THE X-LINKED *LSP1α* GENE OF *DROSOPHILA MELANOGASTER* IS NOT ACETYLATED BY MOF, BUT IS SEX-SPECIFICALLY REGULATED BY INDIVIDUAL COMPONENTS OF THE MSL COMPLEX

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Genetics

at Massey University, Palmerston North,
New Zealand

Vikki Marie Weake
2005

ABSTRACT

Male *Drosophila melanogaster* double the transcription of most of the genes on their single X chromosome, to equal that from the two female X chromosomes, in a process termed dosage compensation. This process is mediated by the MSL complex, consisting of both protein and non-coding RNA components. This complex is only active in males due to the presence of MSL2, which is not translated in females.

The X-linked *Lsp1α* gene of *Drosophila melanogaster* appears to escape dosage compensation, and exhibits two-fold higher levels of expression in females compared to males. The apparent lack of dosage compensation of *Lsp1α* could be due to the promoter being more active in females than in males, or to a lack of regulation by the MSL complex. In this study, the mechanism by which this happens has been addressed. *Lsp1α* is expressed exclusively in the fat body tissue of third instar larvae, and forms part of a multi-protein complex that acts as a nutrient reservoir during pupariation. In this study it has been shown that transgenes, in which the reporter gene, *lacZ,* is under the control of the *Lsp1α* promoter, exhibit variable levels of increased activity in female compared to male third instar larvae. At high levels of transgene expression, activity of the transgene is equal in female and male larvae. When the expression of the transgene is low, the activity of the transgene is much higher in female compared to male larvae. This increased sensitivity of the *Lsp1α* promoter to position effects in females appears to be mediated by one or more components of the MSL complex. Females ectopically expressing MSL2 exhibit decreased levels of transgene activity. Furthermore, overexpression of MSL1 causes an increase in the activity of transgenes subject to strong position effects.

Despite these findings, the sex-specific regulation of the *Lsp1α* promoter does not account for the non-dosage compensated appearance of *Lsp1α.* Instead, unlike control dosage compensated X-linked genes, *Lsp1α* is not enriched for a histone modification, acetylation of lysine 16 of histone H4 that is essential for dosage compensation by the MSL complex.

The developmental stage at which the four genes flanking *Lsp1α* are expressed has been determined using northern RNA hybridization. Expression of the gene immediately 3' of *Lsp1α* could not be detected at any developmental stage using northern RNA hybridization or in adults by RT-PCR. However, the two genes flanking *Lsp1α* are expressed in equal levels in male and female *Drosophila* as determined by quantitative RNase protection analysis. Furthermore, the regions between *Lsp1α* and these flanking dosage compensated genes do not prevent dosage compensation of an X-linked *arm-lacZ* reporter gene.

Bioinformatic analysis shows that *Lsp1α* is present in three species closely related to *D. melanogaster* but is absent in more distantly related species. It is probable that because of its recent evolutionary origin, the *Lsp1α* gene lacks the DNA sequences that are required to attract the MSL complex. More generally, a model is proposed in which dosage compensation involves binding of the MSL complex to DNA sequences in actively transcribed regions with possible limited spreading to closely associated active genes.

ACKNOWLEDGEMENTS

Thanks to my supervisor Dr. Max Scott, for his unflagging support, cheerful optimism, and willingness to listen to my most outrageous ideas for new experiments. Thanks also to my co-supervisor Dr. Kathryn Stowell, for her support and encouragement, and for all her practical advice throughout my entire time at university. Special acknowledgment must be made to all the past and present members of the Flyspot lab: Asela, Emma, Xuelei, Simon, Wayne, Wouter, Abhi, Anja, Corey, Carolina, Julia, Bradley, Esther, Jess and Fang. Special thanks go to my family and to Karl for all their support during my seemingly endless studies, and for at least trying to be interested in my crazy little flies.

I would like to acknowledge the funding I have received from the Foundation for Research, Science and Technology in the form of a Doctoral Scholarship, and from the Royal Society of New Zealand Marsden Fund.

Thanks also to Dr. Maria Flores and Professor Phil Crozier at the Auckland Medical school for the use of the Storm phosphorimager. I especially thank Dr. Tim Parnell and Dr. Edwin Smith for their advice on chromatin immunoprecipitation analysis, Dr. Brendon Monahan for assistance with Southern hybridization analysis, and Dr. Pete Lockhart for help with the phylogenetic analysis.

The creation of transgenic *Drosophila* lines in this study has been approved under the protocol number GMO 00/MU/51 by the Massey University Genetic Technology Committee.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ATP	adenosine 5'-triphosphate
BDGP	Berkeley *Drosophila* Genome Project
bp	base pair
BSA	bovine serum albumin
°C	degrees Celsius
ca.	approximately
cDNA	copy deoxyribonucleic acid
CES	chromatin entry site
ChIP	chromatin immunoprecipitation
cpm	counts per minute
CsCl	cesium chloride
CTD	C-terminal domain
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2-deoxycytidine 5'-triphosphate
DF	dilution factor
dGTP	2'-deoxyguanosine 5'-triphosphate
DHS	DNaseI hypersensitive site
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
g	gram
GFP	green fluorescent protein
h	hour
H3K20	lysine 20 of histone H3
H3S10P	phosphorylated serine 10 of histone H3
H4K16ac	acetylated lysine 16 of histone H4
HA	hemagglutinin

HAT	histone acetyl transferase
HCl	hydrochloric acid
HDAC	histone deactylase
HMT	histone methyl transferase
IGEPAL	octylphenylpolyethylene glycol
kb	kilobase-pairs
kDa	kilodaltons
KOAc	potassium acetate
L	litre
LB	Luria-Bertani (media or broth)
M	molar, moles per litre
mg	milligram
μL	microlitre
mL	millilitre
Milli-Q water	water purified by a Milli-Q ion exchange column
μM	micromolar, micromoles per litre
mM	millimolar, millimoles per litre
min	minute
M_r	relative molecular mass (g mol^{-1})
mRNA	messenger ribonucleic acid
MSL	male specific lethal
NaOAc	sodium acetate
nmol	nanomole
nt	nucleotide
ORF	open reading frame
PCR	polymerase chain reaction
pH	-Log $[H^+]$
poly(A)$^+$	polyadenylated
RNA	ribonucleic acid
RNAPII	RNA polymerase II
RNase	ribonuclease
RNAi	RNA interference
RT-PCR	reverse transcriptase-polymerase chain reaction
rpm	revolutions per minute

s	second
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism
ss	single stranded
TE	10 mM Tris, 1 mM EDTA (pH 8.0)
TEMED	N, N, N', N'-tetramethyl-ethylendiamin
Tris	tris(hydroxymethy)aminomethane
Tween-20	polyoxyethylenesorbitan monolaurate
UTR	untranslated region
UV	ultra violet light
V	volt (m^2 kg s^{-3} A^{-1})
WT	wild-type
v/v	volume per volume
w/v	weight per volume
w/w	weight per weight

1. INTRODUCTION

1.1 Chromatin structure

The DNA in the nucleus of eukaryotic organisms is arranged into a highly organized chromatin structure, which has an integral role in the regulation of gene expression. Chromatin is comprised of basic repeating structural units known as nucleosomes, which consist of two superhelical turns of DNA (*ca.* 146 bp) wrapped around an octamer of the core histones H2A, H2B, H3 and H4 (Kornberg, 1974; Kornberg and Thomas, 1974; Luger *et al.*, 1997). The DNA linker between nucleosomes is bound to additional histone proteins such as H1. The small, highly basic histone proteins have C-terminal and amino-terminal extensions (tails) which may be post-translationally modified by acetylation, phosphorylation, methylation, ubiquitination, sumoylation and ADP-ribosylation, as can the core of the histones (Berger and Felsenfeld, 2001; Xu *et al.*, 2005). The existence of these modified histones, and their association with particular epigenetic states has led to the 'histone code' hypothesis. This hypothesis proposes that the site-specific modification of histone-tails specifies a distinct level of organisation of the chromatin structure (Jenuwein and Allis, 2001; Turner *et al.*, 1992). The isolation and characterisation of many of the enzymes that perform these post-translational modifications has confirmed the basic tenets of this histone code hypothesis. In a broad sense, it has become clear that modifications, which loosen the nucleosome structure, tend to promote gene expression, while those modifications that tighten and condense the nucleosome structure are inclined to repress expression. This hypothesis does not however preclude the possibility that certain histone modifications may provide a surface for interaction with other proteins (Berger, 2002).

1.1.1 Histone acetylation

Acetylation of lysine residues located in the amino-terminal tails of the histone proteins neutralises the charge on these tails, and is hypothesised to weaken the interaction between DNA and the histones thus increasing the accessibility of DNA to the transcriptional machinery (Roth *et al.*, 2001). Acetylation may also weaken the interaction between nucleosomes within the stacked solenoid structure (Luger *et al.*,

1997). This modification is catalysed by histone acetyltransferases (HAT), of which there are two classes: cytoplasmic B-type HATs and nuclear A-type HATs. The type A HATs include Gcn5, CREB-binding protein (CBP), p300, $TAF_{II}250$ and MOF, and are associated with transcriptional activation. MOF is involved with dosage compensation in *Drosophila* (section 1.5.5) (Hilfiker *et al.*, 1997). Acetylation can be reversed by the action of histone deacetylases (HDAC), which tend to be associated with the repression of transcription. However, the chromatin state specified by histone acetylation is completely dependent upon the position of the acetylated residue, as some acetylated lysines such as H4K12 are associated with heterochromatin (Turner *et al.*, 1992). More recently residues within the globular domain of the histones have been shown to be acetylated, for example K56 of histone H3 in yeast (Xu *et al.*, 2005). It is postulated that such modifications, which are positioned at the entry-exit points where the DNA wraps around the nucleosome, may regulate nucleosome remodeling.

