, or in parallel to Raf; phyllopod (phyl), which encodes a nuclear protein implicated in neural differentiation and yan, a transcription factor which represses photoreceptor development.

One particularly striking feature of the RTK→Ras→Raf→MEK→MAP kinase cascade is the evolutionary conservation of this pathway. Homologs are present in signal transduction pathways from yeast, C. elegans, Drosophila and mammalian cells, as shown in table 5.1 (see Feig, 1993; Egan & Weinberg, 1993; Crews & Erikson, 1993; Pelech, 1993 for reviews).

<table>
<thead>
<tr>
<th>ORGANISM</th>
<th>Drosophila</th>
<th>Drosophila</th>
<th>Mammals</th>
<th>C. elegans</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTK</td>
<td>Sev</td>
<td>Torso</td>
<td>PDGF</td>
<td>Let-23</td>
</tr>
<tr>
<td>ADAPTOR</td>
<td>Drk</td>
<td>Grb2</td>
<td>SH-PTP2</td>
<td>Sem-5</td>
</tr>
<tr>
<td>PTP</td>
<td></td>
<td>Csw</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRF</td>
<td>Sos</td>
<td>Sos</td>
<td>Sos-1, 2</td>
<td></td>
</tr>
<tr>
<td>GAP</td>
<td>GAP-1</td>
<td>Sos</td>
<td>Sos</td>
<td>Ras GAP</td>
</tr>
<tr>
<td>G-PROTEIN</td>
<td>Ras1</td>
<td>Ras1</td>
<td>p21^rasl</td>
<td>Let-60</td>
</tr>
<tr>
<td>S/T KINASE</td>
<td>D-Raf</td>
<td>D-Raf</td>
<td>Raf-1</td>
<td>Lin-45</td>
</tr>
<tr>
<td>Y/T KINASE</td>
<td></td>
<td>D-Sor1</td>
<td>MEK</td>
<td></td>
</tr>
<tr>
<td>S/T KINASE</td>
<td>Rolled</td>
<td></td>
<td>MAP</td>
<td></td>
</tr>
<tr>
<td>NUCLEAR EFFECTORS?</td>
<td>Sina</td>
<td>Taillness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Huckebein</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1
Conservation of proteins involved in Ras-mediated signaling. PTP=Protein Tyrosine Phosphatase, GRF=Guanine-nucleotide Releasing Factor, GAP=GTPase Activating Protein.
In *Drosophila*, the proteins of the MAP kinase cascade are conserved in at least two systems; the development of the R7 photoreceptor and the determination of the terminal poles of the developing embryo. As we have already discussed, determination of the R7 photoreceptor begins with a signal from Boss to the RTK Sev. The SH2/SH3 adaptor protein Drk acts to couple the activated Sev RTK to the guanine nucleotide releasing factor Sos, which brings about activation of Ras. This is antagonised by the GTPase-stimulating protein GAP-1. Ras activation leads to phosphorylation (by some unknown kinase) of the Raf S/T kinase which, in turn, leads to phosphorylation of the *Drosophila* MEK kinase homolog, Rolled. Mutation of the gene *seven in absentia* (*sina*) results in a similar phenotype to *sev* and *boss* alleles (Carthew & Rubin, 1990). This gene encodes a nuclear putative transcription factor and may be one of the terminal targets of the Sev signaling pathway. The *seven-up* (*svp*) gene encodes a protein with homology to the steroid receptor family, is expressed in photoreceptors R1, R3, R4 & R6, and misexpression of *svp* causes R7 cells to develop as the photoreceptors which usually express *svp* (Hiromi et al., 1993). Thus, Svp may act to repress the effects of the Sev pathway in non-R7 photoreceptors.

The terminal effect genes are involved in determining the positional identity of the extreme ends of the developing embryo and have been discussed in detail in Chapter 1. The RTK Torso (Tor) is thought to be activated at the ends of the embryo be the extracellular ligand Torsolike (Tsl). Downstream of the Tor RTK, the product of the gene *l(1)pole-hole* (*l(1)ph*) regulates the expression domains of the gap genes *tailless* and *huckebein* which are responsible for the positional identity of the embryo's terminal structures (Johnston & Nüsslein-Volhard, 1992). The gene *l(1)ph* in fact encodes the *Drosophila* homolog of *raf* which is clearly important in the determination of embryonic terminal structures, as well as in specification of the R7 photoreceptor (Melnick et al., 1993). Indeed, the Tor RTK signaling pathway involves *Drosophila* homologs of *ras* (*ras1*), *PP2A* (*corkscrew*), MAP kinase (*rolled*) and MEK kinase (*D-sor1*), as well as the Sos SH2/SH3 adaptor protein (Perrimon, 1993). Recent evidence using germline mosaic flies, mutant for different
genes in the Tor RTK-mediated pathway, suggests that the Raf-MAP kinase cascade can also be activated from Tor in a separate pathway which is independent of Drk, Sos and Ras (Hou et al., 1995).

In general, it appears that the Ras-mediated signaling pathway is brought into play by a number of different RTK's which effect different changes within the cell (Feig, 1993). The fact that different RTK's can bring about different functions while still using the same set of highly conserved kinases presents something of a paradox. We might expect a sex specifically-expressed kinase, such as STK61, to be involved in modulating the effects of the Raf-MAP kinase cascade in a sex-specific way. Indeed, there may be a whole range of sex-specific and tissue-specific factors which are expressed in precisely defined areas which function to "funnel" the general driving force of the MAP kinase cascade down very specific pathways. The fact that STK61 contains OPA repeats, which are found in proteins encoded by genes with clearly defined expression patterns, suggests that this protein may be involved in such a tissue-specific function.

In fibroblasts, increasing the levels of cAMP leads to an inhibition of Raf activation by peptide hormone stimulation (Burgering et al., 1993). This indicates that an interplay between Ras-mediated and second messenger-mediated signaling can occur. Indeed, it is most likely that transduction of a particular signal to produce a specific effect involves a number of different pathways which act together to result in the appropriate response.

5.3.2 PROTEIN KINASES IN CELL CYCLE REGULATION

The oncogenic effects of hypermorphic alleles of genes involved in signal transduction pathways shows that a major purpose of the transduction of extracellular signals is to bring about changes in the cell division cycle. Indeed, the yeast cell cycle-regulatory phosphatase CDC25 was the first guanine nucleotide exchange factor identified in the Ras signaling pathway. A vast amount of work has
been done on the molecular control of the cell cycle and a simplified model derived from the obtained results is shown in figure 5.15 (reproduced from figure 1.6).

![Diagram showing the major gene products involved in cell division cycle regulation](image)

**Figure 5.15**
Diagram showing the major gene products involved in cell division cycle regulation (reviewed in Pines & Hunter, 1990; Woodgett, 1991; Enoch & Nurse, 1991; MacNeill et al., 1991; Glover, 1991). A great deal of this work was carried out on the fission yeast *S. pombe*, but homologous genes have been found in a number of organisms, including humans, and have been shown to be functionally interchangeable in a number of cases. S=S-phase (DNA synthesis phases), M=M-phase (Mitotic division phase), G1=Gap-phase 1, G2=Gap-phase 2.

The CDC2/p34 Serine/Threonine (S/T) kinase is thought to be the main 'workhorse' of cell cycle regulation, acting to drive the cell through the divisions between the phases of the mitotic cycle, dependent upon its phosphorylation state and association with the cell cycle-regulated Cyclin proteins. Dephosphorylation of CDC2 in the CDC2/CyclinB complex in G2-phase enables transition from G2 into M-phase. Dephosphorylation of CDC2 is brought about by the CDC2-activator, CDC25 phosphatase. The Weel protein (and its redundant homologue Mik1) is a Serine/Threonine/Tyrosine (S/T/Y) kinase which inhibits cell cycle progression and may phosphorylate CDC2. The Nim1 S/T kinase is thought to negatively regulate Weel via phosphorylation. Suc1 protein physically associates with CDC2 and regulates G2/M transition (negatively) and mitotic progression (positively).

Not only does the entire cycle depend upon the phosphorylation state of a protein, but this protein is itself a kinase; the CDC2 kinase. The targets of active CDC2 kinase include all manner of proteins which are required for the process of M-phase, such as nuclear lamins (for nuclear breakdown/reassembly), histone H1 (for
chromatin decondensation), the HMG proteins (involved in heterochromatin formation) and the myosin light chain (involved in spindle operation). The proteins involved in cell cycle regulation (including the kinases) are ubiquitous, with homologs in organisms as diverse as *Xenopus, Drosophila* and yeast. These proteins have been shown to be functionally interchangeable in a number of cases (MacNeill et al., 1991; Pines & Hunter, 1990; Nurse, 1990).

In *Drosophila* embryogenesis there is a particularly interesting example of cell cycle regulation (reviewed by Glover, 1991; Enoch & Nurse, 1991). After the male and female pronuclei fuse in the embryo, the first 10 nuclear divisions are the most rapid in the animal kingdom. At this stage there is no G1 or G2 phases but just rapid S-M cycling. The giant nuclei (*gnu*) gene is amongst a number of genes, including the cell cycle-regulated genes *polo, mh* and *fs(I)Ya*, which supply protein maternally to the syncitial blastoderm. Embryos from mothers mutant for the *polo* gene have disrupted early divisions with spindles not forming correctly. *mh* and *fs(I)Ya* mutations have the effect of preventing male and female pronuclear fusion which can result in the production of haploid nuclei in the embryo. Females mutant for *gnu* produce syncitial blastoderm embryos containing nuclei that cannot enter M-phase. However, these nuclei do not stall but go through repeated S-phase cycles, giving rise to giant nuclei. This indicates that, at this stage of development, the S and M phases are uncoupled, with entry into M-phase being independent of completion of S-phase. In agreement with this idea of uncoupling, treatment of the embryo with the DNA synthesis inhibitor aphidicolin results in nuclei that go through repeated M-phases without S-phases. This uncoupling may be due to the naturally high levels of maternally supplied String/CDC25 phosphatase present in the embryo, acting to drive CDC2 kinase into M-phase irrespective of the status of S-phase.

At nuclear division 14, when cellularisation has taken place, G2 phase begins and S-phase becomes coupled to M-phase. G1 phase doesn't enter the cycle until division 17. Embryos which are mutant for *string* stall in G2 at division 14, since maternally-supplied String is degraded by this time. Short pulses of ectopic
String/CDC25 phosphatase from a P-element construct, controlled by a heat shock promoter, can allow morphogenesis to occur fairly normally until mid embryogenesis, when death occurs (Edgar & O'Farrell, 1990). Thus, at division 14 the regulation of the cell cycle is brought under zygotic, rather than maternal, control. In wild type embryos, phosphorylated CDC2 kinase (inactive) accumulates during G2, awaiting the production of zygotic String/CDC25 phosphatase which will activate it and drive the cells into S-phase (Edgar et al., 1994).

As well as it's functions in signal transduction, the PP2A phosphatase is also required for cell cycle regulation, with Drosophila PP2A mutants showing defects in certain aspects of anaphase (Mayer-Jaekel et al., 1993). It is thought that the PP2A phosphatase may act in a similar way to String/CDC25 phosphatase, activating CDC2 kinase, or may regulate the activity of String/CDC25 phosphatase itself.

The cell cycle-regulatory genes are equally important in the male gonad, where they are required to direct the cell cycle during meiosis, although less is known about exactly what factors are regulating the cell cycle in this tissue. In contrast to the case of oocytes, spermatocytes do not contain large quantities of parentally-derived String/CDC25 phosphatase, and a second CDC25-like phosphatase, the Twine/CDC25 phosphatase, has been shown to be required for CDC2 kinase activation in this tissue (Sigrist et al., 1995). Both male and female Drosophila, mutant for twine, exhibit meiotic defects which lead to sterility. It appears that the String/CDC25 phosphatase is able to compensate for loss of Twine/CDC25 phosphatase in oocytes since some meioses do occur. In males, however, neither meiotic division occurs in twine mutants. This evidence suggests that meiosis in the testis may require gene products which are not required for cell cycle regulation in other tissues.
S/T kinases are also involved in the differentiation of segment boundaries in *Drosophila*. The genes which set up the parasegments in the developing embryo are called the segmentation genes. Figure 5.16 summarises the roles which have been proposed for these genes in maintaining segment boundaries through development.