1.1.2 Histone methylation

Methylation of histones occurs on either arginine or lysine residues. Generally methylation of arginine residues is associated with gene activation, and histone methylases (HMT) that modify arginine residues can be regarded as coactivators for example the CARM1/PRMT4 family of HMTs (Kouzarides, 2002). In contrast, lysine methylation can be associated with both gene activation and repression. For instance, H3K9 methylation is associated with heterochromatin silencing and transcriptional repression, while H3K4 methylation is associated with transcriptional activation (Kouzarides, 2002; Lachner and Jenuwein, 2002). The catalytic activity of HMTs resides within the SET domain, which is common to all families of HMT enzymes.

Histone methylation can be reversed by the histone demethylase LSD-1 (Shi *et al.*, 2004) or by arginine deimination, producing citrulline (Cuthbert *et al.*, 2004; Hagiwara *et al.*, 2005; Hagiwara *et al.*, 2002; Nakashima *et al.*, 2002). H4K20 methylation is associated with silent chromatin (Nishioka *et al.*, 2002b) and is under-represented on the male *Drosophila* X chromosome (section 1.5.9).

1.1.3 Histone phosphorylation

Several histone residues are subjected to phosphorylation, including serine 10 of histone H3. H3S10P has been shown to be associated with transcriptional activation, and also

occurs during the chromosome condensation of mitosis (Cheung *et al.*, 2000). In *Drosophila*, H3S10P is dramatically increased at the heat shock loci during the heat-shock response, while histone acetylation is unaltered, intimating that the increased transcription of the heat-shock genes is in part induced by phosphorylation (Nowak and Corces, 2000). Supporting this, functional heat shock transcription factor activity is required for the increase in phosphorylated H3 at the heat shock loci (Nowak and Corces, 2000). Furthermore, many data suggest that prior phosphorylation may be a requirement for acetylation, with H3S10P and H3K14Ac associating together in regions of actively expressed genes (Lo *et al.*, 2000). Significantly, H3 peptides phosphorylated at Ser-10 are not able to be methylated at lysine 9 by SUV39H1, indicating that prior phosphorylation can prevent methylation (Rea *et al.*, 2000). Phosphorylation of H3S10 by the kinase JIL-1 is associated with dosage compensation in *Drosophila* (section 1.5.6) (Jin *et al.*, 2000; Jin *et al.*, 1999).

1.1.4 Histone ubiquitination

Ubiquitination, although commonly associated with protein degradation, may also have a role in the regulation of transcription (Zhang, 2003). $TAF_{II}250$ in the Tata box-binding protein-associated complex $TF_{II}D$ of *Drosophila* exhibits histone H1 monoubiquitination activity (Pham and Sauer, 2000). Furthermore, mutation of the conserved ubiquitination site of H2B in the yeast, *Saccharomyces cerevisiae*, confers defects in mitotic growth and meiosis (Robzyk and Recht, 2000). The SAGA acetyltransferase complex contains an H2B ubiquitin ligase (Grant *et al.*, 1997). The RING finger motif has been associated with E3 ubiquitin ligase activity in several enzymes including the Rad6-associated protein, Brc1 (Zhang, 2003).

1.1.5 Chromatin assembly and remodeling

Chromatin can be remodeled to facilitate transcription by several different classes of enzymes including the SWI/SNF, ISWI, and Mi-2 families based on proteins originally identified in yeast, *Drosophila* and humans respectively (Kadam and Emerson, 2002). These enzymes utilize the energy derived from ATP hydrolysis to disrupt the interaction between DNA and the histone proteins, thus increasing the accessibility of DNA to the transcriptional machinery. In *Drosophila*, three different protein complexes have been

identified that contain the ISWI ATPase as their catalytic core: ACF[1] (Ito *et al.*, 1997), CHRAC[2] (Varga-Weisz *et al.*, 1997) and the Nucleosome remodeling factor (NURF) (Tsukiyama *et al.*, 1995). While all three complexes catalyze energy-dependent nucleosome sliding *in vitro*, ACF and CHRAC are involved in the assembly of nucleosome arrays while NURF disrupts them (Corona *et al.*, 1999; Hamiche *et al.*, 1999; Langst *et al.*, 1999). For example, regularly spaced nucleosomes can be assembled *in vitro* on plasmid DNA in an ATP-dependent manner in the presence of recombinant ACF[3] and recombinant nucleosome assembly protein 1 (NAP-1) (Ito *et al.*, 1999). While ACF has a catalytic role, depositing histones into nucleosomes periodically, NAP-1 acts stoichiometrically to add two histone H2A-H2B dimers onto DNA already possessing a tetramer of histones H3 and H4. In contrast, NURF, together with transcription factors acts to mobilize nucleosomes at promoters, thus facilitating transcription (Mizuguchi *et al.*, 1997; Tsukiyama and Wu, 1995).

NURF functions in dosage compensation in *Drosophila* (section 1.5.8). It consists of four polypeptide subunits: NURF140, which is an ISWI ATPase (Tsukiyama *et al.*, 1995); NURF55, a WD-40 repeat protein (Martinez-Balbas *et al.*, 1998); NURF38, an inorganic pyrophosphatase (Gdula *et al.*, 1998); and NURF301, which has PHD finger and bromodomain motifs (Tsukiyama and Wu, 1995; Xiao *et al.*, 2001). The histone tails are important for the interaction of NURF with chromatin, but this does not involve any of the potentially acetylated lysines at positions 5, 8, 12 or 16 on the histone H4 tail (Georgel *et al.*, 1997). The interaction of NURF with chromatin may be mediated by the bromodomain of NURF301, or by NURF55, as a WD-40 repeat protein, WDR5, has recently been shown to specifically bind histone H3 methylated at lysine 4 (Wysocka *et al.*, 2005). NURF301 forms a scaffold on which the NURF complex assembles (Xiao *et al.*, 2001). NURF interacts with transcription factors including GAGA factor (Xiao *et al.*, 2001), but NURF and GAGA factor do not overlap significantly on immunostained polytene chromosomes (Deuring *et al.*, 2000).

[1] ATP-utilizing chromatin assembly and remodeling factor

[2] Chromatin accessibility complex

[3] ATP-utilizing chromatin assembly and remodeling factor. This consists of Acf1 and the ISWI (imitation switch) ATPase.

1.2 Transcription

RNA polymerase II (RNAPII) generates a mature mRNA molecule from a chromatin template in a process involving multiple steps that include preinitiation, initiation, promoter clearance, elongation and termination (Sims *et al.*, 2004). At the beginning of each transcription cycle, a preinitiation complex consisting of $TAF_{II}D$, $TAF_{II}B$, $TAF_{II}E$, $TAF_{II}H$ and RNAPII together with additional cofactors assembles at the promoter. The movement from this initiation step into productive transcription elongation, promoter clearance, requires NTPs and is ATP-dependent. At promoter clearance, the C-terminal domain (CTD) of RNAPII is phosphorylated on serine 5, switching the polymerase into elongation mode. Phosphorylation of serine 2 of the RNAPII-CTD increases as the polymerase moves towards the 3' end of genes, and is associated with later elongation and termination (Sims *et al.*, 2004). Histone modifications are involved in the switch to transcription elongation, and the Elongator histone acetyltransferase complex is known to associate with phosphorylated RNAPII. Finally, following transcription termination, the mRNA is cleaved, polyadenylated and transported to the cytoplasm.

1.3 Chromatin insulators and boundary elements

Chromatin insulators and boundary elements are DNA sequences defined mechanistically by their ability to block enhancer-promoter interactions when positioned between them, and to buffer transgenes from chromosomal position effects. These properties imply a role for insulators in regulating the correct temporal and spatial transcription of genes by preventing elements in neighbouring domains from interacting with a non-target gene (Gerasimova and Corces, 2001).

1.3.1 The *SCS/SCS'* insulators

The first insulator elements were identified in *Drosophila* at the borders of the 87A7 heat shock puff, and consist of the *SCS* (*specialized chromatin structure*) and *SCS'* elements surrounding the *hsp70* heat shock locus (Udvardy *et al.*, 1985). These insulator elements, when placed either side of a *mini-white* reporter gene, protect this gene from both negative and positive chromosomal position effects (Kellum and Schedl, 1991). Furthermore, either insulator element is capable of blocking an enhancer from activating an *hsp70-lacZ* reporter gene when interposed between the enhancer and *hsp70* promoter (Kellum and Schedl, 1992). The sequences required for the enhancer blocking activity

of the *SCS* and *SCS'* elements reside within two nuclease-hypersensitive regions (Vazquez and Schedl, 1994). The *SCS* and *SCS'* elements interact with different proteins to mediate their insulator activity. The SBP (*SCS* binding protein), encoded by the *zeste-white 5* gene, interacts directly with a 24 bp region of *SCS in vitro*, and also associates with *SCS in vivo* as part of a nucleoprotein complex (Gaszner *et al.*, 1999). Furthermore, mutations in *zw5* reduce the enhancer blocking activity of multimers of *SCS* fragments containing the 24 bp binding site (Gaszner *et al.*, 1999). Two related proteins, BEAF-32A and BEAF-32B (boundary element associated factor of 32 kDa) interact with the *SCS'* but not the *SCS* element (Hart *et al.*, 1997; Zhao *et al.*, 1995). The transcription factor DREF competes with BEAF-32 for binding to sites within the *SCS'* element (Hart *et al.*, 1999). Some evidence indicates that ZW5 and BEAF-32 are capable of interacting *in vitro*, supporting the model that insulators function by creating higher order chromatin domain structures (Blanton *et al.*, 2001). Significantly with regard to the proposed work, transgenic lines in which a construct consisting of the *mini-white* reporter gene flanked by the *SCS* and *SCS'* elements is inserted on the X chromosome still exhibit elevated *mini-white* expression in males, indicating that the *SCS* and *SCS'* elements are unable to block dosage compensation (Henry *et al.*, 2001; Vazquez and Schedl, 1994). This is consistent with the finding that the *SCS* and *SCS'* elements reside within, but not at, the borders of the heat shock puff at *87A*, and that these elements do not prevent the spread of chromatin decondensation resulting upon activation of a *hsp70-lacZ* transgene (Kuhn *et al.*, 2004). In fact both these elements include the promoter sequences of nearby non-*hsp70* genes (Kuhn *et al.*, 2004). These findings imply that the *SCS* and *SCS'* elements do not prevent the spread of the active chromatin domain formed upon induction of the heat shock locus. Instead these elements maintain transcriptional fidelity in this region, and perhaps prevent inappropriate increases in the level of transcription of the nearby non-*hsp70* genes (Kuhn *et al.*, 2004).

1.3.2 The *gypsy* insulator

A second well-characterized insulator element of *Drosophila* is present in the 5' UTR of the *gypsy* retrovirus (Gdula *et al.*, 1996). Similar to the *SCS* and *SCS'* insulators, the *gypsy* insulator blocks enhancer-promoter interactions when positioned between them, while not inhibiting the enhancer from directing the expression of other adjacent genes. It is also able to buffer transgene expression from position effects, and contains strong DNaseI hypersensitive sites (Cai and Levine, 1995; Chen and Corces, 2001; Gdula *et*

al., 1996; Roseman *et al.*, 1993; Scott *et al.*, 1999). Within this 350 bp element lie 12 copies of a 26 bp sequence to which the zinc finger Su(Hw)[1] protein binds (Geyer *et al.*, 1988). Su(Hw) interacts with a second protein, Mod(mdg4)[2], and this interaction is essential for the enhancer-blocking activity of the *gypsy* insulator (Cai and Levine, 1997; Gdula *et al.*, 1996).

1.3.3 The bithorax complex insulators

The insulator elements present in the bithorax complex of *Drosophila* act in a somewhat different manner to the *SCS/SCS'* and *gypsy* insulators. In this complex, the *Ultrabithorax* (*Ubx*), *Abdominal-A* (*Abd-A*) and *Abdominal-B* (*Abd-B*) genes are expressed in a specific pattern within the embryo along the anterior-posterior axis. The regulatory sequences that determine the segment in which each gene is expressed are separated by insulator elements. Mutations that remove any one of these insulators cause the fusion of adjacent segments, resulting in a distinctive mutant adult phenotype (Mihaly *et al.*, 1998). For example, deletion of the *Fab-7* insulator, which separates the *iab-6* and *iab-7* regulatory sequences that direct the expression of *Abd-B* in parasegments PS11 and PS12, leads to cross-talk between these regulatory elements resulting in homeotic phenotypes in the adult fly (Hagstrom *et al.*, 1996; Zhou *et al.*, 1996). This *Fab-7* insulator also contains DNaseI hypersensitive sites within a 1.2 kb region of DNA (Hagstrom *et al.*, 1996). The insulators of the bithorax complex differ from those discussed previously in that the regulatory sequences are able to overcome them to activate genes when appropriate. This effect may be mediated by a PTS (promoter-targeting sequence) element, such as that present in the *Fab-8* element, which enables distal enhancers to overcome the blocking effect of the *Fab-8* insulator (Barges *et al.*, 2000; Zhou and Levine, 1999). Supporting this, replacement of *Fab-7* with the *SCS* or *gypsy* insulators, while preventing interactions between *iab-6* and *iab-7*, disrupts regulation of *Abd-B* by more distal enhancers (Hogga *et al.*, 2001).

1.3.4 Vertebrate insulators

Insulator elements have also been identified in vertebrates, for example the chicken 5' β-globin locus (Chung *et al.*, 1993). This insulator, the 5' HS4 element, contains a

[1] Suppressor of Hairy-Wing

[2] Modifier of mdg4

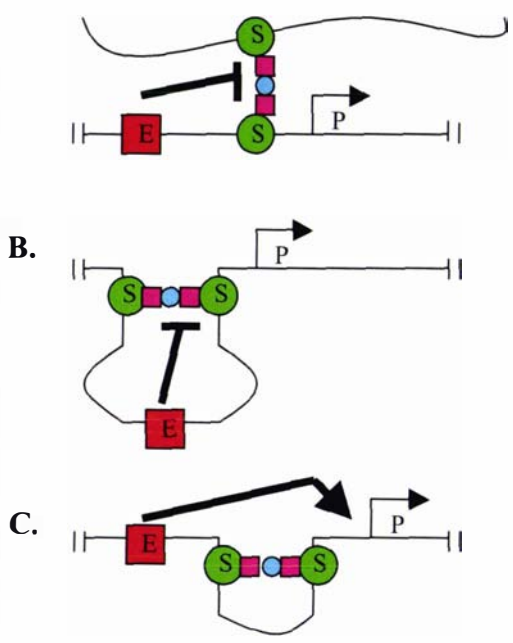
strong DNaseI hypersensitive site, and marks the boundary between the more open acetylated chromatin of the β-globin locus and the more condensed hypoacetylated chromatin outside of the locus (Hebbes *et al.*, 1994; Reitman and Felsenfeld, 1990). The enhancer-blocking activity of the 5' β-globin insulator is entirely due to the presence of a single 42 bp binding site for the protein CTCF within the 5' HS4 element (Bell *et al.*, 1999). A second insulator is present at the 3' end of the locus, which also contains binding sites for CTCF (Saitoh *et al.*, 2000). The chicken 5' β-globin insulator reduces enhancer-directed reporter gene activity when placed either side of the reporter construct on a circular plasmid, but only blocks enhancer-directed activity on a laniaries plasmid when inserted between the enhancer and promoter (Recillas-Targa *et al.*, 1999).

The 5' HS4 element activity resides within a 250 bp *core* element containing five protein binding sites identified by DNaseI footprinting (Chung *et al.*, 1997). While the CTCF binding site is required for the enhancer blocking activity (Bell *et al.*, 1999), this is entirely separable from the barrier activity of the 5' HS4 that blocks the spread of chromosomal silencing (Recillas-Targa *et al.*, 2002). Instead USF proteins are required for barrier activity, with knockdown of USF1 expression by RNAi leading to a loss of the recruitment of active histone modifications such as histone acetylation and H3K4 methylation, and a subsequent invasion of the repressive chromatin modification, H3K9 methylation (West *et al.*, 2004).

1.3.5 Models for insulator function

The models proposed for insulator function can be separated into two types, which are not mutually exclusive. The first proposes that insulators block enhancers from directing the action of promoters either by interacting with the promoter in a 'promoter-decoy' model, or by looping out intervening sequences containing the enhancer so it cannot interact with the promoter. The second proposes that insulators alter localized higher-order chromatin structure, thus preventing an enhancer from directing the promoter by encasing it in a chromatin environment that is not accessible to transcription factors. Further, alteration of localized chromatin structure at an insulator blocks the long-range spread of chromatin modifications required for silencing, or activation. The former model is supported by studies showing that the *gypsy* insulator is able to block enhancer-promoter interactions only when positioned between the enhancer and the promoter, and functions most effectively when flanking the enhancer (Figure 1).

Furthermore, the insertion of two tandem *gypsy* insulators between the enhancer and promoter abolishes this blockage of enhancer activity, and is due to their tandem arrangement rather than close proximity or spacing (Cai and Shen, 2001; Muravyova *et al.*, 2001). Other insulator elements such as those at the chicken 5' β-globin locus appear to function via the latter model, with the insulator activity requiring the constitutive recruitment of particular histone modifications (West *et al.*, 2004).



(A) The Su(Hw) insulator (S) interacts via proteins with another distant Su(Hw) insulator, preventing the enhancer (E) from activating the promoter (P). (B) The flanking Su(Hw) insulator sites interact via proteins, looping the enhancer away from the promoter thus preventing interactions. (C) The tandem Su(Hw) sites interact with each other via proteins, looping out the intervening chromatin between the enhancer and promoter, thus facilitating their interaction, and promoter activation. Based on (Cai and Shen, 2001).

Figure 1. Chromatin looping model of insulator action

1.4 Dosage compensation in mammalian cells

In female mammalian cells, one of the two X chromosomes is inactivated to equalize the dose of X-linked genes to that of males, with one X chromosome (Bernstein and Allis, 2005). One of the two female X chromosomes is randomly inactivated in each of the somatic cells during embryogenesis. This inactivation is initiated at a single site on the X chromosome, known as the X inactivation centre (Xic). The non-coding sense *Xist* and anti-sense *Tsix* RNAs are transcribed from this site on the X chromosome that is to be inactivated. These non-coding RNAs coat the inactive X chromosome, initiating a chain of chromatin remodeling events that are required for silencing. They are essential for the initiation, but not the maintenance of the heterochromatic state on the

inactive X chromosome (Bernstein and Allis, 2005). These include the methylation of H3K9, H3K27 and H4K20, and the hypoacetylation of H3 and H4. The H2A histone variant macroH2A is incorporated into the inactive X chromosome. Furthermore, the inactive X chromosome is characterized by extensive DNA methylation. Together, these epigenetic marks result in a stable heterochromatic state for the inactive X chromosome that is propagated throughout development (Bernstein and Allis, 2005). Despite the parallels between dosage compensation in *Drosophila* and X inactivation in mammalian cells (non-coding RNAs, chromatin remodeling), these processes are remarkably different. Dosage compensation involves hypertranscription of the male X chromosome rather than inactivation of the female X, and *Drosophila* do not exhibit significant DNA methylation in comparison to mammalian cells.