**Figure 5.16**
Diagram showing model for functions of segment polarity genes (Arias et al., 1988; Ingham, 1991; see main text for further details). In the early embryo, differential concentrations of the products of the *fushi-terazzu* and *even-skipped* pair-rule genes determines cells to be 'competent' to express either *wingless* (*wg*) or *engrailed* (*en*). Thus, *wg*-competent and *en*-competent bands of cells alternate along the length of the embryo. However, *Wg* and *En* proteins are expressed only in the cells bounding the parasegment boundary, maintaining the positional identity of cells within the parasegment. This localised expression pattern is maintained as shown above. The *wg* gene is repressed via the Patched (*Ptc*) receptor. Expression of *hedgehog* (*hh*) occurs only in *en*-expressing cells and alleviates *Ptc*-mediated repression. The *en* gene may be activated via the Armadillo (*Arm*) protein, which relocalises from the plasma membrane to the cytoplasm under the action of *Wg* signaling. The *dishevelled* (*dsh*) gene also appears to be required for reception of this signal. Naked (*Nkd*) and Shaggy (*Sgg*) proteins are believed to prevent *Arm* function in non-boundary cells. Mutation of the *sgg* gene allows *Arm* relocalisation in the absence of *Wg* signal, and *Sgg* has been shown to phosphorylate *Arm*. Thus, activation of *Arm* may occur by dephosphorylation, under the action of *Wg* signal from neighbouring cells.
Two S/T kinases have been implicated in this process. These kinases are encoded by the genes *fused* (*fu*) and *shaggy* (*sgg*, or *zeste-white 3*). Wild type embryos exhibit clear segmentation, with the anterior portion of the segment containing a denticle belt and the posterior composed of naked cuticle. Embryos carrying mutant alleles of *fu* develop a mirror image of the anterior denticle belt in the posterior of each segment (Nüsslein-Volhard & Weischaus, 1980). Other genes which produce this phenotype are *arm, dsh, en, hh* and *wg*. We can see from figure 5.16 that these last five genes are all thought to be involved in the production or transduction of the *wg* signal which allows *en* expression in the anterior band of each parasegment. Since *wg* expression is not maintained in embryos mutant for *fu* (Limbourg-Bouchon et al., 1991), the Fused S/T kinase is thought to have a role in the transcriptional activation of the *wg* gene. This could either be a direct activation of *wg* transcription, or an indirect modulation, possibly by activating the En protein which would lead to Hh-repression of Ptc, thereby allowing expression of *wg*.

Embryos carrying mutations in the genes *nkd* or *sgg* exhibit the opposite phenotype to that described above, with a mirror image of the posterior naked cuticle developing in the anterior of each parasegment (Nüsslein-Volhard & Weischaus, 1980). As shown in figure 5.16, these genes are believed to be involved in the repression of *en* expression in the absence of *wg*-induced signaling. Both *wg*-induced signaling and mutations of *sgg* result in cytoplasmic accumulation of the Arm protein, which is normally membrane-localised (Peifer et al., 1994a). Additionally, Sgg kinase is capable of phosphorylating Arm protein, and this phosphorylation is suppressed by Wg signaling (Peifer et al., 1994b). This strongly suggests a model whereby the Sgg S/T kinase is responsible for maintaining the phosphorylation state of Arm, until the Wg-signal causes it to be de-phosphorylated, enabling it to dissociate from the plasma membrane and lead to activation of *en* expression.
Kinase-mediated signaling is also vital to the establishment of the dorsal-ventral axis of the *Drosophila* embryo. In essence, this system involves the transduction of a signal from the overlying follicle cells which is received by receptors encoded by the gene *toll* on the ventral surface of the embryo. This results in a nuclear gradient of the morphogen, Dorsal protein, with ventral nuclei containing higher levels of this protein (Roth *et al.*, 1989). The downstream targets of the putative transcription factor Dorsal are *zerknüll* and *decapentaplegic*, which are activated at low Dorsal concentrations and *twist* and *snail*, which are activated at high Dorsal concentrations (see chapter 1 for details). When Toll is inactive, a S/T kinase encoded by the *pelle* gene is found near the plasma membrane, complexed with a protein of unknown function, Tube, and Dorsal is localised in the cytoplasm in a complex with the protein Cactus. Recent co-transfection assays in *Drosophila* culture cells have suggested that the S/T kinase Pelle is involved in mediating Dorsal nuclear localisation (Norris & Manley, 1996). These studies show that Pelle is capable of enhancing both nuclear localisation and transcriptional activity of Dorsal. It was also shown that these Pelle-mediated effects were enhanced by over-expression of the *toll* gene. The Pelle kinase is capable of phosphorylating the Tube protein *in vitro* (Grosshans *et al.*, 1994). The Tube protein has been shown to be essential for the differentiation of dorsal-ventral polarity (Letsou *et al.*, 1993). Norris & Manley, 1996, showed that Tube could re-localise to the nucleus and that this was dependent upon Toll and Pelle function. Thus, it seems likely that activation of Toll receptor leads to activation of Pelle kinase which phosphorylates Tube causing it to permit dissociation of Dorsal from Cactus and relocalisation of Dorsal to the nucleus.

One of the targets of Dorsal, the *decapentaplegic (dpp)* gene, encodes a growth factor with homology to the TGF-β family and is itself involved in kinase-mediated signaling. The S/T receptor kinase encoded by the gene *saxophone* is thought to be a Dpp receptor, mediating it's functions in signaling (Xie *et al.*, 1994). Generally, growth factor ligands of the TGF-β family are specific to receptor kinases which are
serine/threonine-specific, with Dpp acting via a number of receptors of this type (Ruberte et al., 1995).

Another gene which has a role in the differentiation of dorsal-ventral polarity in Drosophila is the torpedo locus. This gene encodes the Drosophila homolog of the mammalian epidermal growth factor RTK. Mutations at this locus produce highly pleiotropic phenotypes, including embryo ventralisation, germ band extension defects and CNS defects (Clifford & Schüpbach, 1992). Since mutations in the genes encoding the Sos protein can suppress or enhance the eye phenotypes of certain torpedo alleles, it seems that the Torpedo RTK leads to transduction of signals via a Ras-mediated pathway in at least some cases (Rogge et al., 1991). Alleles of torpedo which cause embryo ventralisation are maternal in nature, with embryos from mutant mothers exhibiting the phenotype. Although the precise details of the basis of this phenotype are unclear, the role of Torpedo is believed to be the repression of the Toll pathway, at the dorsal side of the embryo (Pawson & Bernstein, 1990). The current model for this Torpedo function states that a signal is sent to the dorsal follicle cells from the embryo which activates the Torpedo RTK. Active Torpedo then transduces this signal resulting in the production of another signal which is sent to the embryo and acts to repress the Toll pathway. RNA from the genes K10 and gurken is found localised around the dorsal nuclei in the syncitial blastoderm and a gurken null mutation results in torpedo-like ventralisation. It is possible that these genes are responsible for the initial signal to the overlying follicle cells (Pawson & Bernstein, 1990).

5.3.5 DIRECT SIGNAL TRANSDUCTION VIA PROTEIN KINASES

Thus far we have been discussing kinase-mediated signaling pathways which involve fairly complex interactions between several factors. However, signals can also be transduced from extracellular ligands into nuclear effects by surprisingly direct means. The mammalian transcription factor p91 directly associates with the activated epidermal growth factor RTK via it's SH2 domain (Fu & Zhang, 1993).
This leads to activation of p91 by tyrosine phosphorylation. When activated, p91 re-locates to the nucleus and activates transcription of the \textit{c-fos} gene (Fu & Zhang, 1993). Additionally, a nuclear localisation sequence (NLS) has been identified in several proteins and appears to be sufficient for nuclear relocation of proteins (Whiteside & Goodbourn, 1993). It is possible that there are a number of signal transduction proteins which utilise systems such as these to bring about direct transduction of extracellular signals.

5.3.6 PROTEIN KINASES IN CYTOSKELETAL INTEGRITY

Another interesting S/T kinase has been found which is required for the correct development of the eye in \textit{Drosophila}. The \textit{nina-c} locus encodes two proteins with putative kinase domains joined to myosin heavy chain-like domains suggesting a role involving interaction with cytoskeletal proteins. Mutants at this locus have eye photoreceptors with reduced rhodopsin levels which gives rise to unusual neural responses. This lack of rhodopsin appears to be due to the reduction in size of the rhabdomeres (Membranes containing rhodopsin and other phototransduction molecules) present in these mutants (Seigfried \textit{et al.}, 1990). It is likely that the \textit{nina-c} kinase acts upon cytoskeletal proteins in the cell to maintain the integrity of the rhabdomeres.

5.3.7 NON-RECEPTOR TYROSINE KINASES

Non-receptor tyrosine kinases are also important in \textit{Drosophila} development as indicated by the studies on the \textit{abl} locus (Hoffman, 1991). The \textit{abl} gene encodes a predicted protein with SH2, SH3 and kinase homology. Mutants have fairly mild phenotypes, with hypomorphic alleles allowing flies to develop to adulthood but with irregular, roughened eyes. Even individuals carrying complete null \textit{abl} alleles develop as far as the late pupal stage. It would appear that functional redundancy is the cause of the mildness of these phenotypes, as embryos which are also mutant for the genes \textit{disabled}, \textit{failed axon connections}, \textit{prospero} and the \textit{fasciclin I} gene.
(Fasciclin I is a cell adhesion molecule produced in the CNS at the same time as abl when axon connections are forming) have much more severe phenotypes of the CNS and muscle resulting in late embryonic or early larval death (Hoffmann, 1991). The function of Abl is to some extent independent of its kinase activity, since complete rescue of abl/abl mutant flies is achieved by germline transformation of a P-element construct which ectopically expresses kinase-deficient Abl protein. Kinase function is required for other processes, however, since only a construct expressing the active kinase fully rescues the double mutant abl/abl, disabled/disabled (Hoffmann, 1991).

5.3.8 PROTEIN KINASES IN NEUROGENESIS

S/T kinases have been shown to be involved in postembryonic differentiation of the Drosophila CNS. Studies on cell-cell interactions involving the Notch and Delta neurogenic loci have provided a great deal of information regarding the mechanisms of early embryonic neurogenesis (for reviews see Vaessen et al., 1990; Artavanis-Tsakonas & Simpson, 1991). However, much less is known about which genes are responsible for directing later postembryonic differentiation of the CNS. Mutant alleles of the minibrain (mnb) gene result in a reduction in size of both male and female brain, with female brains showing slightly more reduction than males (Tejedor et al., 1995). The overall architecture of brains from mnb mutant flies seems to be fairly well preserved, although the optic lobes are disproportionately reduced in size and some neuron loss is seen. As might be expected, mnb adults exhibit some behavioural abnormalities, including loss of attraction to a vertical stripe, and loss of odour discrimination. Courtship behaviour has not been examined. The mnb gene encodes a S/T kinase with homology to cell cycle regulatory kinases and the mnb phenotype seems to result from a reduction in the number of progeny cells produced from certain neuroblasts. This suggests that the Mnb kinases may be involved in regulating neuron-specific cell division.

The analysis of the mnb gene has interesting implications with regard to stk61. Clearly, male and female Drosophila exhibit sex-specific behaviours which must
have a foundation in CNS differences. Indeed, the CNS structure is sexually dimorphic, with females having larger numbers of Kenyon fibres and males exhibiting extra divisions of the terminal abdominal neuroblasts (Technau, 1984). These extra male neuroblast divisions are dependent upon the \textit{dsx} gene, however, making it unlikely that \textit{stk6l} has a significant role to play in this particular process. RNA \textit{in situ} evidence from our lab suggests that \textit{stk6l} may be expressed in the brain of third instar larvae (D. Clyde, pers. comm.), although it is not known whether fully processed transcript is present or if active protein is produced in this tissue. However, \textit{stk6l} may have some role in neuroblast regulation, influencing the differentiation of subtle sexually dimorphic CNS characteristics.

5.3.9 CONCLUSIONS

There are examples of the requirement for kinases in every major developmental process, from the gradual restriction of totipotency (as in the cases of \textit{fu} and \textit{sgg}) to terminal differentiation (as in the action of \textit{nina-c}). Based upon these observations, some specific predictions were made concerning possible functions of the OPA repeat-containing S/T kinase encoded by the \textit{stk6l} gene.

In general, it is not hard to imagine how S/T kinase-mediated processes such as signal transduction and cell cycle control could contribute to the establishment of sexually dimorphic characteristics. Indeed, the ubiquitous nature of protein kinase-regulated pathways in development makes it reasonable to assume that sex differentiation processes will involve these proteins. However, since the \textit{yp} genes remain the only cloned sex differentiation genes, a role for protein kinases in sex differentiation is still to be demonstrated. For this reason, we hope that studies on the sex-specifically regulated S/T kinase gene, \textit{stk6l}, will provide an inroad into the regulatory processes required for differentiation of the sexually dimorphic features of \textit{Drosophila}.
In *Drosophila* there are many processes in which a role for a sex-specifically expressed S/T kinase, such as that encoded by *stk61*, could be envisaged. Clearly, initial rounds of nuclear division in the syncitial blastoderm during embryogenesis requires close control of the cell division cycle. The ovary-specific 3.5kb *stk61* transcript may express STK61 protein in this tissue which could certainly play a part in this process. Similarly, in the developing female embryo an inductive signal from the somatic component of the gonad is required for the germ cells to be determined as female (see chapter 1 for details). Again, an ovary-specific kinase may be required to allow female germ cells to respond to this signal. Control of the cell division cycle is equally important in male germ cell development, where a testis-specific kinase (possibly encoded by the testis-specific 3.0kb *stk61* transcript) may come into play. Close control of the cell division cycle, as well as modulation of positional information, would also be required to bring about the sex-specific structures of male and female *Drosophila* abdominal tissue. The fact that STK61 contains OPA repeats indicates that it may be localised to a specific tissue or area in the female body, where it may play the part of a sex-specific "filter" of extracellular signals, allowing signals to be interpreted in a female-specific way. This could be a developmental role, where *stk61* may act to define a specific group of cells as female, thereby influencing their development. Alternatively, *stk61* may be active in the adult fly, perhaps acting to transduce pheromonal, or other sex-specific signals. In the next chapter 1 will discuss what strategies might shed light on the functions of the *stk61* gene.
Chapters 3, 4 and 5 describe the cloning and characterisation of a gene which we have named Serine/Threonine Kinase 61 (stk61). This gene was isolated via a sex-specific differential screen designed to detect genes which produce non-gonadal sex-specific transcripts. The stk61 gene produces several sex-specific transcripts, including a 4.5kb female carcass-specific transcript, a 3.5kb ovary-specific transcript and a 3.0kb testis-specific transcript. cDNAs representing the testis-specific and female carcass-specific transcripts were cloned and sequenced. Both cDNAs contain an identical long open reading frame, encoding a OPA-repeat-containing serine/threonine-specific protein kinase. This open reading frame is functionally relevant, since apparently full-length protein can be translated in bacterial cells. Northern blot analysis indicates that production of the fully processed female carcass-specific transcript is dependent upon the tra sex determination gene. Consensus sequences known to be required for Tra/Tra-2 regulation of dsx transcripts have been delimited in the sequence of the stk61 cDNAs. Thus, the female carcass-specific transcript is likely to be under direct control of the Tra and Tra-2 proteins. The cDNA representing the testis-specific transcript contains within its 5'UTR two copies of a male germline-specific translational control element (TCE) which mediates translational repression until late in spermiogenesis.
6.2 IMPLICATIONS OF RESULTS AND FUTURE WORK

6.2.1 FEMALE CARCASS-SPECIFIC TRANSCRIPT

The apparent mode of regulation of the female carcass-specific transcript suggests that the *stk61* gene may be involved in sex-specific differentiation of non-gonadal tissue.