1.5 Dosage compensation in *Drosophila melanogaster*

The male heterogametic sex (XY) of *Drosophila* exhibit similar X-linked phenotypes to the female homogametic sex (XX) despite the apparent dose difference of X-linked genes. This phenomenon of *dosage compensation* was first noted by Muller with regard to X-linked eye colour mutants of *white apricot* (w^a) in which the level of eye pigmentation is directly proportional to the dose of the allele (Muller *et al.*, 1931). In these mutants, homozygous w^a females with two doses of the allele have a level of pigmentation equivalent to that of w^a males carrying only a single copy of the allele. This observation led to the conclusion that a mechanism exists in *Drosophila* to equalize the expression of X-linked genes, thereby compensating for the difference in dosage between the heterogametic and homogametic sex. This mechanism is termed dosage compensation.

The enzyme activities of certain X-linked genes, including *glucose-6-phosphate* (G6PD; *Zw*) and *6-phosphogluconate dehydrogenases* (6PGD; *Pgd*) have been measured in flies carrying varying doses of the X chromosomes (Seecof *et al.*, 1969). No significant difference in activity is observed between diploid females (two doses) and males (one dose) for either G6PD or 6PGD. Dosage compensation occurs at the level of transcription, as the amount of steady-state RNA expressed from X-linked genes such as the salivary gland secretion polypeptide gene (*Sgs-4*) is equivalent in diploid male and female *Drosophila* (Breen and Lucchesi, 1986). Furthermore, transcription from the

entire single male X chromosome, as measured by incorporation of [^{3}H]uridine, is approximately equivalent to that from the paired female X chromosomes (Mukherjee and Beermann, 1965). Both of the female X-chromosomes are actively expressed within the *Drosophila* nucleus, as females heterozygous for X-linked mutant genes affecting cuticular structures do not display phenotype mosaicism, but rather uniformly demonstrate the dominant (wild-type) phenotype. If the X chromosome bearing the dominant gene had been inactivated in some cells during development, it would be expected that patches exhibiting the mutant phenotype would be generated (Lucchesi and Manning, 1987).

The evidence from these studies led to the proposal that the mechanism of dosage compensation in *Drosophila* operates via a two-fold increase in the level of transcription of the single male X chromosome relative to the paired female X chromosomes. This transcriptional upregulation equalizes the yield of gene products from the single male X chromosome relative to the two female X chromosomes, hence compensating for the two-fold difference in dosage. An alternative hypothesis has been proposed by Birchler and colleagues, which states that dosage compensation is caused by reduced transcription of the autosomes in male flies due to a relocation of the protein MOF from the autosomes to the X chromosome (Bhadra *et al.*, 1999). Many key pieces of evidence do not support this alternative hypothesis, including the increased transcription and puffy chromatin structure of the male *Drosophila* X chromosome.

1.5.1 Male-specific lethal mutations

The validity of the mechanism for dosage compensation proposed above was tested by the hypothesis that mutations in genes involved in regulating dosage compensation in *Drosophila* would be lethal to one sex (Belote and Lucchesi, 1980a). For example, a mutation might be lethal to males if the mutant protein had been essential for the upregulation of the transcription of the single male X chromosome, whereas females would remain unaffected due to receiving a normal dose of X-linked gene expression. Such mutations were found in field-collected strains, inspiring extensive screens of ethyl methanesulfonate (EMS)-treated second and third chromosomes for these sex-specific lethal mutations. Four recessive male-specific lethal mutations were isolated; all of these were lethal to homozygous males in the late pupal or larval stage while the viability of homozygous females remained unaffected. The first male specific lethal

mutant, *maleless*^{ts}, is a temperature sensitive allele of *maleless* (*mle*), a mutation that had previously been isolated (Fukunaga *et al.*, 1975; Golubovsky and Ivanov, 1972). The second and third mutations represented alleles of a novel loci, *male-specific lethal-1* (*msl1*), and the fourth, also corresponding to a new loci, was designated *male-specific lethal-2* (*msl2*) (Belote and Lucchesi, 1980a). A second screen of EMS-treated third chromosomes yielded two allelic male-specific mutations of a fourth loci, *male-specific lethal-3* (*msl3*) (Lucchesi *et al.*, 1982). This mutation was allelic to the previously discovered *maleless on the third-132* (*mle(3)132*) (Uchida *et al.*, 1981).

The activities of several X-linked enzymes were found to be reduced in *msl* males to approximately 60% of the levels of those in wild-type males or in *msl* females, whereas autosomal-linked enzyme activities were unaffected in *msl* males relative to *msl* females or wild-type males (Belote and Lucchesi, 1980b). This finding conclusively demonstrated that *msl* males die due to a failure to dosage compensate. Furthermore, this correlates with a significant reduction in the transcription of the single male X chromosome in *mle*^{ts} homozygous mutants in comparison to *mle*^{ts} females (Belote and Lucchesi, 1980b; Breen and Lucchesi, 1986). Together, these results confirmed that the mechanism of dosage compensation in *Drosophila* operates via a two-fold increase in the level of transcription of the single male X chromosome relative to the paired female X chromosomes.

The genes encoding the MSL proteins have been cloned, and include a fifth *msl*, *males-on-the-first* (*mof*) (Table 1). MSL2 differs from the other MSL proteins in that it is present exclusively in wild-type males and is completely absent in wild-type females (Bashaw and Baker, 1995; Kelley *et al.*, 1995; Zhou *et al.*, 1995). It is regulated at the translational level by the female-specific protein SXL (Bashaw and Baker, 1995), which prevents stable association of the *msl2* mRNA with the ribosome (Gebauer *et al.*, 2003).

Gene	Protein size	Protein Family	Features	Reference
msl1	1039 aa	-	X chromosome localization (residues 1-84) Interaction MSL2 (residues 85-186) Interaction MOF & MSL3 (residues 759-1039) 2 N-terminal acidic stretches >40 phosphorylation sites 4 *N*-glycosylation sites	(Copps *et al.*, 1998; Palmer *et al.*, 1993; Scott *et al.*, 2000)
msl2	773 aa	-	N-terminal (residues 1-190) RING zinc finger domain (residues 37-87) involved in X chromosome localization and interaction with MSL1 C-terminal metallothionein-like cysteine rich domain	(Bashaw and Baker, 1995; Kelley *et al.*, 1995; Lyman *et al.*, 1997; Palmer *et al.*, 1993; Scott *et al.*, 2000; Zhou *et al.*, 1995)
msl3	512 aa	chromo superfamily	Interacts with MSL1 directly 6 *N*-glycosylation sites 15 phosphorylation sites chromodomain (residues 437-516) interacts with RNA *in vitro*	(Akhtar *et al.*, 2000; Gorman *et al.*, 1995; Koonin *et al.*, 1995; Scott *et al.*, 2000)
mof	827 aa	histone acetyl transferase family	chromodomain zinc finger (important interaction chromatin) and histone acetyl transferase (MYST) domain (residues 518-827) Interacts *roX2* RNA *in vivo*	(Akhtar and Becker, 2001; Akhtar *et al.*, 2000; Hilfiker *et al.*, 1997; Smith *et al.*, 2000)
mle	1293 aa	ATP-dependent DEAH subfamily of helicases	Helicase and RRM domains, and the C-terminal glycine-rich GGY motifs (residues 940-1293) are important for targeting to X chromosome via *roX* RNA helicase domains	(Kuroda *et al.*, 1991; Richter *et al.*, 1996)

Table 1. The male-specific lethal proteins

This table is based on the properties of the male-specific lethal proteins as reviewed in (Akhtar, 2003).

1.5.2 Evidence for a complex consisting of the MSL proteins

Early studies of the *msl* mutations revealed that flies carrying multiple *msl* mutations were no more severely affected than flies carrying only a single mutation (Belote,

1983). This indicated that the MSL proteins functioned together with regards to dosage compensation, and further evidence for this was provided by immunolocalization studies. Immunolocalization refers to the staining of polytene chromosome preparations with antibodies for the MSL proteins, with the localization of the MSL proteins on the chromosomes visualized by secondary staining with fluorescent antibodies. This technique has been used to demonstrate that MSL1, MSL2, MSL3, MLE and MOF bind exclusively along the entire length of the male X chromosome but are completely absent from the female X chromosomes, and do not localize to any chromosomes in their respective mutants (Bashaw and Baker, 1995; Gorman *et al.*, 1995; Gu *et al.*, 1998; Kelley *et al.*, 1995; Kuroda *et al.*, 1991; Palmer *et al.*, 1993; Zhou *et al.*, 1995). This preferential association with the male X chromosome occurs despite the presence of the MSL proteins (with the exception of MSL2) in both sexes (Bashaw and Baker, 1995; Gorman *et al.*, 1995; Kelley *et al.*, 1995; Kuroda *et al.*, 1991; Palmer *et al.*, 1993; Zhou *et al.*, 1995). These immunolocalization experiments were expanded in several different studies to demonstrate that the MSL proteins in fact colocalize along the entire length of the male X chromosome preparations (Bashaw and Baker, 1995; Bone *et al.*, 1994; Gorman *et al.*, 1995; Gu *et al.*, 1998). This provided initial evidence that the MSL proteins form a complex that preferentially associates with the male X chromosome to mediate dosage compensation.