Mutations in the *tra* or *tra-2* genes result in female somatic development being subverted to the male pathway, with *XX/tra* or *XX/tra-2* flies developing as pseudo-males. Since the complete splicing of the female carcass-specific *stk61* transcript appears to be under direct control of the *tra* gene product, it is reasonable to assume that this *stk61* transcript produces a protein which is involved in transducing the 'ON' state of the *tra* gene into female-specific development. This is of particular interest because the majority of sex-specific features which are under *tra/tra-2* control represent not a direct dependence upon *tra* or *tra-2*, but rather a dependence upon the *dsx* gene, which is regulated by *tra* and *tra-2*.

Certain aspects of male courtship behaviour have been hypothesised to be repressed directly by Tra and Tra-2 (see chapter 1 for details). Since *stk61* encodes a S/T-specific protein kinase, it is likely that STK61 protein is involved in production or reception of an extracellular signal of some kind. A possible role for *stk61* in reception of sex-specific behavioural signals (e.g. pheromones), or modulation of global signals in a sex-specific way, cannot be ruled out. There are two sex-specific processes in *Drosophila* which are known to require intercellular signaling. The first of these is the differentiation of the male-specific muscle of Lawrence (MOL) in the fifth abdominal segment. In this case, both *tra* and *tra-2* genes are required to prevent development of the muscle, but *dsx* is not. Since the determining factor seems to be the sex of the MOL innervating axons, rather than the sex of the MOL itself, *tra* and *tra-2* must be having their effect in these axons. Clearly, a signal must be sent from the innervating axons to the muscle precursor cells, in the developing...
embryo, which either initiates or represses the development of the MOL. In the segment polarity system (see figure 5.16), the Shaggy S/T kinase is directly involved in preventing production of a signal which acts upon neighbouring cells. In the same way, the STK61 S/T kinase may act in the female MOL-innervating axons to prevent production of a MOL-inducing signal. The presence of OPA glutamine repeats in the STK61 protein suggests that the protein is localised to a specific tissue or subset of cells. At present, work leading to the production of STK61 antibodies is underway in the lab, which will enable the expression pattern of the protein to be determined. It may be that STK61 protein will be found to be localised to certain PNS axons, which would support a role for $s_{1k61}$ in MOL repression. As discussed previously, another OPA-containing protein, the Notch protein, is known to function in early neurogenesis. This lends further weight to a possible neurogenic role for STK61.

6.2.2 TESTIS-SPECIFIC TRANSCRIPT

At present it is unclear from the available whole mount RNA in situ hybridisation data whether $s_{tk61}$ RNA is localised to the somatic or to the germline component of the testis. However, the presence of TCE elements in the 5' UTR of the testis-specific transcript strongly suggests that STK61 protein functions in the male germline. Precise cell-division-cycle regulation is vital in development of spermatocytes and a number of protein kinases which are involved in spermatocyte cell cycle regulation have already been discussed. However, it is unlikely that STK61 kinase is involved in this process in the male germline, since TCE-containing transcripts only become translationally active late in the elongation phase of spermiogenesis, when haploidisation has already occurred. It is much more likely that STK61 plays some part in the later processes which lead to morphological differentiation of the spermatozoon. This may involve transduction of signals sent from the somatic component of the testis, or even regulation of the integrity of cytoskeletal structures (as with the Nina-C kinase). Again, determination of the precise expression pattern of testis-specific STK61, via anti-STK61 antibody in situ analysis, will hopefully provide some clues as to the role of the protein in this tissue. Similarly, antibody in
situ analysis could be used to examine the function of the TCE elements present in the testis-specific transcript. No germ cells beyond the primary spermatocyte stage are found in 3rd instar larvae. As with the Mst(3)CGP gene family, we would not expect to find STK61 protein in these cells, unless the TCE sequences are mutated.

6.2.3 OVARY-SPECIFIC TRANSCRIPT

As yet, we have not isolated a cDNA which represents this transcript. However, Northern blot analysis shows that it does not contain the long 3' UTR sequence which is specific to the female carcass-specific transcript. It does contain sequences covering the open reading frames of both of the other transcripts. It is possible that the ovary-specific transcript contains a 3' UTR sequence which is not found in either of the other two sex-specific transcripts.

The ovary-specific transcript was not detected in XY/trd^{FEM} pseudofemale flies. However, examination of the gonads of the pseudofemales used for this analysis showed that they were not well transformed towards female-like ovaries, still containing accessory gland-like structures (figure 4.4). More complete transformation to 'pseudo-ovaries' can be achieved by heat-shock which increases production of ectopic Tra in flies carrying the trd^{FEM} construct. We would expect such pseudo-ovaries to contain the ovary-specific stk61 transcript, only if it is expressed in the somatic component of the ovary.

Determination of the precise localisation and sequence of this transcript is a high priority for future work. If the stk61 ovary-specific transcript is expressed in the somatic component of the ovary, it would be reasonable to suppose that this transcript would be under the direct control of the tra gene product, in the same way as is proposed for the female carcass-specific transcript. This is particularly interesting in the light of evidence which shows that the somatic inductive signal, which is required for female germline differentiation, is under control of tra-2, but not dsx (see chapter 1 for details). It is attractive to propose that the role of the stk61
ovary-specific transcript may be in the production of this signal. However, the \textit{tra} gene does not appear to be required for this signal to be produced. Since the female carcass-specific \textit{stk6l} transcript appears to be more dependent upon \textit{tra} than \textit{tra-2} (or, at least, equally dependent), this may not be the function of \textit{stk6l} in this tissue. The ovary-specific transcript may, instead, be germline-specific, possibly acting to transduce signals within the oocyte. The elucidation of the functions of this transcript awaits cloning of the cDNA and precise determination of the RNA and protein localisation within the female gonad.

6.2.4 FUTURE PROSPECTS

The work presented here strongly suggests that at least some of the transcripts from the \textit{stk6l} gene are regulated directly by Tra/Tra-2 proteins. Clear confirmation of this model by Northern analysis has proved extremely difficult, for reasons discussed earlier. One approach which could prove more effective is RNAase protection analysis. This would involve subcloning a cDNA fragment bounding the putative Tra/Tra-2-regulated splice site into a transcription vector. Thus, a radiolabelled riboprobe could be generated, by \textit{in vitro} transcription. This riboprobe would then be hybridised with RNA extracts from male and female carcass tissue and treated with RNase, to remove any single-stranded molecules. Analysis of these reactions by polyacrylamide gel electrophoresis and autoradiography would reveal a band diagnostic of fully-spliced \textit{stk6l} transcript. If the female carcass-specific transcript is produced by removal of intron 4, as proposed, we would only expect to see this band in reactions performed with female carcass RNA. This same technique could then be used to examine the involvement of \textit{tra} and \textit{tra-2} gene products in removal of intron 4, by performing RNAase protection analysis using RNA extracts from \textit{tra} or \textit{tra-2} mutant flies.

In the longer term, a direct role for Tra and Tra-2 in \textit{stk6l} intron 4 removal could be demonstrated by use of \textit{Drosophila} cell culture co-transfection assays. This technique has been used to demonstrate a direct requirement for Tra and Tra-2
proteins in splicing regulation of transcripts from the \textit{dax} gene (see chapter 1 for details). An expression vector containing \textit{stk6l} genomic DNA, encompassing intron 4, would be transfected into culture cells either with or without expression vectors encoding Tra and/or Tra-2 proteins. The \textit{stk6l} transcription products could then be analysed using RT-PCR, or RNAase protection assays. If the intron 4 splice acceptor site is indeed regulated by Tra/Tra-2, we would only expect to see this intron removed in cells expressing Tra and Tra-2 proteins.

6.2.4.1 P-Element Mutagenesis.

In order to definitively assign a function to the \textit{stk6l} gene, it will be necessary it introduce mutations into the gene. It is possible to interfere with the function of a specific gene in \textit{Drosophila} by introduction of a P-element transposon into the locus. When males with P-elements present in the genome are mated to females lacking P-elements, the P-elements are mobilised in the germ cells of the resulting embryo, 'hopping' to new positions in the genome. The efficiency of this process can be increased by mating flies which have mobilisable transposase-deficient P-elements to flies with un-mobilisable P-elements (\textit{A2-3} elements) which produce transposase (Zhang & Spradling, 1993). This process is random and there is no way of predicting where the P-elements will relocate. However, use of a P-element-specific oligonucleotide primer and a gene-of-interest-specific primer enables identification of flies in which a P-element has relocated close to a gene of interest, via PCR analysis (Kaiser & Goodwin, 1990). Males of the mobilisable P-element strain are mated to females of the \textit{A2-3} transposase-producing strain. F1 males are collected and mated to wild-type females to produce true-breeding F2 flies. The eggs from F2 females can be collected and analysed by PCR. One fly with a P-element inserted close to the gene of interest can be detected in a DNA sample representing 1000 flies. By dividing these flies into smaller and smaller groups, re-testing each time, single flies with the P-element insertion can be isolated. Conversely, P-elements which hop from close to the gene of interest can result in deletions in the locus. These can be detected by Southern blot analysis.
6.2.4.2 Germline Transformation Studies.

Another approach which can provide information on the function of a gene is to mis-express the gene ectopically in inappropriate tissues, or in flies of the inappropriate sex. This can be done in *Drosophila* by using germline transformation. All that is required for integration of a P-element into the genome of *Drosophila* is the 31bp inverted terminal repeats which lie at the extreme ends of the P-element and a supply of transposase protein. Thus, sequences encoding STK61 protein can be introduced into the germline of embryos under the control of a heat-shockable promoter, as shown in figure 6.1. A 'helper' element is co-injected which provides transposase, but cannot itself integrate into the genome due to mutated terminal repeats.

![Figure 6.1](image)

**Figure 6.1**
This experiment is currently being carried out by Diane Harbison, in our lab. It is difficult to predict exactly what phenotypes ectopic expression of \textit{stk6l} would produce. However, based upon the predictions discussed above, we might expect to see a number of sex-specific phenotypes such as gametogenic, muscular, cuticular and behavioural defects. The results of this experiment will allow predictions to be made about what phenotypes may be produce by 'site-selected' P-element mutagenesis (see above).

The ability to ectopically express \textit{stk6l} would also enable identification of genes which genetically interact with \textit{stk6l}, as shown in figure 6.2.

![Diagram](image)

\textbf{Figure 6.2}
Use of ectopic expression of \textit{stk6l} to identify genetic interactions. goi=gene of interest. See text for details.

Clearly, choice of the 'gene of interest' shown in figure 6.2 will depend upon evidence such as the localisation of \textit{STK61} protein and the phenotypes caused by \textit{stk6l} miss-expression, which will help to identify particular genetic pathways in
which stk6l might play a part. Since stk6l encodes a S/T kinase, possible genes of interest might include genes known to be involved in signal transduction pathways, such as the ras gene which is involved in a number of different transduction processes in the developing embryo. We might expect ectopic stk6l expression to cause a modulation of ras phenotypes. Such ras-mediated processes might include specification of the terminal regions of the embryo, or specification of the R7 photoreceptor.

In the longer term, the transcriptional regulation of the stk6l gene could also be investigated by germline transformation. Constructs containing LacZ reporter gene sequences under control of genomic DNA sequences, 5' to the transcriptional start site of cloned cDNA's, may help to determine whether transcriptional regulation plays any part in the production of sex-specific stk6l transcripts. It is possible that separate promoter elements are responsible for production of the various stk6l transcripts, which influences the splicing pattern of these transcripts. This type of regulation is seen in the Sxl gene, where transcription from the early promoter, PE, results in the production of active fully-spliced Sxl transcript, despite the lack of active Sxl protein (See chapter 1 for details). Transcripts produced from the late promoter, P, however, require the action of Sxl protein for active fully-spliced Sxl transcript to be produced. We have seen that intron 4 is spliced out of the testis-specific stk6l cDNA, even in pseudo-testes, which lack active Tra-2 and Tra (figure 4.3). Yet, the removal of this intron appears to require Tra and Tra-2. It may be that a system analogous to that of Sxl exists, whereby transcription of stk6l from a testis-specific promoter somehow bypasses the need for splicing regulation. When specific stk6l promoter sequences have been delimited, factors which bind to these elements could be identified by gel-shift assays and protein expression library screening.

Germline transformation could also be used to construct flies expressing LacZ reporter gene transcript containing the stk6l TCE elements in 5' UTR sequences.
This would enable a convenient assay to test the translational control of these TCE elements, and the effect of mutation of the elements.

In addition, the function of the cDNA11-specific 3'UTR could be investigated by germline transformation. Expression of a LacZ-encoding transcript, fused to the cDNA11-specific 3'UTR, would reveal whether the 3'UTR acts to mediate translational control or specific mRNA localisation.

Clearly, the stk6l locus is a complex one, which will require a great deal of further study to fully understand its regulation and function. The work presented here identifies the stk6l gene as a good candidate for a sex differentiation gene which has a number of features which are unique and will make this gene highly interesting as a possible inroad into the largely unknown territory of somatic sex differentiation in Drosophila.
REFERENCES


237


239


*Drosophila* yolk protein genes: different segments of one enhancer have
different cell-type specificities that interact to give normal expression. *Genes
Dev.* **4**, 613-623.

**191**, 28-36.

Lucchesi, J. C. and Skripsky, T. (1981). The link between dosage compensation and


consequences of alternate splicing in sex determination and differentiation in

Dev.* **1**, 307-312.


between sex-determination and sexual differentiation in *Drosophila

maternal-effect loci required for both the anteroposterior and dorsoventral

247


252


253


255


257


258


Wang, J., and Bell, L. R. (1994). The Sex-lethal amino acid terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. Genes Dev. 8, 2072-2085.


APPENDIX

PUBLICATIONS
The Developmental Consequences of Alternate Splicing in Sex Determination and Differentiation in Drosophila

Colin MacDougall, Diane Harbison, and Mary Bownes
Institute of Cell and Molecular Biology, University of Edinburgh, Darwin Building, The King's Buildings, Edinburgh EH9 3JH, United Kingdom

INTRODUCTION

Many eukaryotic genes generate alternately spliced transcripts which can produce different proteins or which have altered translational controls. One of the most direct demonstrations that alternately spliced forms of transcripts lead to different developmental consequences lies within the sex determination pathway of Drosophila. The doublesex (dsx) gene at the end of this pathway in somatic cells encodes two differently spliced transcripts, one specific for males and one specific for females (Burtis and Baker, 1989). These encode proteins with a common DNA binding region and a sexually unique carboxy terminus; DSXM and DSXF act as transcription factors and have opposing activities. The main developmental consequences are the repression of a set of downstream female-specific differentiation genes by DSXM in males and the repression of certain male characteristics by DSXF in females (for reviews see Slee and Bownes, 1990; Steinmann-Zwicky et al., 1990; Ryner and Swain, 1995).

Many of the sexual differences between male and female Drosophila are controlled by the two alternate products of the dsx gene (Burtis and Baker, 1989). They direct both the determination of sex-specific characteristics in the imaginal cells and the maintenance of determination throughout subsequent cell divisions. This regulation leads to the final differentiation of male or female genitalia, the differences in pigmentation patterns in the abdomen of each sex and in other sex-specific bristle patterns, such as the sex comb on the first leg of the male. Differences between the sexes in the pattern of nerve cell divisions are also directed by the two related DSX proteins (Taylor and Truman, 1992), as is the selection between the male or female development of the gonad into a testis or ovary and its subsequent differentiation (Szabad and Nöthigter, 1992). The yolk protein genes which are expressed in the adult female fat body are targets of the DSX protein (Burtis et al., 1991) and their yolk protein products are essential for oocyte development.

However, the use of alternate splicing in sexual development in Drosophila is not limited to dsx. Alternate splicing of dsx transcripts is controlled by the products of the transformer (tra) and transformer-2 (tra-2) genes (Nagoshi et al., 1988). tra RNA is also alternately spliced, in this case with the dramatic consequence that in males no functional protein product is made (Butler et al., 1986; McKeown et al., 1987), whereas in females an RNA binding protein is produced which interacts with the tra-2-encoded RNA binding protein (Belote and Baker, 1982), directing the female-specific splicing of dsx. It seems likely that tra and tra-2 have other targets in addition to dsx that are important for sexual development, since several aspects of sexual dimorphism depend upon the tra/tra-2 genes but are independent of dsx (Taylor et al., 1994). This includes courtship behaviour, the development of abdominal cells, which produce a female pheromone, and the correct innervation of nerves needed for the development of a male-specific muscle (Lawrence and Johnston, 1986; Taylor, 1992).

The female-specific splicing of tra RNA is itself directed by an alternately spliced gene product. This is encoded by the Sex-lethal (Sxl) gene that is at the head of the sex-determination hierarchy (Cline, 1984, 1993). Sxl produces many transcripts, including several specific to female somatic cells, that generate a functional RNA binding protein. The male mRNA from Sx does not encode a functional protein (Bell et al., 1988). The SXL protein is known to direct the female-specific splicing of its own RNAs as well as that of tra (Inoue et al., 1990; Bell et al., 1991; Horabin and Schedl, 1993a,b; Samuels et al., 1994).

The female product of Sxl also directs suppression of the
Activation of Sex-lethal

The Y-chromosome does not play a role in sex determination, which leads to sex-specific lethality when Sxl is inappropriately activated or repressed. Female differentiation of the germ-line also depends upon Sxl (Steinmann-Zwicky, 1992). From its multiple functions, some of which are not mediated through the activities of tra and tra-2, it seems that Sxl, like tra, must have more target genes that remain to be discovered, namely those involved in dosage compensation, size (females are larger than males), and germline differentiation. Thus this pathway provides excellent insight into how alternate splicing can be used to ensure that very different developmental decisions are taken during development. Its branched nature, with multiple targets at each point in the hierarchy, shows how a complex network of interrelated processes can be controlled at the level of mRNA processing. Figure 1 shows the morphological sexual differences between adult males and females, and Fig. 2 summarises the sex determination pathway in somatic cells.

SOMATIC SEX DETERMINATION

The primary determinant of sex in flies is the number of X-chromosomes to sets of autosomes, the X:A ratio (see reviews by Slee and Bownes, 1990; Steinmann-Zwicky et al., 1990, Cline, 1993; Ryner and Swain, 1995). Those flies that have one X-chromosome to two sets of autosomes (X:A = 0.5) are male, whilst those that have an X:A ratio of 1 (i.e., two X-chromosomes to two sets of autosomes) are female. The process of somatic sex determination is largely cell autonomous; there appears to be no hormonal component. This can be inferred from studies on gynandromorph flies possessing both XO and XX cells which develop into male and female tissue autonomously. Since the male and female cells are exposed to the same compounds circulating in the haemolymph, a hormonal influence can be excluded. The Y-chromosome does not play a role in sex determination, unlike the process in mammals; rather it carries genes which are required to complete spermiogenesis (Fuller, 1993).

Activation of Sex-lethal

The X:A ratio is assessed by a number of zygotic loci that are located on the X-chromosome. These activate Sex-lethal and are known as numerator elements (sisterless-a (sis-a), sisterless-b (sis-b), and runt). There are also repressor elements (or denominators) on the autosomes (deadpan (dnp)) (for review see Parkhurst and Meneely, 1994). An additional X-linked locus known as sisterless-c (sis-c) has also been identified; this appears to act as a numerator element although its effects are weaker than those of either sis-a or sis-b (Cline, 1993). Numerator elements act as feminising elements, while the known denominator acts as a masculinising element (Cline, 1993). Alterations in the dosage of these elements leads to sex-specific lethality. A reduction in the numerator dose results in female-specific lethality, while an increase is lethal to males (Cline, 1988; Younger-Shepherd et al., 1992). The denominator element exhibits the reciprocal phenotype, an increase in dosage resulting in female lethality. The assessment of the X:A ratio affects the activation of the gene Sxl. Cells that have an X:A ratio of 1 (i.e., female) activate Sxl, while those that have an X:A ratio of 0.5 (i.e., male) do not activate Sxl.

Several of the numerator elements and the denominator element deadpan have been cloned and characterised at the molecular level. sis-a is a member of a family of transcription factors known as basic leucine zippers (Erickson and Cline, 1993a). sis-b (which corresponds to scute α/Tα, a member of the achaete/scute complex) encodes a basic helix–loop–helix protein (bHLH) (Cline, 1988; Torres and Sánchez, 1989; Erickson and Cline, 1993b). The pair-rule segmentation gene runt has also been implicated in sex determination (Duffy and Gergen, 1991). The predicted runt protein shows homology with a family of transcriptional regulators, including the polyoma enhancer binding protein. Interestingly, this gene, unlike sis-a and sis-b which act throughout the embryo, appears to affect sex determination only in the central region of the trunk. To date, only one denominator element has been identified, the proneural gene dnp, which encodes a bHLH protein. As expected for a denominator element, alterations in the gene dosage of this locus result in sex-specific lethality (Younger-Shepherd et al., 1992; Cline, 1993).

In common with many other processes in Drosophila, sex determination relies upon maternally contributed RNAs laid down in the oocyte during oogenesis. One of these is encoded by the gene daughterless (da) (Caudy et al., 1988; Crommiller et al., 1988). As its name suggests, females mutant at one allele of this locus (da<sup>1</sup>) do not produce any female progeny. Again, da encodes a protein that contains a bHLH domain. The neural locus extramacrochaetae (emc) has been predicted to act as a negative regulator of Sex-lethal. The protein contains a HLH domain but does not contain the basic residues which are essential for DNA binding (Younger-Shepherd et al., 1992; Bier et al., 1992).

FIG. 1. Morphological differences between males and females. (a) Male and female flies; note the male is smaller than the female. (b) Male and female foreleg; note the sex comb on the male. (c) Male muscle and equivalent segment in female (photographs courtesy of Dr Peter Lawrence, MRC, Cambridge). (d) Male and female genitalia and abdominal pigmentation; note the more extensive pigmentation in the male. (e) Male and female gonad. A, accessory gland; MOL, muscle of Lawrence; SC, sex comb; T, testis.
In this case, maternally contributed emc may act by binding with the other HLH-containing proteins to form nonfunctional heterodimers. Recently, another protein which interacts with hairy-related bHLH proteins has been identified (Paroush et al., 1994). The protein encoded by groupcho, has been shown to interact with dpn and may act as a transcriptional corepressor in conjunction with other bHLH proteins. The hermaphrodite (her) locus has multiple roles during sex determination. Maternally contributed HER appears to act as a positive regulator of Sxl activation and also affects the process of dosage compensation. The zygotic function of her is not rescued by the constitutive expression of either Sxl or tra. dsx splicing is unaffected in interssexual flies resulting from her zygotic mutants. This implies that the zygotic function of her may be similar to that of intersex (ix), acting in parallel with or downstream of doublesex (Pultz et al., 1994; Pultz and Baker, 1995). her has recently been characterised at the molecular level. It encodes a zinc finger protein which may function as a transcription factor (Ryner and Swain, 1995). These proteins activate Sex-lethal at the level of transcription (Keyes et al., 1992). The various bHLH proteins are able to interact to produce homo- or heterodimers which then bind to DNA, activating transcription. The ability of these proteins to form heterodimers in Sxl activation was demonstrated by the inappropriate expression of another bHLH protein, the pair-rule gene hairy (Parkhurst et al., 1990). When hairy protein is expressed ectopically under the control of the hunchback promoter, it is a female lethal. It was shown that HAIRY interacts with the other HLH proteins to form heterodimers which could not bind to DNA, preventing Sxl from being activated in females. The gene sans fille (snf) [known also as liz (Steinmann-Zwicky, 1988) and fs(1)1621] represents a maternal effect gene that is required for activation of Sxl in both the germline and the soma. snf has been characterised at the molecular level and shows significant sequence and functional homology with the U1A snRNP protein (Flickinger and Salz, 1994). The virilizer locus has also been implicated in the process of sex determination (Hilfiker and Nöthiger, 1991). However, it is not yet clear where in the hierarchy vir acts. It appears to function upstream of tra to modulate Sxl activity but whether it acts directly on tra, Sxl, or both has not been established.

It is interesting to note that several of the genes involved in the assessment of the X:A ratio also have a function later in development, during neurogenesis. The maternal product of daughterless, for example, is required during sex determination; the zygotic function is required later for development of the peripheral nervous system (Caudy et al., 1988). sis-b [a member of the achaete/scute complex], deadpan [proneural gene], and extramacrochaetae [neural gene] are also involved in both of these processes. These bHLH proteins may interact to constitute a genetic switch, such that the ratio of positive [numerator] to negative [denominator] regulators determines a cell's fate by affecting the transcriptional activity of downstream genes.