Supporting this proposal, each of the MSL proteins requires the wild-type gene product of the other four MSL proteins to bind to the male X chromosome in the typical heavily banded pattern (Bashaw and Baker, 1995; Gorman *et al.*, 1995; Gorman *et al.*, 1993; Palmer *et al.*, 1994). Additionally, the stability of each MSL protein is reduced in each of the other *msl* backgrounds (Gorman *et al.*, 1995; Palmer *et al.*, 1993; Palmer *et al.*, 1994). Furthermore, many of the MSL proteins co-fractionate on gel filtration and anion exchange chromatography and interact in a yeast two-hybrid system (Copps *et al.*, 1998). Finally, conclusive evidence for the formation of a complex by the MSL proteins has been provided by the coprecipitation of MSL proteins by antibodies raised against other MSL proteins (Copps *et al.*, 1998; Kelley *et al.*, 1995). The entire complex, including the *roX2* RNA[1] (section 1.5.7) has been partially purified from male tissue

[1] *roX1* RNA is not expressed in S2 cells.

culture S2 cells using immunoprecipitation carried out under "RNA-friendly" conditions (Smith *et al.*, 2000).

1.5.3 Ectopic expression of *msl2* in females induces association of the MSL proteins with the X chromosome

Kelley and colleagues constructed transgenic lines of *Drosophila* in which the *msl2* gene was present under the control of the *heat shock protein 83* (*hsp83*) promoter. Using this system, *msl2* was expressed at constitutive low levels in both female and male flies, and could be induced to high levels after heat shock (Kelley *et al.*, 1995). Although the viability of females carrying this construct is reduced due to inappropriate dosage compensation, enough survive to the third instar larval stage to enable preparation of polytene chromosomes for immunolocalization experiments. These experiments revealed that the MSL1, MSL2 and MLE proteins colocalize on the X chromosomes of females carrying the *hsp83-msl2* construct (Kelley *et al.*, 1995). Furthermore, this system enabled the localization of the MSL proteins in the various *msl* mutant backgrounds to be determined directly, as the *msl* wild-type gene products are not required for viability in female flies.

1.5.4 Chromatin structure of the male X chromosome

The enlargement of the single male X chromosome in polytene chromosome preparations so that its width is comparable to that of the paired female X chromosomes has been observed in many studies (Mukherjee and Beermann, 1965; Offermann, 1936). It appears that this diffuse morphology is due entirely to the presence of the MSL complex, as it is completely reversed in a *msl1* or *msl2* background, but can be induced by the ectopic expression of *hsp83-msl2* in transgenic females (section 1.5.3) (Gorman *et al.*, 1993).

1.5.5 Histone acetylation is associated with the presence of MSL complex and is essential for dosage compensation

A histone acetyl transferase (HAT) was predicted to be associated with the MSL complex even prior to the identification of the MOF HAT as a *male-specific-lethal* gene. It had been observed that a specific acetylated isoform of histone H4, H4K16ac, in which histone H4 is acetylated at lysine 16, predominantly localizes to the X chromosome (Turner *et al.*, 1992). This high-level accumulation of H4K16ac on the

male X chromosome is reduced in *msl1*, *msl2*, *msl3* or *mle* mutants to levels comparable to that of the autosomes or female X chromosomes (Bone *et al.*, 1994). In addition, immunolocalization experiments had shown that the H4K16ac isoform colocalizes with MSL1 and MLE along the male X chromosome (Bone *et al.*, 1994).

The functional confirmation of MOF as the HAT associated with dosage compensation came from experiments demonstrating that purified MOF catalytic domain peptide could acetylate *Drosophila* histones, with a preference for histone H4 *in vitro* (Smith *et al.*, 2000). Furthermore, partially purified MSL complex consisting of the five MSL proteins and *roX* RNA specifically acetylates lysine 16 of histone H4 *in vitro* (Smith *et al.*, 2000). This histone acetylation is entirely ascribable to MOF, as partially purified MSL complex containing mutant protein produced by the mof^1 allele, which has a point mutation in the most highly conserved residue of the acetyl-CoA binding domain, does not acetylate histones (Smith *et al.*, 2000). Furthermore, histone acetylation is essential for dosage compensation as males hemizygous for the mof^1 allele die (Hilfiker *et al.*, 1997).

1.5.6 A kinase, JIL-1 is involved in the complex

The presence of phosphorylation sites on two of the members of the MSL complex, MSL1 and MSL3, had led to some speculation that a kinase might be involved in dosage compensation (Gorman *et al.*, 1995; Palmer *et al.*, 1993). A novel serine/threonine tandem kinase, JIL-1, cloned in an unrelated study, provided a probable candidate for this MSL complex-associated kinase (Jin *et al.*, 1999). JIL-1 colocalizes to the X chromosome with the MSL proteins and is present at a two-fold greater level on the male X chromosome relative to each of the female X chromosomes (Jin *et al.*, 2000; Jin *et al.*, 1999). Furthermore, JIL-1 coimmunoprecipitates from *Drosophila* cell extracts with MSL1, MSL2 and MSL3 (Jin *et al.*, 2000). Additionally, the two-fold upregulation of JIL-1 on the male X chromosome can be induced in females ectopically expressing *msl2* (section 1.5.3), and this upregulation is abolished in a *msl1* or *msl3* background (Jin *et al.*, 2000).

Mutations in *jil-1* significantly decrease viability of both sexes, affecting males more severely than females indicating that JIL-1 is required for dosage compensation (Wang *et al.*, 2001). Significantly, JIL-1 is able to phosphorylate serine 10 of bovine histone

H3 *in vitro*, indicating that the amino-terminal tail of *Drosophila* histone H3, which is identical to that of bovine histone H3, may be a substrate for JIL-1 *in vivo* (Jin *et al.*, 1999). In *jil-1* mutants, polytene chromosome morphology is severely disrupted with the chromosome arms becoming highly condensed, and the male X chromosome being more severely affected than the autosomes (Wang *et al.*, 2001). However, the binding of MSL1, MSL2 and MSL3, and the presence of H4K16ac on the male X chromosome remains unaffected in the absence of JIL-1 (Wang *et al.*, 2001). These results imply a role for JIL-1 in maintaining the open chromatin structure of the autosomes and the male X chromosome. This is consistent with the finding that an upregulation of JIL-1 on the male X chromosome correlates with a colocalized increase in the level of H3S10P (Wang *et al.*, 2001). An interaction between JIL-1 and a splice variant of the complex *lola* locus during early embryogenesis has been reported, although the effect of this interaction on the role of JIL-1 in dosage compensation is unclear (Zhang *et al.*, 2003).

1.5.7 Noncoding RNAs form part of the MSL complex

The identification of MLE as a member of the RNA-dependent ATPase superfamily provoked considerable speculation that an RNA component might also be associated with the MSL complex (Kuroda *et al.*, 1991; Richter *et al.*, 1996). This proposal was confirmed by the finding that treatment of salivary glands with RNaseA prior to the preparation of polytene chromosomes abolishes the association of MLE with the X chromosome, but does not affect MSL1 or MSL2 localization (Richter *et al.*, 1996). It appears that the association of MOF with the X chromosome may also be RNase-sensitive (Akhtar and Becker, 2000). However when MSL1 and MSL2 are overexpressed in a double *roX* mutant background, MSL1, MSL2, MSL3 and MOF bind to many sites along the X chromosome (Oh *et al.*, 2003). Furthermore MOF, MLE and MSL3 have the ability to bind RNA *in vitro* (Akhtar *et al.*, 2000; Buscaino *et al.*, 2003; Richter *et al.*, 1996). These data suggest that the interaction of MLE, and perhaps also MOF with the MSL complex and the male X chromosome is dependent on the presence of a particular species of RNA.

Two of the RNA components of the MSL complex have been isolated, *roX1* and *roX2* (*RNA on the X chromosome*) (Amrein and Axel, 1997; Meller *et al.*, 1997). These genes encode male-specific transcripts of 3.7 and 0.6 kb respectively that do not contain open reading frames longer than 60 amino acid residues, are unspliced, and hence are

unlikely to be translated (Amrein and Axel, 1997; Park *et al.*, 2003). The *roX1* and *roX2* RNAs share only a small 30 nt region of significant homology, with no other significant primary or secondary structure homology (Franke and Baker, 1999), although the genes encoding these *roX* RNAs both contain male specific DNaseI hypersensitive sites (DHS) (Kageyama *et al.*, 2001)

Both *roX1* and *roX2* are regulated by MSL2, independent of its RING finger (Amrein and Axel, 1997; Meller *et al.*, 1997; Rattner and Meller, 2004). The DHS are dispensable for this regulation by MSL2 (Rattner and Meller, 2004) and are not required for the stability or function of the RNAs (Park *et al.*, 2003; Stuckenholz *et al.*, 2003). However there is some evidence by quantitative RT-PCR that the DHS positively regulates *roX* RNA levels in males, and that MSL1, MSL3 and MLE but not the histone acetyltransferase activity of MOF are required for this positive regulation (Bai *et al.*, 2004). Furthermore, MLE regulates *roX2* transcription (Lee *et al.*, 2004).

The *roX* RNAs associate with the MSL complex, as demonstrated by colocalization with the MSL proteins to the X chromosome (Franke and Baker, 1999; Meller *et al.*, 2000; Meller *et al.*, 1997) and RT-PCR analyses of partially purified MSL complexes (Meller *et al.*, 2000). The *roX* RNAs are required redundantly for the association of the MSL complex with the male X chromosome, as MSL proteins do not localize to the X chromosome in double mutants lacking both *roX1* and *roX2* (Franke and Baker, 1999; Meller and Rattner, 2002).

1.5.8 ISWI functions antagonistically to MSL complex

In *Drosophila* larvae homozygous for null mutations in *ISWI* (section 1.1.5), a chromatin remodeling ATPase, the male X chromosome appears more diffuse than in wild-type, while the autosomes and female X chromosomes remain unaffected (Deuring *et al.*, 2000). This aberrant X chromosome morphology is also apparent in *Drosophila* carrying a null mutation in *nurf301*, a component of the NURF complex, indicating that this is the ISWI-containing complex required for X chromosome condensation (Badenhorst *et al.*, 2002). However, this aberrant X chromosome morphology in *ISWI* mutant males can be completely suppressed by blocking acetylation of H4K16. Additionally, induced acetylation of H4K16 in *ISWI* mutant females results in diffuse X chromosome morphology that is indistinguishable from that observed in the *ISWI*

mutant males (Corona *et al.*, 2002). Furthermore, increased expression of MOF enhances the mutant X chromosome phenotype in partial loss of function *ISWI* mutant males. In addition, H4K16 acetylation diminishes the ability of ISWI to interact effectively with its substrate. Together these findings indicate that the H4K16ac modification present along the dosage compensated male X chromosome may inhibit ISWI from chromatin compaction and the consequent transcriptional repression. Recent work indicates that the Iswl ATPase in yeast functions to delay the release of initiated RNA polymerase II into the elongation phase of transcription (Morillon *et al.*, 2005). At the yeast *MET16* locus H3K4, H3K36-di and H3K36-tri methylations are present during a postinitiation regulatory phase prior to the appearance of H3K4-di and H3K79-di methylations at the onset of transcription elongation (Morillon *et al.*, 2005). This suggests that the MSL complex may function as an anti-termination factor that promotes transcription elongation. Intriguingly, an H3K4 methyltransferase, MLL1, has been shown to interact with the zinc finger domain of MOF in human cell lines (Dou *et al.*, 2005). However, this type of interaction may not be present in *Drosophila* as the same zinc finger domain of MOF interacts with MSL1, which is critical for the correct targeting of MOF to the male X chromosome (Morales *et al.*, 2004).

1.5.9 Histone methylation of H4K20 is under-represented on the male X chromosome

The PR-Set7 human HMT specifically methylates H4K20 (Nishioka *et al.*, 2002b). In *Drosophila*, PR-Set7 methylates H4K20, and mutants in this gene suppress variegation indicating it also has a role in silencing gene expression (Karachentsev *et al.*, 2005).

The presence of H4K20 methylation inversely correlates with H4K16ac, indicating that these modifications may be mutually exclusive, and indeed these modifications are competitive *in vitro* (Nishioka *et al.*, 2002b). PR-Set7 localizes to mitotic chromosomes in human cells, which may provide a mechanism for the epigenetic inheritance of this methylated mark (Rice *et al.*, 2002). Transcriptionally active regions lack H4K20 methylation, and this modification is under-represented on the male X chromosome of *Drosophila*, while enriched in condensed chromatin regions including the chromocentre (Nishioka *et al.*, 2002b).

1.5.10 The helicase activity of MLE is essential for dosage compensation

Early work had indicated that MLE could be a member of the ATP-dependent DEAH subfamily of helicases (Kuroda *et al.*, 1991). Purified MLE demonstrates both RNA and DNA helicase activities *in vitro*, forming complexes 3-4 fold more efficiently with RNA (Lee *et al.*, 1997). This helicase activity is dependent on ATP, and translocates in the 3' to 5' direction (Lee *et al.*, 1997). An allele of MLE generated by site-directed mutagenesis (MLE^{GET}) lacking RNA/DNA helicase activity could not rescue homozygous *mle* males, indicating that the helicase activity of MLE is essential for dosage compensation (Lee *et al.*, 1997). However, the mutant MLE^{GET} protein still associates with the male X chromosome, albeit in a reduced fashion (Lee *et al.*, 1997). Interestingly, the pattern of MSL1 binding to the X chromosome was also reduced in the mle^{GET} mutants, in a manner similar to that observed in other *mle* mutants (Lee *et al.*, 1997). This suggests that the helicase activity of MLE is required for the wild-type pattern of MSL complex association with the male X chromosome. Significantly, MLE is required for the incorporation of the *roX* RNAs into the MSL complex, even though partial complexes lacking MLE or the *roX* RNAs still bind several sites on the X chromosome, and complexes without MLE still may contain the *roX* RNAs (Akhtar *et al.*, 2000; Gu *et al.*, 2000; Meller *et al.*, 2000). Due to the ability of MLE to act on RNA/DNA heteroduplexes *in vitro* (Lee *et al.*, 1997), there remains the possibility that MLE directs *roX* RNA-mediated DNA recognition of sequences within the X chromosome (Kelley, 2004). However this is unlikely as sequences within the *roX* RNA homologous to the MSL high affinity binding site (section 1.5.12.1) are not required for this binding (Park *et al.*, 2003).

1.5.11 Assembly of the MSL complex

The first clues as to the order in which the MSL complex assembles and binds the male X chromosome came from immunolocalization studies of the MSL proteins in the various *msl* mutant backgrounds (Bashaw and Baker, 1995; Gorman *et al.*, 1995; Gorman *et al.*, 1993; Gu *et al.*, 1998; Lyman *et al.*, 1997; Palmer *et al.*, 1994). This evidence suggested that MSL1 and MSL2 assemble first in the complex, then bind together to the X chromosome, followed by MSL3, MOF and MLE, which bind co-operatively. Furthermore, recombinant MSL1 and MSL2 interact in a yeast-two hybrid system with the RING finger domain of MSL2 being an essential requirement for this interaction (Copps *et al.*, 1998). In addition, transgenic flies carrying *msl2* alleles with

point mutations within the RING finger domain are male-lethal, whereas lines carrying alleles with point mutations in other domains within the protein still support male development (Lyman *et al.*, 1997). From these results it can be concluded that MSL1 and MSL2 interact via the RING finger domain of MSL2, and that this interaction is required prior to X chromosome localization. There is also the potential that an uncharacterized enzymatic activity of the RING finger domain of MSL2 is required for dosage compensation, as the RING finger domain in other enzymes has been shown to possess E3-Ubiquitin ligase activity (Zhang, 2003).

MSL1 and MSL2 also interact with MSL3 but not MLE, as demonstrated by yeast-two hybrid assays (Copps *et al.*, 1998). The N-terminal and C-terminal regions of MSL1, but not the middle region are essential for its association with other proteins in the MSL complex, as determined genetically (Scott *et al.*, 2000). Significantly, truncated MSL1 that lacks the first 84 amino acid residues cannot bind to the male X chromosome, in contrast to full-length FLAG-tagged MSL1 (Scott *et al.*, 2000). Further genetic analyses indicated that the N-terminal region of MSL1 interacts with MSL2, while the C-terminal region of MSL1 interacts with MOF and MSL3. These interactions were confirmed *in vivo* by affinity chromatography (Scott *et al.*, 2000). The kinase domain of JIL-1 associates with MSL1 and MSL3, as is consistent with the presence of many phosphorylation sites on both the MSL1 and MSL3 proteins (Jin *et al.*, 2000). However, interactions with MOF, MLE, MSL2, or the *roX* RNAs have not yet been examined, hence a potential association of JIL-1 with these components of the complex cannot be ruled out.

There are two main candidates for the component of the MSL complex with which the *roX* RNAs interact, MOF and MSL3, although MLE also interacts with RNA. Both MOF and MSL3 interact nonspecifically with RNA preferentially over DNA in electrophoretic mobility shift assays via their chromodomains as shown by site-directed mutagenesis (Akhtar and Becker, 2000). However, partially purified MSL complexes that lack MLE still contain *roX2* RNA, as detected by RT-PCR, indicating that MLE is not essential for association of *roX2* RNA with the MSL complex (Akhtar and Becker, 2000). Supporting this, *in vivo* transient expression of MOF derivatives, some of which are mutated in their chromodomains, revealed that the chromodomain is essential for the specific co-immunoprecipitation of *roX2* RNA (Akhtar and Becker, 2000). However,

MOF, MSL3, MSL1 and MSL2 can be co-immunoprecipitated independently of RNase treatment (Akhtar and Becker, 2000). This indicates that their association is either dependent solely on protein-protein interactions, or that the RNA species required for the interaction is protected within the complex from RNase digestion. Significantly, the patterns of both *roX1* and *roX2* localization on the male X chromosome are disrupted in *mle* mutants, so that they are located only at their site of synthesis. This indicates that the MLE helicase is required to package the *roX* RNAs into MSL complexes (Kageyama *et al.*, 2001; Meller *et al.*, 2000). Furthermore, in mle^{GET} mutants, in which MLE protein is produced that lacks helicase activity, MSL complexes that lack the *roX* RNAs assemble only at the 30-40 high affinity sites (Gu *et al.*, 2000). Somewhat surprisingly, the pattern of *roX1* but not *roX2* localization is also disrupted in *msl3* mutants, with *roX1* being present only at its site of synthesis whereas *roX2* also localizes to the 30-40 high affinity sites (Meller *et al.*, 2000). However, *roX2* localization is unaffected in the *roX1* mutant background indicating that the RNAs do not require each other for inclusion in the MSL complex (Meller *et al.*, 2000).
Interestingly, MOF appears able to autoacetylate both itself and MSL3 at K116 near its chromodomain (Buscaino *et al.*, 2003). These authors suggested that this acetylation decreased the affinity of MSL3 for *roX2* RNA, inducing dissociation from the X chromosome, which is RNA-dependent. The affinity of MSL3 for *roX* RNA and hence the X chromosome is maintained by deacetylation by RPD3. However, this K116 residue is not conserved in MSL3 from other *Drosophila* species (Kelley, 2004).