**Sex-lethal**

Sex-lethal plays a pivotal role in the processes of both somatic and germline sex determination as well as in the process of dosage compensation [Fig. 2] (Baker, 1989; Cline, 1993). The structure of Sex-lethal is complex, with 10 exons and two promoter regions dispersed over a region of 25 kb (Samuels et al., 1991) [Fig. 3]. The use of the different promoters, different exons, and different polyadenylation sites leads to the production of at least 10 different RNA species, with varying patterns of expression. Three transcripts are specific to the male (4.3, 3.3, and 2.1 kb) and four to the female (4.1, two transcripts of 3.1, and 1.9 kb). One of the 3.1 kb transcripts and the 1.9-kb transcript are probably germline-dependent, since their levels of expression are reduced in abdomens that contain no ovaries. In addition, the initial activation of Sex-lethal results in the production of early transcripts in the female embryo [Fig. 3]. These transcripts are derived from the early promoter [PE] in response to the X:A signal [X:A = 1]. Alternate splicing and the use of different polyadenylation signals gives rise to three transcripts of 3.7, 2.6, and 1.6 kb. In situ hybridisation to whole-mount embryos indicates that these early transcripts are present in the embryo prior to pole-cell formation [Keyes et al., 1992] [Fig. 4]. The signal peaks in embryos at about nuclear division 12 and begins to decline until germ band extension, when it can no longer be detected. Neither early transcripts nor protein are detected in the pole cells [these are the germline primordia, Fig. 4]. Not surprisingly, these transcripts are not present in flies mutant for da. The early proteins have been suggested to act in establishing the positive feedback loop for Sex-lethal autoregulation as they are present in the embryo at the time when the early functions of Sex-lethal occur (Salz et al., 1989).

The next step in the regulation of Sex-lethal occurs at the level of RNA splicing. The main difference between the male transcripts and those of the female is in the incorporation of exon 3 [male-specific exon] in the male. This exon contains several translational stop codons, resulting in the formation of a truncated protein in the male. In the female, XL protein translated from the early transcripts directs the splicing pattern of the later transcripts such that the male-specific exon is spliced out, enabling a full-length protein to be produced. These proteins maintain the productive mode of splicing.

Sequence analysis of the female cDNAs indicates that they contain a long open reading frame (ORF) which extends from exon 2 to at least exon 8 [see Fig. 3]. This gives a predicted protein product of approximately 354 amino acids [Bell et al., 1988; Samuels et al., 1991]. The sequence shows two conserved domains, RMM1 and RMM2, which show significant sequence homology to a conserved RNA binding domain found in other RNA binding proteins [RNP] [Bell et al., 1988]. This family of proteins is able to bind both RNA and single-stranded DNA and functions by binding to various RNA species [including its own] to direct their
SEX DETERMINATION IN FEMALE SOMATIC CELLS

- **Courtship behaviour**
- **Correct nerve innervation to repress male muscle**
- **Promotes YP expression in females**
- **Cell divisions in neurons**
- **Continuous repression of male genital primordia**

Female differentiation and behaviour

SEX DETERMINATION IN MALE SOMATIC CELLS

- **Repression of YPs in adult fat body**
- **Continuous repression of female genital primordia**
- **Promotes male development in leg disc?**

Male differentiation and behaviour

**FIG. 2.** Diagrams of sex determination in female and male somatic cells. X:A, X chromosome to autosome ratio; SXL, Sex-lethal transcript is spliced in female mode and makes a functional protein; TRA, transformer transcript is spliced in female mode in the presence of SXL protein and encodes a functional protein; TRA-2, transformer-2 product is required; DSX, doublesex transcript is spliced in female mode in the presence of TRA and TRA-2 protein and makes a female-specific protein which regulates the expression of downstream genes; DSX\textsuperscript{M}, in the absence of TRA protein, doublesex transcript is spliced in the male mode to produce a male-specific protein that regulates the expression of downstream genes; IX, intersex product interacts with DSX\textsuperscript{F} to regulate downstream genes in females and affects courtship behaviour in males; HER\textsubscript{ZYG}, the zygotic product of the hermaphrodite (her) gene is required for female differentiation independently of DSX. Arrows represent interactions shown experimentally. Arrows with question marks demonstrate possible interactions. *tra* does not induce germ cells but is required for differentiation of female germ cells.
**FIG. 3.** The structure of the *Sex-lethal* transcription unit. The positions of the two promoters (early $P_E$ and late $P_L$) are indicated. In the female, early proteins from $P_E$ direct the splicing of the late transcripts from $P_L$ into the female mode. Different forms of the protein are produced by the use of different splice and polyadenylation sites. These proteins all contain the ribonucleoprotein (RNP) binding domain that is required for Sxl function. In the male, where no early transcripts are produced, no splicing of the late transcripts occurs. Consequently the third exon (which contains stop codons) is included, producing a truncated protein that does not contain the RNP domain.

Splicing pattern (Samuels *et al.*, 1994). The mechanism by which this sex-specific splicing is achieved will be discussed in the section dealing with the *tra* and *tra-2* genes.

As expected, the presence of SXL protein differs between the two sexes. It is present in female embryos and absent in male embryos (Bopp *et al.*, 1991) (see Fig. 4). In addition, the open reading frames differ among the different classes of female-specific RNAs, depending on the different splice sites and polyadenylation signals. Most of these differences occur at the carboxy terminus of the protein and all of the

**FIG. 4.** Developmental distribution of SXL protein. (a) In the embryo [i] SXL is found in all the syncytial nuclei except those which form the pole cells. At cellular blastoderm [ii], SXL remains localised to the nuclei. During the later stages of development, SXL can be detected in probably all somatic tissue such as third instar larval salivary gland nuclei [iii]. (b) Embryos stained using anti-SXL antibody. Antibody is stained pink and so female embryos appear pink and male embryos remain blue. No SXL protein can be detected in the pole cells (white arrow). (c) SXL protein can also be detected in male and female adults. In the female, two major species of protein (36 and 38 kDa) can be detected in all tissues. Additionally, two minor species of 40 and 42 kDa can be detected in the ovary. In the male, two smaller species of 33 and 35 kDa can be detected in the head and thorax. The functional significance of these protein variants is not known.
a) No SXL in germ cell nuclei

SYNCYTIAL BLASTODERM

No SXL in germ cells

CELLULAR BLASTODERM

SXL protein preferentially localised in nuclei of somatic cells in cellular blastoderm

SXL protein found associated with chromosomal intro sin

(b) MALE

FEMALE

SXL Proteins of Different Sizes:
- 42 kD
- 40 kD
- 38 kD
- 36 kD
- 35 kD
- 33 kD

Ovary
predicted female products contain both RNP domains. The distribution of the SXL protein was assessed using monoclonal antibodies raised against sequences present downstream of the male-specific exon. Two prominent proteins (38 and 36 kDa) were detected in the adult female but not in the male. These two variants derive from alternate utilisation of the 3'splice site of exon 5. There was a differential distribution of these proteins within the fly. Both these proteins could be detected in the ovaries and head. In dissected carcasses containing no ovaries, the 38-kDa protein was the major species, with lower levels of the 36-kDa protein present. The predominant species in the thorax was the 38-kDa protein. In addition to these two species, two other minor forms of the protein (40 and 42 kDa) could be detected in ovaries and early embryos (Bopp et al., 1991).

Surprisingly, SXL protein could also be detected in the adult male. Two species of 35 and 33 kDa are detected in the thorax and head; these are more prominent in the head than in the thorax and cannot be detected in the abdomen. Protein levels are 20-40 times lower than those detected in females. These proteins have been demonstrated to be derived from the Sxl locus, since they are absent in males carrying deletions for Sxl. They are similar to the low-abundance SXL proteins, which are also detected in females. It is possible that these proteins are produced from translational initiation codons downstream of exon 3. However, they do not appear to function in splicing (Bopp et al., 1991), since no female-like Sxl or tra transcripts are detected in males.

SXL protein has been shown to be preferentially localised to the nucleus. In the interphase nucleus during embryogenesis, SXL protein can be detected as regions of intense staining superimposed on more diffusely staining nucleoplasm, suggesting that high levels of SXL protein have been accumulated (Bopp et al., 1991).

**Sex-lethal and Germline Sex Determination**

In contrast to the process of sex determination in the soma, the mechanism by which sex determination occurs in the germline is still unclear. The known factors are shown in Fig. 5. As described previously, the **Sex-lethal** early transcripts [and protein] are not present in the germline primordia (Bopp et al., 1991; Keyes et al., 1992). In addition, the numerator elements **sis-a**, **sis-b**, and **runt** have been shown by pole-cell transplantation experiments not to be essential for **Sxl** activation in the germline. Pole cells from embryos mutant for all three of these loci have been shown to produce functional germ cells when transplanted to a wild-type background (Granadino et al., 1993; Steinmann-Zwicky, 1994). In contrast to the somatic process, the genes **tra**, **tra-2**, and **dsx** have no function in female germline development (Marsh and Wieschaus, 1978; Schüpbach, 1982). **tra-2**, however, is required in males for the production of motile sperm (Fuller, 1993). Germ cells that are 1X:2A [i.e., male] attempt to develop as spermatocytes irrespective of the sex of the surrounding somatic tissue. In a female somatic background, these germ cells assume an intersexual identity. **Sxl** activation in the germline, in common with the process in the soma, is cell autonomous and is essential for female development. However, 2X:2A germ cells (i.e., female) also require an inductive signal to start and/or complete sexual development, which depends upon the phenotypic sex of the surrounding soma. Consequently, intersexual cells are produced in a male somatic background while oocytes develop in a female background. Any spermatocytes which are produced arrest in the primary spermatocyte stage. This suggests [at least in the female] that inductive interactions are required between the soma and the germline to direct the correct sex determination of the germ cells (Nöthiger et al., 1989; Steinmann-Zwicky, 1992). This signal is established during embryogenesis after somatic sex determination has occurred and therefore the correct expression of the downstream sex-determining genes is required. The inductive interactions may be mediated through the action of **tra-2**. XX flies mutant for both **dsx** and **tra** express significant levels of female **Sxl** activity. In **tra-2** mutant XX individuals, **Sxl** is expressed in the male mode. Since **tra-2** is not expressed sex-specifically, it has been postulated that it interacts with an unknown gene (which may be under the control of **Sxl**) to control the inductive signal from the soma to the germline (D. Bopp, personal communication). This inductive signal has been postulated to act in one of two ways: either 2X germ cells undergo male development unless they receive an inductive signal from the female soma resulting in **Sxl** activation, thus promoting female development; or 2X germ cells undergo oogenesis unless they receive an inducing signal from the male soma to repress **Sxl**, allowing male development to occur.

Several loci have been isolated which appear to function during germline sex determination. Mutations in the genes **snf** (Salz, 1992), **fl(2)id** (Granadino et al., 1990), and some alleles of the **ovarian tumour (otu)** locus (Pauli et al., 1993) result in the formation of multicellular cysts in the ovary. This phenotype is similar to that observed with several **Sxl** germline-specific mutations. **SXL** protein is absent in the ovaries of flies mutant for **snf** or **otu** [genes (Bopp et al., 1993). As described previously, the **snf** locus encodes a protein which exhibits homology to the U1A snRNP protein. It is therefore likely that the **SNF** protein acts to establish the correct splicing of **Sxl** transcripts in both the soma and the germline (Bopp et al., 1993; Oliver et al., 1993; Flickinger and Salz, 1994). Although **otu** has been cloned and sequenced, no significant homology to other sequences has been obtained. It has been postulated that these genes may act in establishing **Sxl** autoregulation. Mutations in the gene **bag-of-marbles** (McKearin and Spradling, 1990) result in sterility in both males and females, with production of undifferentiated cysts observed in both the testis and the ovary. The germ cells in these cysts are morphologically similar to germline stem cells, gonial cells or, in some cases, spermatocytes. Ovaries from these flies have normal levels of **SXL** but the localisation of the protein is perturbed (Bopp et al., 1993). Unlike the previous classes of mutants, mutations at the **ovo** locus are still able to produce rudimentary...
RNA Splicing in Sex Determination

FIG. 5. A schematic representation of the process of germline sex determination. In the female, the process of germline sex determination differs from the process in the soma. Inductive interactions with the soma (mediated by the genes of the somatic sex determination hierarchy) are required, as well as a cell autonomous signal mediated by SXL. The signal from the X:A ratio activates SXL. This, in conjunction with the OVO and ovarian tumour (OTU) protein and the inductive interactions from the soma, determines the sex of the germline. In the male, there does not appear to be a repressive inductive interaction with the soma. Germline sex determination appears to be partly cell autonomous. DSX⁺, doublesex female-specific protein; DSX⁻, doublesex male-specific protein; OTU, ovarian tumour protein; OVO, ovo protein; SNF, Sans file protein; TRA, transformer protein; TRA-2, transformer-2 protein.

Sex-lethal and Dosage Compensation

In some organisms in which one sex is heterogametic, the process of dosage compensation is essential to compensate for the functional aneuploidy that exists in the heterogametic sex. This dosage compensation can be achieved in several ways. One of the X-chromosomes may be inactivated; the 2X-chromosomes may be transcribed at a lower rate than the single X-chromosome; or the single X-chromosome may be transcribed at a faster rate than the 2X-chromosomes. In Drosophila, in which the male is the heterogametic sex, dosage compensation is achieved by hypertranscription of the single male X-chromosome (Mukherjee and Beermann, 1965). The incorporation of uridine into transcripts derived from the male X-chromosome relative to the autosomes was shown to be substantially higher than that of the female X-chromosome. This ensures the equalisation of levels of gene products in the male and female.