Together, these studies suggest that the MSL complex is based around MSL1, which binds via an N-terminal amphipathic α-helical domain to the RING finger domain of MSL2 prior to X chromosome localization (Figure 2). MSL3 and MOF then bind the C-terminal domain of MSL1. Both MSL3 and the MOF HAT interact with the *roX2*, and potentially *roX1* RNAs within the complex. MSL3, which is subject to acetylation by MOF, must be deactylated by RPD3 in order for it to associate with the *roX2* RNA and the complex. It is likely that MSL1 and MSL3 are phosphorylated by and/or bind to the JIL-1 kinase domain. MLE probably requires the presence of the *roX* RNAs to associate with the MSL complex, and its RNA helicase activity is required for the incorporation of *roX* RNA into the complex. MSL3 is required for the association of *roX1* RNA with the complex.

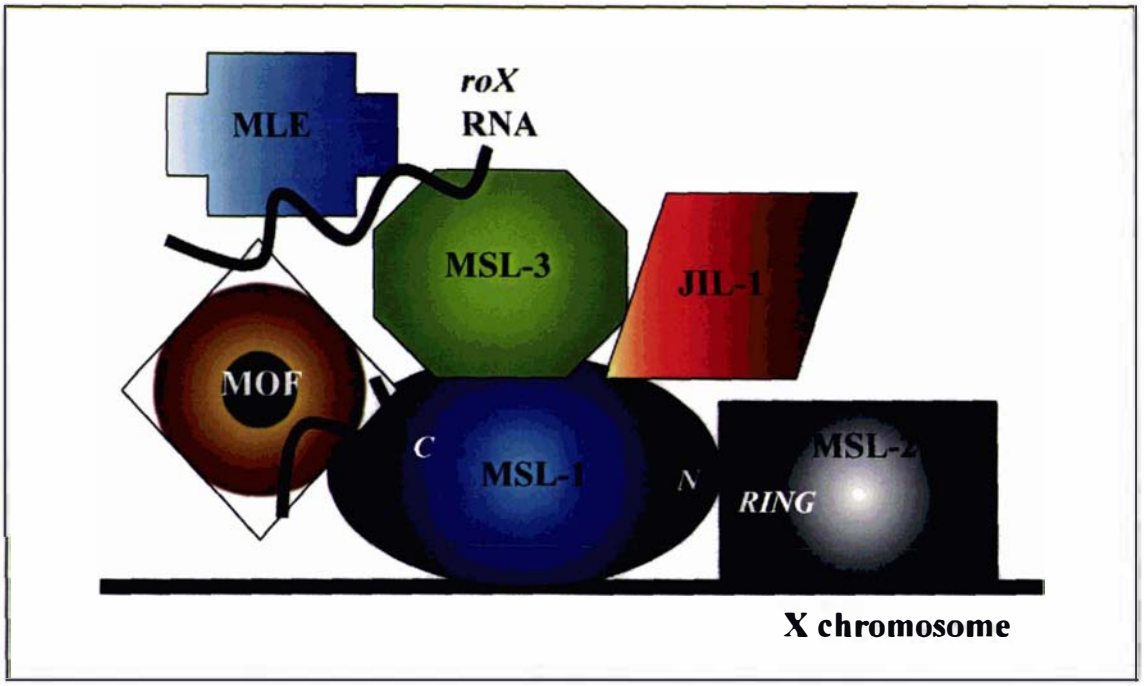


Figure 2. Model of the MSL complex

1.5.12 Recognition of the X chromosome by the MSL complex

There are two major models proposed to describe how the MSL complex distinguishes the X chromosome from the autosomes. The first of these proposes that the MSL complex recognizes short enhancer-like sequences near most X-linked genes, that are absent from autosomal genes (Lucchesi and Manning, 1987). This is supported by studies in which large regions of the X chromosome translocated to autosomes retain dosage compensation, while regions of the autosomes translocated to the X chromosome do not become dosage compensated (Lucchesi and Manning, 1987). Furthermore, an autosomal segment translocated onto the X chromosome still resembles an autosome, and does not take on the appearance of the more enlarged and diffuse X chromosome (Lakhotia, 1970). However, many autosomal genes inserted on the X chromosome become dosage compensated, while only some X-linked genes, such as *white* retain partial dosage compensation on an autosome (Lucchesi and Manning, 1987; Qian and Pirrotta, 1995). This indicates that enhancer-like sequences are not present on every individual X-linked gene. A $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ synthetic polynucleotide recognizes more X-linked than autosomal sites in many *Drosophila* species that have the ability to dosage compensate (Pardue *et al.*, 1987), but this sequence has not been proven to have any function in dosage compensation. An alternative form of this sequence-recognition model postulates that these enhancer-like sequences may target

the MSL complex to multiple genes within a limited domain (Lucchesi and Manning, 1987). This study provides support for an amended version of this model (section 8).

The second model proposes that the MSL complex recognizes a small set of X-linked *chromatin entry sites*, and spreads from these sites in *cis* into the rest of the chromosome. The evidence supporting this model will be outlined in the following sections.

1.5.12.1 A partial MSL complex bind to 30-40 high affinity sites that include the *roX* genes

MSL1 and MSL2 bind weakly in a reproducible fashion to a subset of 30 - 40 sites along the X chromosome in the *msl3*, *mle* and *mof* and double *roX1 roX2* mutant backgrounds (Gorman *et al.*, 1995; Gu *et al.*, 1998; Lyman *et al.*, 1997; Palmer *et al.*, 1994). It has been proposed that these high affinity binding regions constitute *chromatin entry sites* (CES), at which the MSL complex assembles before spreading into flanking regions of chromatin (Kelley *et al.*, 1999; Lyman *et al.*, 1997).

Two of these chromatin entry sites correspond to the *roX1* and *roX2* genes (Kelley *et al.*, 1999). Both genes had previously been mapped to polytene chromosome bands corresponding to CES, with *roX1* mapping to 3F and *roX2* to 10C (Amrein and Axel, 1997; Meller *et al.*, 1997). These genes, when inserted on an autosome in transgenic flies provide novel CES for the MSL complex as determined by immunolocalization in an *msl3* background (Kelley *et al.*, 1999). Transcription of the *roX* genes is not required for this association of the MSL complex, as a *roX2* cDNA transgene lacking a promoter, and a *roX1* partially-deleted transgene that produced no detectable RNA still bound both complete and partial MSL complexes (Kelley *et al.*, 1999). Male *roX1 roX2* double mutants can be rescued by introduction of either *roX1* or *roX2* as a transgene on an autosome (Meller and Rattner, 2002). This suggests that the function of the *roX* genes as a CES is secondary to their role within the MSL complex in regards to dosage compensation, and that other CES on the X chromosome are sufficient to recruit the MSL complex.

The CES activity of *roX1* was mapped to a 217 bp fragment within the exons of the gene that does not include the 30 nt region of homology to *roX2* (Kageyama *et al.*,

2001). This fragment contains a male-specific DNaseI hypersensitive site (DHS), and is sufficient for both recruitment and spreading of the MSL complex even though it does not produce any detectable transcripts (Kageyama *et al.*, 2001). A similar male-specific DHS is present in the *roX2* gene that is sufficient for recruitment of the MSL complex (Park *et al.*, 2003). MSL proteins bind to this 217 bp region of the *roX1* gene by chromatin immunoprecipitation (Kageyama *et al.*, 2001). This shows that nascent RNA transcribed from the *roX* genes is not required for association of the MSL complex with their CES. However, MSL binding to the *roX* CES is dependent on the presence of *roX* RNA in the complex, although the sequences within the RNA homologous to the DHS are not required for this binding (Park *et al.*, 2003). This indicates that a simple model in which MLE catalyses RNA-mediated recognition of the CES is unlikely. The DHS of *roX1* and *roX2* share some consensus, which is conserved between species, and consists of the palindromic sequence $GAGAGN_{4-5}TC[T/C]CTCTC$ (Park *et al.*, 2003). Interestingly, this consensus includes several GAGA sequences (Park *et al.*, 2003). GAGA sequences are recognized by proteins encoded by the *trithorax-like* (GAGA factor) and *pipsqueak* genes (Farkas *et al.*, 1994; Lehmann *et al.*, 1998). GAGA factor may target NURF to regulatory regions to maintain chromatin in an accessible conformation (Cavalli, 2002), although GAGA factor is not male-specific so this does not explain the male-specificity of the DHS. Candidates for further CES identified by sequence homology to this consensus did not demonstrate MSL binding, indicating that this consensus sequence is not sufficient to target the MSL complex (Park *et al.*, 2003).