Sex-lethal and Male-Specific Lethals

Trans-acting regulators of dosage compensation have been identified. These comprise a group of four autosomal loci known collectively as the male-specific lethals. These are the male-specific lethal-1 (msl-1), male-specific lethal-2 (msl-2) (Belote and Lucchesi, 1980a), male-specific lethal-3 (msl-3) (Lucchesi et al., 1982), and maleless (mle) loci (Belote and Lucchesi, 1980b). As their name suggests, mutations in these genes are lethal to males but not females. It

Copyright © 1995 by Academic Press, Inc. All rights of reproduction in any form reserved.
FIG. 8. The segmental origins of the imaginal disc primordia are shown. These three groups of primordial cells fuse to form the discs of the larvae and differentiate into the male and female genitalia and analia. Blue shows the derivatives of the female primordial cells, orange the male primordial cells, and yellow the analia which differentiate into different structures in male and female flies. The female primordial cells do not grow or differentiate in males and the male primordial cells do not grow or differentiate in females. They do, however, remain an integral but small population of cells within the genital disc.
**FIG. 9.** The role of DSX in yp regulation. In the embryo DSX directs the decision as to whether the gonad will develop along a male or female developmental pathway and maintains this in the larva. In the pupa the ovary or testis differentiates and then, under the control of unknown tissue-specific factors, the yp genes are repressed in the testis but expressed in specific follicle cells at defined stages of oogenesis. The expression of dsx is unknown during fat body development. The adult fat body differentiates during metamorphosis. DSX then directs repression of the yp genes in the male fat body. Expression requires fat body-specific factors. Using a lacZ reporter the DNA sequences flanking the yp genes which direct this sex- and tissue-specific expression have been identified. Examples are shown for yp3. A fragment from -285 to +49 bp directs expression in the ovary but not the fat body. A fragment from -704 to -285 directs female fat body expression. This same construct is not expressed in the male fat body. These regions, either independently or, as shown in the figure, together, show no expression in the testis. fb, fat body; ov, ovary, hsp 70-lacZ, heat-shock protein-70 gluc promoter fused to a lacZ reporter gene; DSXF, doublesex female protein; DSXM, doublesex male protein.
has been postulated that these genes function in a common pathway, since mutations in one of these loci is just as detrimental as mutations in more than one. Mutations in mle, msl-1, and msl-2 result in a 50–60% reduction in the levels of gene transcription of the X-chromosome evident in wild-type males (Baker et al., 1994).

The mle, msl-1, and msl-3 genes have been characterised at the molecular level. Proteins from these genes are expressed in both males and females; however, their pattern of distribution differs greatly between the sexes (Baker et al., 1994). They are seen to associate with many sites on the X-chromosome, but are not present on the female X-chromosome. This association leads to a significant increase in the presence of acetylated histone H4. This observation suggests that these proteins act directly to regulate the process of dosage compensation and, in addition, that all these proteins may regulate the same set of genes. mle, in contrast to the rest of the msl proteins, is able to associate with other chromosomal locations without the interactions of the other loci. It is however unclear whether this additional binding to the chromosomes is part of dosage compensation or whether it represents another function. mle is allelic to nap, a gene which affects the activity of sodium channels (Kernan et al., 1991).

Regulation by Sex-lethal

The function of Sxl in dosage compensation was inferred by the reciprocal male and female lethal phenotypes of gain of function and loss of function mutations. The immediate target of Sxl during dosage compensation is msl-2 (Zhou et al., 1995). Sxl functions by preventing the splicing out of a female-specific leader sequence in the msl-2 transcripts; consequently no MSL-2 protein is produced. In the male, this sequence is spliced out, resulting in the production of functional MSL-2 protein (Zhou et al., 1995). In flies which are mosaic for Sxl expression, MSL proteins were only associated with the X-chromosome in cells which were not expressing Sxl. mle produces multiple transcripts during development (Kuroda et al., 1991), msl-1 codes for three transcripts (Palmer et al., 1994), and msl-3 has at least three transcripts (Baker et al., 1994). All of these appear to be equivalent in both males and females, showing that the regulation of these transcripts by Sxl is not direct. The msl-2 4-kb transcript can also be detected in both males and females. However, it appears that this transcript is more abundant in males than in females. Also, the female transcript appears to be slightly larger than that of the male. As described previously, this is due to the presence of an untranslated leader sequence found in the female transcript but spliced out in males. The predicted MSL-2 protein contains both a RING finger motif and a metallothionein-like domain. This protein is absent from females. MSL-2 protein has also been demonstrated to bind to the X-chromosome at the same sites as MLE and MSL-1. As discussed previously, regulation of msl-2 may be mediated by the female-specific leader sequence. The 5’ and 3’ ends of this intron contain stretches of thymine residues similar to those found in the Sxl consensus binding site (Zhou et al., 1995).

In polytene chromosome squashes the male X-chromosome is more open and diffuse than that of the female. It has been proposed that this altered chromatin configuration is important in allowing hypertranscription to occur. Sequence analysis of the mle locus has shown that it exhibits similarity with both RNA helicase A and the DEAH RNA helicases (Kuroda et al., 1991). It has been proposed that mle forms a complex with the other msl proteins (similar to the spliceosome complex), facilitating hypertranscription by either increasing the rate of elongation of transcripts or removing RNA from the transcription start sites.

Other mechanisms of dosage compensation have been proposed to exist. It appears that the proteins from the Sxl early promoter (which, as discussed earlier, directs the autosplicing of Sxl transcripts) may function in directing early stages of dosage compensation which are not regulated by the msl loci (Gergen and Wieschaus, 1986; Bernstein and Cline, 1994). Females which are homozygous for all four mlo loci and a null allele of Sxl still die. If this lethality was due to inappropriate activation of the msl loci, then it would be expected that the additional mutations in the msl loci would suppress this lethality by preventing hypertranscription. This suggests that Sxl may act on loci other than the known msls to direct dosage compensation. The X-linked gene runt is an example of a gene which appears to be regulated by the action of Sxl and not the msls. Therefore loci may exist which are also regulated by Sxl and act on a different set of genes. It is therefore interesting to note that alleles of Sxl exist that affect dosage compensation later in development which are not suppressed by mutations in mle or msl-1. As described previously, it has been suggested that all four msl act on the same targets. Consequently it is possible that other dosage compensation loci exist which act on a different set of target genes.

Autoregulation of Sxl

A number of uridine runs have been identified in the introns both upstream and downstream of the Sxl male-specific exon (exon 3) and have been implicated in SXL regulation of the female-specific splicing event (Sakamoto et al., 1992; Horabin and Schedl, 1993a,b; Samuels et al., 1994; Wang and Bell, 1994). Sakamoto et al. (1992) used a cell culture transient expression system to show that deletion of several of these U-rich motifs disrupts SXL regulation. It was observed that regulation could be restored by replacing the deleted sequences with synthetic oligonucleotides. In the first real binding study of SXL, Samuels et al. (1994) used gel-shifts, footprinting, and UV cross-linking with purified SXL protein to demonstrate directly that SXL binds to poly-(U) runs in RNA. Both these studies suggested that SXL may bind cooperatively to adjacent poly-(U) motifs and cooperativity has been subsequently demonstrated by Wang and Bell (1994). This study showed that SXL protein binds to many sites around the male-specific exon and that,
when bound, the proteins interact cooperatively via their N-termini. An increase in cooperativity was observed when longer RNA molecules with multiple U-rich motifs were used as a binding substrate. The authors suggest that the N-terminus of SXL may interact directly with other splicing regulatory proteins. In support of this, the hnRNP A1 has an N-terminus with a similar amino acid constituency to SXL and interacts with the splicing factor SF2/ASF.

Progress has recently been made in the elucidation of the type of mechanism used by SXL to prevent the inclusion of exon 3 in the processed female transcript. Horabin and Schedl (1993a,b) have used germline transformation to introduce altered SxI minigene constructs into flies. Using RT-PCR to analyse the spliced RNA products of these mutant minigenes, they were able to ask very specific questions about what sequences around the male-specific exon are actually required for SXL regulation. Their findings have further demonstrated the importance of the poly-(U) runs both upstream and downstream of exon 3. Interestingly, it was shown that deletion of the five poly-(U) runs in the downstream intron disrupted SXL regulation much more drastically than deletion of the U runs in the upstream intron. This suggests that the critical step in preventing the inclusion of exon 3 is the blockage of the downstream intron’s 5’ splice site. This is supported by the observation that when the exon 3 5’ splice site is deleted, all product is spliced in the male mode. Horabin and Schedl showed that when the exon 3 5’ splice site is deleted, all product is spliced in the male mode.

**Sxl Regulation of tra Splicing**

As well as modulating the splicing of its own transcript, SXL protein also regulates the splicing of the primary transcript from the gene transformer (tra). Cloning and characterisation of the tra gene has shown that, in female flies only, the choice of a downstream splice acceptor site prevents the inclusion of a translational stop codon. This facilitates the production of the active 211-aa TRA protein (Butler et al., 1986; McKeown et al., 1987, 1988; Boggs et al., 1987—see Fig. 6). The observation that the area around the non-sex-specific splice site of the tra gene contains a uridine octamer sequence originally suggested that tra may be directly under the control of SXL. Transformation experiments have subsequently shown that SXL is responsible for the sex-specific splicing of tra nascent RNA and that the splice acceptor site containing this uridine octamer is required for this regulation (Sosnowski et al., 1989). This was shown by introducing various constructs containing tra genomic DNA (transcribed by the hsp70 promoter) into a tra background via germline transformation. Deletion of the non-sex-specific splice site led to a degree of Sxl-independent feminisation of the male, as would be expected if the function of SXL in the female is to prevent the use of this site. Deletion of the sex-specific splice site resulted in accumulation of unspliced RNA in females, a lack of female-specific RNA, and an inability of this construct to either rescue tra females or transform males.

SXL blocks the non-sex-specific splice site by antagonising the essential splicing factor U2AF which binds to the same U-rich sequences as the SXL protein. SXL, however, lacks the arginine-serine "RS" repeat which is present in U2AF (Zamore et al., 1992, Zhang et al., 1992) as well as other splicing factors such as SF2/ASF (Ge et al., 1991; Krainer et al., 1991), SC35 (Fu and Maniatis, 1992), SRp20, SRp75 (Zahler et al., 1992), and suppressor of white apricot (Sn[W]; Chou et al., 1987). The 70K U1 snRNP also contains these repeats (Theisen et al., 1986; Spritz et al., 1987). If the RS motif is introduced into the SXL protein, it becomes constitutively active as a splicing factor, causing splicing from the same splice site which it normally blocks (Valcárcel et al., 1993).

The predicted protein product from the female-specific tra transcript also contains an RS motif (Boggs et al., 1987), indicating that the function of TRA may be to modulate splicing in the female. Su[W] protein can be rendered non-functional by deletion of its RS motif and when this is replaced with the TRA RS motif this function is restored, indicating that this motif may play a similar role in both proteins (Li and Bingham, 1991). Deletion of the Su[W] RS motif seemed to affect the nuclear localisation of the protein, which may suggest a possible role for this region. However, the U2AF RS motif was also shown to be essential for its in vitro splicing activity (Zamore et al., 1992), suggesting that this motif may have more than one function.

**Regulation of tra-2 Splicing**

The cloning of transformer-2 (tra-2) has revealed that it encodes four transcripts which are alternately spliced (Amrine et al., 1988; Goralski et al., 1989, Mattax et al., 1990, Amrine et al., 1990), as shown in Fig. 7. These transcripts potentially encode proteins with a common C terminus containing both the RS motif and the 80- to 90-amino-acid RNP motif, which is also found in SF2/ASF, SC35, SRp20, SRp75, and U2AF65, as well as in the U1A and U2B’ snRNPs (reviewed in Kenan et al., 1991), indicating that tra-2 may also encode a splicing regulator.

**tra-2 Function in the Male Germine**

TRA-2 protein is required in the male germline, as shown by the fact that nonfunctional sperm are produced in flies mutant for this gene (Belote and Baker, 1983). Indeed, most abundant expression of tra-2 is seen in this tissue where two male germline-specific transcripts are produced. In wild-type flies, the splicing of the M1 intron (shown in Fig. 7) is normally an inefficient process such that the concentration of the M1-containing transcript is higher than the
(a) SxI Primary RNA

1 2 3 4 5 6 7 8 9 10

→ SxI Processed RNA

1 2 3 4 5 6 7 8 9 10

→ tra Primary RNA

1 2 3

→ tra Processed RNA

1 2 3

→ dsx Primary RNA

1 2 3 4 5 6

→ dsx Processed RNA

1 2 3 4 5 6

COURTSHIP REPRESSION OF MUSCLE

EXUPERANTIA

(b) SxI Primary RNA

1 2 3 4 5 6 7 8 9 10

→ SxI Processed RNA

1 2 3 4 5 6 7 8 9 10

→ tra Primary RNA

1 2 3

→ tra Processed RNA

1 2 3

→ dsx Primary RNA

1 2 3 4 5 6

→ dsx Processed RNA

1 2 3 4 5 6

SPERMATOGENESIS
**tra-2 Regulates the Splicing of exu**

It has been shown that one function of tra-2 in the male germline is to regulate the production of male-specific transcripts from the gene exuperantia (exu). The exu gene has functions in both the male and the female germline (Hazeldrig et al., 1990). In females, exu has a maternal effect, regulating the localisation of the bicoid gene product in the oocyte. In males, exu is required in the germline for correct spermatogenesis, since mutation of the gene results in male sterility. Two sex-specific and germ-line-dependent exu transcripts have been identified; a 2.9-kb transcript which is male germline-specific and a 2.1-kb transcript which is female germline-specific. These transcripts appear to be initiated from different promoters and exhibit sex-specific processing of the 3' UTR region (Hazeldrig and Tu, 1994). In tra-2 mutants' production of the exu male-specific transcript is much less efficient, although not totally abolished. Male-specific 3' UTR sequences are important for spermatogenic exu function since deletions in this region result in male sterility.

**Somatic Functions of tra-2**

Although tra-2 is clearly important for regulating male germline sexual differentiation, no somatic function has been demonstrated for tra-2 in the male. In the female, however, tra-2 acts in concert with tra to direct most aspects of female-specific differentiation (see Slee and Bownes, 1990, for review). Mutant alleles of tra or tra-2 result in transformation of females into pseudomales exhibiting male characteristics. These include male pigmentation, cuticular structures, and rudimentary testes. They are, however, of female size and are infertile due to nonfunctional sperm. Mutations in the tra gene have no effect on males, and tra-2 mutants show no male somatic transformations. Null alleles of the gene dsx result in intersexuality of both males and females, with the phenotype seeming to result from an expression of the genes responsible for both...
male-specific and female-specific development at the individual cell level.

The epistatic relationship among tra, tra-2, and dsx was first shown by the construction of flies which carried double homozygous mutations in various combinations (Baker and Ridge, 1980). The epistatic gene could then be identified by virtue of its phenotype being manifest in the fly. This showed that dsx is epistatic to tra. In another set of experiments tra expression was shown to be unable to cause development along the female pathway in flies mutant for tra-2 and dsx. In addition, molecular evidence shows that tra and tra-2 are required for production of female-specific dsx transcripts (Nagoshi et al., 1988). The above evidence shows that dsx is epistatic to tra and tra-2 and that tra and tra-2 are required for dsx to be expressed in the female mode, while the male functions of dsx are independent of tra and tra-2.

**Regulation of dsx Splicing**

The genetic evidence suggests that dsx is differentially active in both males and females, acting primarily to repress genes required for differentiation of the opposite sex. Analysis of the dsx transcripts showed how tra and tra-2 enable this to occur [Baker and Wolner, 1988; Burtis and Baker, 1989]. Examination of cDNAs representing the 3.9-kb (male-specific) and 3.5-kb (female-specific) dsx transcripts shows that these messages are differentially spliced and polyadenylated but are both capable of producing large, functional proteins with sex-specific carboxyl termini. Hence it would be quite feasible for there to be differential activity in both sexes.

Germline transformants containing the female-specific tra cDNA fused to the hsp70 promoter have their male soma transformed to female soma. This transformation is correlated with the production of the female-specific dsx transcripts in the male soma, while tra-2, XX flies produce only male dsx mRNA [McKown et al., 1988]. tra-2 is also required for production of the female-specific dsx transcripts but not for production of male-specific dsx transcripts [Nagoshi et al., 1988]. Tissue culture cotransfection experiments [Hoshijima et al., 1991; Ryner and Baker, 1991] allow the effects of TRA and TRA-2 proteins upon dsx pre-mRNA to be assessed directly by analysing the spliced products of dsx pre-mRNA in the presence or absence of TRA and TRA-2. These studies show that TRA and TRA-2 act in concert to positively promote the usage of the female-specific splice acceptor site, as shown in Fig. 6.

A region lying just downstream of the exon 4 acceptor site has been implicated as being involved in tra and tra-2 regulation of the dsx pre-mRNA [Nagoshi and Baker, 1990]. Lying in this region are six 13-nt repeats [dsx repeat element dsxRE], the deletion of which results in a loss of female-specific product in the cotransfection system described above [Hoshijima et al., 1991; Ryner and Baker, 1991]. It has been shown that TRA, TRA-2, and some SR proteins bind to the dsxRE in vitro [as shown in Fig. 6], with TRA-2 binding being dependent upon a purine rich enhancer (PRE) element present within the dsxRE [Hedley and Maniatis, 1991; Lynch and Maniatis, 1995]. The same studies showed that the dsxRE and PRE elements act synergistically. Also, substitution of some of the noncanonical purines present in the polypyrimidine stretch of the female-specific acceptor site causes female-specific splicing independent of TRA and TRA-2 [Hoshijima et al., 1991]. This indicates that this splice site is not used in males because of its non-standard polypyrimidine stretch. In females, TRA and TRA-2 act to stabilise the splicing apparatus at this site and thus promote its use, i.e., default splicing occurs in males, while in females regulation of female-specific splicing occurs by TRA and TRA-2 promoting the use of the nonpreferred site. Indeed, it has been demonstrated that TRA and TRA-2 work by attracting general splicing factors, including some SR proteins, to the dsxRE region, enabling it to function as a splicing enhancer (Tian and Maniatis, 1993).

Sequences homologous to the dsxRE 13-nt repeats have been identified in both exu and tra-2 transcripts. Since, in the male germline, alternative processing of both these transcripts is under the control of tra-2, it would be expected that mutations of the repeat sequences would result in loss of tra-2 regulation. However, mutational analysis of some of these sites has not as yet been able to establish a role for them in sex-specific splicing regulation [W. Mattox and T. Hazelrigg, personal communication].

The RNP motif of TRA-2 has been shown to be essential for its somatic and male germline functions, although it is required but not sufficient to direct RNA binding in vitro (Amrein et al., 1994). Experiments using the yeast dihybrid assay have shown that TRA and TRA-2 physically interact with themselves, with each other, and with the general splicing factor SF2. These three proteins have been shown to be sufficient to cause dsx primary transcripts to be spliced in a female-specific manner. One of the TRA-2 RS motifs, RS2, is essential for in vivo function and for interactions in the dihybrid assay. The RS1 RS motif of TRA-2 is not essential but, if deleted, results in temperature-sensitive mutation in vivo and decreased sensitivity of dihybrid interactions in vitro and so may act as a stabiliser of protein–protein interaction (Amrein et al., 1994).

In addition to promoting the usage of certain splice sites, TRA-2 also appears to be able to prevent the usage of splice sites. This is shown by the fact that tra-2 downregulates the removal of the M1 intron from tra-2 primary transcripts in the male germline [Mattox and Baker, 1991]. This may be a result of the proximity of the splice site relative to the TRA-2 binding site, which may cause bound TRA-2 to interfere stearically with the splicing apparatus. This repressive function may alternatively be due to additional tissue-specific factors which modulate TRA-2 function.

**Additional Functions of tra**

For most aspects of somatic sexual differentiation, dsx is the last member in the hierarchy of regulatory genes. This
view, however, in which the only sex determination function of \( \text{tra} \) and \( \text{tra-2} \) is to direct the \( \text{dsx} \) primary transcript to be spliced in the female mode, needs to be revised. Fresh evidence indicates that \( \text{tra} \) and \( \text{tra-2} \) also govern \( \text{dsx} \)-independent pathways of sex-specific differentiation.

Male flies have a sex-specific pair of muscles known as the muscle of Lawrence (MOL) \cite{Lawrence1984}. These muscles span the fifth abdominal segment and were initially thought to be involved in the curling of the male abdomen during copulation. More recently, however, it has been found that flies lacking the MOL can still copulate \cite{Gailey1991}. \( \text{XX} \) flies carrying null mutations of \( \text{tra or tra-2} \) develop as pseudomales which have this muscle present \cite{Taylor1992}. However, this effect of the \( \text{tra} \) and \( \text{tra-2} \) alleles cannot be due to their function in \( \text{dsx} \) regulation, as the muscle is present in \( \text{XY} \) individuals mutant for \( \text{dsx} \) but is absent in \( \text{XX} \) \( \text{dsx} \) mutant flies. Thus, the repressive function that active \( \text{TRA} \) and \( \text{TRA-2} \) proteins exert on the development of this muscle must act via a pathway that is independent of \( \text{dsx} \). Transplantation of nuclei between males and females has shown that the identity of this muscle is not autonomous but depends upon the sex of the innervating axons \cite{Lawrence1986}.

Recent studies on courtship behaviour have also pointed towards the presence of at least one branch of regulatory genes which are governed by \( \text{tra} \) and \( \text{tra-2} \) but not by \( \text{dsx} \). In wild-type flies, mating involves a number of male-specific courtship behaviours which are readily observable \cite{Spieth1974,Ehrman1978}. Courtship is initiated with the male tapping the female with his forelegs, orienting toward her, and following her. He then begins a courtship song by extending one wing and vibrating it. This is followed by the male extending his proboscis and licking the female’s genitalia and finally copulation is attempted. Females are largely sedentary during mating, although reception and rejection behaviours are observed.

Temperature shift experiments using a temperature-sensitive allele of \( \text{tra-2} \) have shown that absence of male behaviour in the female is dependent upon activity of the \( \text{tra-2} \) gene, such that inactivation of this gene in females from the late larval stage onwards results in appearance of male courtship elements \cite{Belote1987}. Flies of the genotype \( \text{XX} \),\( \text{dsx} \) exhibit no male courtship behaviour suggesting that \( \text{DSX}^\text{Y} \) does not normally act to repress male courtship behaviour in the female. Similarly \( \text{XX} \),\( \text{dsx}^\text{D} \) flies which constitutively express \( \text{DSX}^\text{M} \) do not attempt to court despite the fact that they are male in morphology. This indicates that \( \text{DSX}^\text{M} \) does not activate male courtship in the male \cite{Taylor1994}. These experiments argue that there is no role for \( \text{dsx} \) in regulating courtship behaviour. However, it has been observed that \( \text{XY} \),\( \text{dsx}^\text{D} \) flies court much less than normal males. These flies also elicit more courtship than wild-type adult males, even though such flies do not produce characteristically female feromones \cite{McRoberts1985,Jallon1988}. At first glance, this would indicate a direct role for \( \text{dsx} \) in regulating this behaviour. Taylor et al. \cite{Taylor1994} observed, however, that although a large proportion of \( \text{XY} \),\( \text{dsx}^\text{D} \) flies did not court at all, at least one fly carrying each tested allele exhibited male courtship behaviours up to and including wing extension. The lack of attempted copulation of \( \text{XY} \),\( \text{dsx} \) flies may be explained by their morphology, which makes this a physical impossibility. This led Taylor et al. \cite{Taylor1994} to propose that the anomalous expression of both male- and female-specific genes in \( \text{dsx} \) flies, due to lack of the repressive function of both \( \text{DSX}^\text{M} \) and \( \text{DSX}^\text{D} \) proteins, may lead to developmental abnormalities, resulting in the \( \text{XY} \),\( \text{dsx} \) fly being less able to sense attractive females. Thus, even though the neural identity of the fly remains male, it would be less likely to court. This may also account for the increased sex appeal of these flies. Young wild-type males show both a lower courtship frequency and a higher elicitation of courtship than do wild-type adult males. The developmental burden caused by expression of both male- and female-specific genes in the \( \text{XY} \),\( \text{dsx} \) flies may result in a retardation of maturation which causes these flies to retain their sex appeal and low courtship frequency after the time when wild-type males would have lost theirs. Evidence that the CNS of \( \text{XY} \),\( \text{dsx} \) flies is essentially male comes from analysis of the courtship song which, although not exactly wild type, is still clearly recognisable. This is in marked contrast to the anomalous song produced by gyroandromorphs when much of the thoracic nervous system is diplo-X \cite{Taylor1994}.

Taken together, the above evidence suggests that \( \text{tra} \) and \( \text{tra-2} \) regulate the sexual identity of the fly regarding courtship behaviour, but \( \text{dsx} \) does not appear to be required, implying a branch of regulatory genes under the control of \( \text{tra} \) and \( \text{tra-2} \) but not \( \text{dsx} \).

A candidate gene for being involved in such a pathway is the \( \text{fruitless} \) \( \text{fru} \) gene, which has marked effects on male courtship behaviour \cite{Gailey1991,Taylor1994}. Males with extreme \( \text{fru} \) alleles court vigorously but do not attempt copulation and are unable to curl their abdomen. Defects are also observed in the courtship song of these mutants. With regard to the possibility that this gene may be involved in a pathway of regulation governed by \( \text{tra} \) and \( \text{tra-2} \) but not \( \text{dsx} \), two aspects of the \( \text{fru} \) phenotype are particularly interesting. First, certain alleles result in the absence of the MOL, although this does not account for the inability of these flies to curl their abdomen since weaker \( \text{fru} \) alleles which lack this muscle are still able to copulate. As discussed above, \( \text{tra} \) and \( \text{tra-2} \) are required to prevent the formation of this muscle in the female, while \( \text{dsx} \) and \( \text{ix} \) are not. However, whether \( \text{tra} \) and \( \text{tra-2} \) play any part in \( \text{fru} \) regulation is unknown. Second, severe \( \text{fru} \) alleles result in male flies courting nonspecifically such that males are courted with equal vigour as females. Interestingly, lack of courtship discrimination by male flies has also resulted from ectopically expressing \( \text{tra} \) in the antennal lobes or mushroom bodies \cite{Ferveur1995,O'Dell1995}. Again, although mutation of \( \text{fru} \) and misexpression of \( \text{tra} \) have similar effects with regard to loss of courtship discrim-
ination in the male, it is not known whether these two genes form part of a common pathway.

The localised expression of $tra$ described above was achieved using a $tra$ cDNA under the control of the upstream sequence (UAS). This construct was introduced into the genome via P-element-mediated germ-line transformation. To express $tra$ in a specific tissue, it is necessary to produce the yeast UAS-activator protein GAL4 in that tissue. This was done by transforming embryos with a P-element construct containing a GAL4 cDNA under the control of a weak promoter. This weak promoter requires the "help" of a tissue-specific enhancer to express significant levels of GAL4. Thus, the tissue localisation of GAL4 protein in these "enhancer-trap" strains depends entirely upon which tissue-specific enhancers the P-element construct comes under the control. By selecting strains which express GAL4 brain specifically and crossing them to the UAS- $tra$ strain, brain-specific expression of $tra$ can be achieved.

It is becoming increasingly clear that what was previously thought to be a linear hierarchy of regulatory genes is in fact a branched pathway, with sex differentiation genes lying directly under the control of $tra$ and $tra-2$ as well as $dsx$. It is likely that there are a number of different branches at the level of $tra$ and $tra-2$ and the elucidation of these processes will no doubt be the basis of future work.

The Control of Male and Female Sexual Differentiation by $dsx$

The major morphological differences between males and females are apparent during differentiation of the adult at metamorphosis. Many genes expressed uniquely in adult male or female somatic cells have been identified. These include the components of the vitelline membrane and chorion in females (e.g., Waring and Mahowald, 1979; Fargnoli and Waring, 1982; Kafatos et al., 1985) and components of the accessory gland in males (e.g., Schäfer, 1986; Chen et al., 1985; Monsma and Wolfrin, 1988; DeBenedetto et al., 1990). Yet most of the genes encoding these sex-specific proteins are not directly controlled by the $dsx$ gene in the adult. All of the products mentioned above are made in sexually unique tissues or organs and the regulation of the genes encoding them depends upon the presence or absence of tissue-specific factors. The $dsx$ gene has played its role earlier in development by determining and maintaining the state of determination throughout embryonic and larval growth and is no longer required once the cells differentiate. The exceptions to this are the yolk protein (yp) genes expressed in the female (but not the male) adult fat body (Bownes and Nöthiger, 1981) and the glucose dehydrogenase gene expressed in a specific pattern in the male and female reproductive tract (Feng et al., 1991). Both these examples are of genes expressed in a tissue found in adults of both sexes, but with some unique sexually dimorphic functions, showing that $dsx$ can function after differentiation to control sex-specific gene expression in certain tissues.

The major developmental decisions executed by $dsx$ therefore occur during embryonic and larval development. Its functions are well documented by genetic and developmental studies, such as the analysis of mosaics and gynandromorph flies which contain cells of each sex within the same organism (Schüpbach et al., 1978). The role of $dsx$ is perhaps best illustrated for the genital disc, which gives rise to the adult analia and genitalia and comprises three distinct groups of primordial cells. One group will eventually differentiate into either male or female analia depending upon whether $DSX^X$ or $DSX^M$ is expressed. There is, however, a selection between the other two groups of cells, with one group growing in the female and the other in the male and differentiating into the very different genitalia of the two sexes. This means that $dsx$ is able to select either the repression of or promote the growth of whole primordia as well as cause a single group of cells to select between two alternate developmental pathways (Epper and Nöthiger, 1982). How this is achieved is not clear but presumably depends upon the position of the primordial cells in the fly and upon interaction with the segment polarity and/or the segment identifying homeotic genes. Figure 8 shows the development of the male and female genital primordia, discs, and structures generated at metamorphosis in males and females.

Analysis of mosaic patches of sexually transformed cells and temperature shift experiments with a temperature-sensitive allele of transformer-2 ($tra-2ts$) shows that $dsx$ is required throughout the growth of genital discs, functioning not only to set the cells along a specific developmental pathway, but also to maintain that determined state throughout the subsequent cell divisions. Thus $dsx$ is required to maintain the appropriate sexual determination of cells (Wieschaus and Nöthiger, 1982; Epper and Bryant, 1983). Consistent with these results is the observation that $dsx$ mutant flies differentiate both male and female pattern elements, indicating that both groups of primordial cells develop to some extent when neither $DSX^X$ or $DSX^M$ protein is present (Nöthiger et al., 1987).

The main activity of $DSX^X$ and $DSX^M$ seems, from genetic studies, to be to repress the expression of genes needed in the other sex. There are, however, some indications that the proteins may also promote sex-specific gene expression. There are sex-specific differences in the cell divisions which generate the abdominal neurons of the adult. Once the female neuroblasts stop dividing, the terminal abdominal neuroblasts of the male undergo extra divisions. In the absence of $dsx$ (in contrast to the genitalia) the cells do not divide at all in either sex. Thus $dsx$ is required for the non-sex-specific cell divisions prior to the sexually dimorphic cell divisions (Taylor and Truman, 1992).

The male foreleg carries the sex comb and the pattern of neuronal axons differs between the male and female first legs. Recent experiments using ectopic expression of a $dsx$ male-specific cDNA in flies showed that sex comb morphol-
ology could be induced not only in female forelegs but also on second and third legs of both sexes. This has been used to indicate that DSX\textsuperscript{M} positively promotes the development of male-specific structures of the foreleg (Jursnich and Burtis, 1993). However, this should be viewed with care, since dsx nulls do not have a similar phenotype and the dsx\textsuperscript{M} mutants, which constitutively express the male DSX protein, do not have sex combs on other legs. It is possible that dsx is not normally expressed at all in these cells and high levels of expression may lead to new phenotypes by interactions with new combinations of tissue-specific transcription factors.

One interesting point to note is that a lack of DSX proteins, or an expression of both forms of DSX\textsuperscript{F} and DSX\textsuperscript{M} in the same fly, leads to a similar phenotype, namely intersexual flies. When no DSX protein is present, presumably the sex-differentiation genes of both sexes are derepressed and thus both sets of genes are expressed. When both gene products are present it is possible that they interfere with each other’s function, again leading to a partial derepression of the sex-differentiation genes.

Although dsx is not involved in sex determination within germ cells, its correct expression is essential in somatic cells of the gonad for germ cells to be able to develop along a female developmental pathway (Steinmann-Zwicky, 1992). Thus there is signalling between somatic cells and germ cells during development of the germline and gonad to ensure compatibility between germ cell and somatic cell sex differentiation (see section on Sex-lateral and germline sex determination).

Whilst we know a great deal at the level of genetic decisions about developmental fate control by the alternately spliced products of the dsx gene, we have no idea what the target genes of DSX\textsuperscript{F} and DSX\textsuperscript{M} are within imaginal primordia and histoblasts during development. In fact the only immediately identified downstream targets of dsx are the female-specific yolk protein genes expressed in the adult fat body. The glucose dehydrogenase gene expressed in the reproductive tract may also be a target but this has not been investigated at the molecular level.

For normal female differentiation to occur, two other genes are required. These are ix (Baker and Belote, 1983) and her, both of which have zygotic functions essential for normal female differentiation, yet act downstream and independently of the normal sex-specific splicing of dsx (Pultz et al., 1994; Pultz and Baker, 1995). It seems likely that these gene products interact with DSX\textsuperscript{F} to bring about female-specific differentiation. The molecular nature of ix remains to be elucidated, but it is clear that dsx alone does not generate the developmental decisions during embryonic and larval development.

**Execution of the Signal: Interactions of dsx with Downstream Sex Differentiation Genes**

Despite the wealth of information about the developmental consequences of mutations in dsx we have information only about its molecular interactions with one family of downstream genes, encoding the female-specific yolk proteins. Even with these genes we really don’t understand the nature of sexual selection, since DSX\textsuperscript{F} and DSX\textsuperscript{M} bind to the same DNA fragment and give the same footprint, showing that they contact the same DNA bases (Burtis et al., 1991), yet in vivo this leads to yolk protein gene repression in males and expression in females.

The three yolk protein genes (yp1, yp2, and yp3) are expressed in the follicle cells of the ovary and in the adult female fat body (reviewed in Bownes, 1994). Their expression in the follicle cells is not directly controlled by dsx but is cell-type-specific once the dsx-dependent decision to develop and differentiate an ovary rather than a testis has been made (Bownes et al., 1990). In the fat body, yp gene expression is directly controlled by the sex-determination hierarchy (Belote et al., 1985). XX flies with mutations in tra-2 or dsx, for example, can have their yp genes repressed and XY flies with mutations at dsx can express the yp genes in the fat body. This latter finding suggests that the main function of DSX\textsuperscript{M} is to repress expression in males. There is some evidence that DSX\textsuperscript{F} promotes yp gene expression in females (Coschigano and Wensink, 1993).

A number of cis-acting DNA sequences have been identified which confer female body fat expression of yp1, yp2, and yp3, using a variety of reporter gene systems (Garabedian et al., 1985, 1986; Logan et al., 1989; Logan and Wensink, 1990; Liddell and Bownes, 1991; Abrahamsen et al., 1993; Ronaldson and Bownes, 1995). One of these, a 125-bp fragment located 5’ of yp1, called the fat body enhancer (FBE), has been extensively studied, but it should be noted that it is not the only region flanking yp1 and yp2 that can direct female body fat expression of these two divergently transcribed genes. DSX protein, which contains a zinc finger DNA binding domain, binds four times within the FBE. DSX\textsuperscript{F} and DSX\textsuperscript{M} differ at their carboxy termini, not in the DNA binding region, and can compete for the same DNA binding sites in the FBE (Burtis et al., 1991; Erdman and Burtis, 1993).

A number of other proteins, fat body transacting factors, have been shown in vitro to bind to the FBE (Abel et al., 1992; Falb and Maniatis, 1992). These include both enhancers and repressors, the transcriptional activator box B-binding factor-2, the CCAAT/enhancer binding protein, and the adult enhancer factor 1. These proteins regulate alcohol dehydrogenase gene expression in the fat body and give footprints which overlap with those of DSX in the FBE.

It seems that the interaction with these other trans-acting factors may differ according to whether DSX\textsuperscript{F} or DSX\textsuperscript{M} is present. When neither is present transcription will proceed. When DSX\textsuperscript{F} alone is present, transcription also proceeds, but when DSX\textsuperscript{M} alone binds, transcription is repressed. Perhaps DSX\textsuperscript{M} is less easily displaced by transcriptional activators in the fat body. The presence of both DSX\textsuperscript{F} and DSX\textsuperscript{M} allows transcription of the yp genes, possibly because they interfere with each other’s function.

To understand how this is achieved will require a combi-
nation of in vitro gel shift and footprint assays to investigate which combinations of factors will bind or be displaced by each other. Mutations induced in the binding sequences will then be essential to see if these in vitro studies are valid by analysis of the in vivo function of the altered binding sites. Other, as yet unidentified, gene products, such as those encoded by ix (Baker and Belote, 1983), may also have critical roles to play in the sex specificity of yp gene expression. Figure 9 shows our current understanding of the role of dsx in yp gene expression.

Ovarian development requires interaction between the germ cells and somatic cells. We know that signals sent from the sexually determined female somatic cells to the germ cells are essential for female germline development. A gene which is downstream of dsx in the sex determination pathway and which is essential for ovarian and female germ-line development is small ovaries (Wayne et al., 1995). This mutation maps to the X-chromosome and once cloned would be a good possibility for an additional target for dsx gene regulation.

CONCLUSIONS

Alternate splicing provides a mechanism for generating families of related proteins with either dramatic or subtly different functions, depending upon the system. In Drosophila this has been exploited to regulate sexual determination and differentiation by a genetic cascade of alternately spliced products. The alternate splicing of the transcripts encoding both splicing factors and DNA binding proteins is quite well understood, but there are a number of branches in the pathway that remain to be elucidated, along with understanding the mechanism by which the differentially spliced products of the DNA binding proteins can bring about sex-specific transcription of target genes.

ACKNOWLEDGMENTS

We are grateful to Isobel Black and Josie Priestley for typing and manuscript preparation; to Daniel Bopp, Jean Beggs, Rolf Nöthiger and Neil White for providing valuable comments on the manuscript; and to Daniel Bopp, T. Hazeldrig, and W. Mattrox for providing photographs and sharing unpublished data. Research in our laboratory is supported by the MRC and bbsrc.

REFERENCES


numerator element of the X:A signal that determines the state
of activity of Sex-lethal in Drosophila. EMBO J. 8, 3079–3086.
The protein Sex-lethal antagonizes the splicing factor U2AF to
regulate alternative splicing of transformer pre-mRNA. Nature
362, 171–175.
Wang, J., and Bell, L. R. (1994). The Sex-lethal amino acid terminus
mediates cooperative interactions in RNA binding and is essen-
tial for splicing regulation. Genes Dev. 8, 2072–2085.
of synthesis of chorion proteins in Drosophila melanogaster. Cell
16, 599–607.
characterisation of small ovaries, a gene required in the soma
for the development of the Drosophila ovary and the female
germline. Genetics 139, 1309–1320.
Wieschaus, E., and Nöthiger, R. (1982). The role of the transform-
er genes in the development of genitalia and analia of Drosophila
Younger-Shepherd, S., Vaessin, H., Bier, E., Jan, L. Y., and Jan,
HLH protein, acts as a denominator in Drosophila sex determina-
proteins: A conserved family of pre-mRNA splicing factors.
Genes Dev. 6, 837–847.
and domain structure of the mammalian splicing factor U2AF.
Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I., and
Green, M. R. (1992). Cloning and intracellular localization of
the U2 small nuclear ribonucleoprotein auxiliary factor small
Zhou, S., Yang, Y., Scott, M. J., Pannuti, A., Fehr, K. C., Eisen, A.,
Koonin, E. V., Fouts, D. L., Wrightsman, R., Manning, J. E., and
Lucchesi, J. C. (1995). Male-specific lethal 2, a dosage compensa-
tion gene of Drosophila, undergoes sex-specific regulation and
encodes a protein with a RING finger and a metallothetaein-like
cysteine cluster. EMBO J. 14, 2884–2895.

Received for publication September 6, 1995