Recently, a third CES was identified at 18D10 containing a male-specific DHS that is sufficient to attract a full or partial MSL complex *in vivo* (Oh *et al.*, 2004). This CES is not transcribed, and attracts the MSL complex only as a multimer of 4 copies of the 510 bp sequence (Oh *et al.*, 2004). This CES does not share the consensus motif characterized for the *roX1* and *roX2* DHS, but does bind MSL2 as determined by chromatin immunoprecipitation (Oh *et al.*, 2004).

Approximately 30 – 40 residual bands of MSL2 staining remain associated with the X chromosome in the *roX1 roX2* double mutants[1], and several autosomal sites exhibit

[1] These sites correspond to the 30-40 chromatin entry sites on the X chromosome at which there is residual MSL1 and MSL2 binding in the *mof*, *mle* or *msl3* mutant backgrounds (section 1.5.12.1).

strong MSL2 staining, as do the fourth chromosome and the chromocentre (Meller and Rattner, 2002; Oh *et al.*, 2003; Park *et al.*, 2003). Moreover, the enrichment of H4K16ac on the male X chromosome is abolished in *roX1 roX2* double mutants, with a concomitant increase in the level on autosomes and especially within centric heterochromatin within the chromocentre (Meller and Rattner, 2002). Intriguingly, the autosomal sites to which MSL2 strongly localizes in the double *roX1 roX2* mutants appear puffed, a morphology associated with actively transcribed regions of chromatin (Meller and Rattner, 2002). These results suggest that the MSL complex may still partially assemble and bind to certain sites on the male X chromosome in the absence of either *roX* RNA. The survival of a few (*ca.* 5%) adult *escaper* male *roX1 roX2* double mutants further supports this (Meller and Rattner, 2002), as does the ability to increase the proportion of these double mutant male survivors by overexpressing MSLs (Oh *et al.*, 2003). It appears likely that the *roX* RNAs are required for the correct chromosomal localization of the MSL complex, judging by the relocation of MSL1, MSL2 and H4K16ac to the autosomes and chromocentre in the double *roX1 roX2* mutant (Meller and Rattner, 2002; Oh *et al.*, 2003; Park *et al.*, 2003). It is possible that the MSL complex regulates the transcription of several autosomal genes in a *roX*-independent manner, perhaps sex-specifically, and that the puffed appearance of and increased MSL association with these autosomal sites in the double *roX1 roX2* mutants reflects the increased availability of the MSL complex to these sites as it fails to fully localize to the male X chromosome (Meller and Rattner, 2002). Intriguingly centric heterochromatin contains sequences rich in the repeat $(AAGAGAG)_n$, which resembles conserved elements within the *roX* DHS (section 1.5.12.1), indicating that the MSL complex may have an affinity for these GAGA-type sequences (Park *et al.*, 2003).

1.5.12.2 The MSL complex exhibits limited spreading from CES into flanking regions of the chromosome

Kelley and colleagues noted that multiple bands of MSL staining sometimes flanked the *roX1* transgene inserted on an autosome, indicating that the complex could spread from the initial CES into flanking chromatin (Kageyama *et al.*, 2001; Kelley *et al.*, 1999). The extent and frequency of spreading varied considerably between different transgenic insertion lines, and within nuclei of the same transgenic fly (Kelley *et al.*, 1999). Spreading occurred in both directions but was not contiguous, with distinct bands of MSL binding observed flanking the transgene (Kelley *et al.*, 1999). Furthermore,

spreading can occur from the autosomal transgene *in trans* to a paired homologue, although naturally this could not occur in male flies from the single X chromosome (Kelley *et al.*, 1999). As observed on the male X chromosome, H4K16ac and the *roX1* and *roX2* RNAs also colocalize with the MSL proteins in these flanking bands as determined by immunolocalization and *in situ* hybridization respectively (Kelley *et al.*, 1999). A similar phenomenon was observed with the insertion of a *roX2* transgene on an autosome (Meller *et al.*, 2000). This suggests a model for dosage compensation in which the MSL complex initially assembles at these 30 – 40 CES, which include the *roX1* and *roX2* genes, before spreading into flanking chromatin to mediate dosage compensation of surrounding genes.

This begs the question of how the MSL complex spreads. Several clues to this have come from the analysis of partial MSL complex localization in *msl* mutants. While a partial complex binds to the 30 – 40 CES in *msl3*, *mle* and *mof* backgrounds, this complex does not spread beyond these CES into surrounding chromatin as in the wild-type background (Gorman *et al.*, 1995; Gu *et al.*, 1998; Lyman *et al.*, 1997; Palmer *et al.*, 1994). Additionally, the complex cannot spread beyond ectopic CES induced by the insertion of *roX1* or *roX2* transgenes on an autosome in *msl3* or *mof* mutants (Gu *et al.*, 2000; Kageyama *et al.*, 2001). Furthermore, the complex does not bind the autosomal *roX1* transgene in *mle* flies (Kageyama *et al.*, 2001). This spreading deficiency is not simply due to the lack of complete complexes, because in mof^{1} mutants, which form complete MSL complexes that demonstrate no HAT activity, complexes including mutant MOF protein can access the CES but not spread into flanking chromatin (Gu *et al.*, 2000). In addition, a lack of spreading from CES is also observed in mle^{GET} mutants, which form complexes lacking only helicase activity and *roX* RNA (Gu *et al.*, 2000).

Additional clues as to the mechanism of spreading came from observation of the effect of over-expression of certain members of the MSL complex. Significantly, over-expression of MLE results in the localization of MLE protein to all of the polytene chromosome arms, rather than just the male X chromosome (Richter *et al.*, 1996). Similarly, MOF over-expression not only induces genome wide acetylation of lysine 16 on histone H4, but also causes association of MSL1 and MSL2 with many other chromosomal sites (Gu *et al.*, 2000). This suggests that the acetylation of H4K16 by MOF may provide a binding site for further MSL complexes, enabling them to spread

beyond the initial CES. Additionally, MLE may also function in the spreading of the MSL complex, perhaps by enabling the incorporation of the *roX1* and *roX2* RNAs. Noteworthy with regards to this proposal is the observation previously noted (section 1.5.7) that in double *roX1 roX2* mutants the MSL complex, as well as the H4K16ac isoform is partially relocated to the autosomes and removed from the male X chromosome (Meller and Rattner, 2002).

Although initial experiments indicated that transcription of the *roX* RNAs was not required for their ability to recruit the MSL complex (Kelley *et al.*, 1999), later work shows that the MSL protein/*roX* RNA ratio has a large influence on the extent of MSL spreading into flanking chromatin (Park *et al.*, 2002). In fact, spreading from *roX* transgenes is inversely correlated with the number of expressed *roX* genes, indicating that these CES compete for the MSL proteins required for spreading (Park *et al.*, 2002). Spreading does not occur from an autosomal *roX* transgene in the presence of a wild-type X chromosome, but occurs extensively in *roX* double mutants (Park *et al.*, 2002). Transcription of the *roX* transgenes is essential for this spreading to occur (Park *et al.*, 2002). This suggests that MSL complexes will form at the site of *roX* transcription onto the nascent *roX* transcripts if there are sufficient MSLs present in the nucleus. If the MSLs are instead recruited to the X chromosome by additional interactions distinct from those mediated by the *roX* RNA as well as by the pool of newly transcribed *roX* RNA, they are unavailable to form additional complexes around the *roX* transgene and spread into flanking chromatin. Thus the ratio of MSL protein to *roX* RNA may determine the extent of spreading from the *roX* CES into flanking chromatin. The 18D10 CES is not transcribed, so it is unclear whether it will be affected in a similar manner.

1.5.12.3 The MSL complex is recruited to sequences lacking CES

The MSL complex binds X to autosome transpositions whether or not they contain a putative CES (Oh *et al.*, 2004). Furthermore, the transposed region adopts the enlarged appearance of the male X chromosome as observed in previous experiments (Lakhotia, 1970; Oh *et al.*, 2004). The MSL complex is targeted to actively expressed genes, including X-linked transgenes, and its distribution pattern along the X chromosome varies in polytene preparations from different tissues (Sass *et al.*, 2003). However an autosomal region translocated to the X chromosome did not attract MSL complex

despite the nearby presence of a CES (Lakhotia, 1970; Oh *et al.*, 2004). Further analyses confirmed that X-linked fragments ranging from 0.3 kb to 39 kb are able to attract the MSL complex when inserted on an autosome regardless of the presence of a CES (Fagegaltier and Baker, 2004; Lakhotia, 1970; Oh *et al.*, 2004). These findings demonstrate that a model in which the MSL complex spreads *in cis* from just the *roX* CES, or even the remainder of the *ca.* 35 CES, is insufficient to explain the mechanism of MSL binding to the X chromosome. Mitzi Kuroda has proposed an amended model to explain these findings, in which MSL targeting sites or recognition elements on the X chromosome have varying affinities for the MSL complex (Lakhotia, 1970; Oh *et al.*, 2004). These sites range from high-affinity *e.g. roX* CES, to immediate-affinity *e.g.* 18D10, to low-affinity sites. Immediate and low-affinity sites might require the interaction of several sites to attract the MSL complex, whereas high-affinity sites might be involved in different aspects of gene activation rather than the two-fold dosage compensation, and might bind the complex tightly via the *roX* RNAs. Bruce Baker has gone further with his proposal that spreading of the MSL complex is not involved in dosage compensation and that the CES only exhibit increased binding affinity for the MSL complex (Fagegaltier and Baker, 2004). This is supported by the finding that the number of CES observed is directly related to the concentration of the MSL proteins (Demakova *et al.*, 2003).

1.5.13 The MSL complex upregulates transcription and relieves position effects

The model proposed for dosage compensation requires evidence that the recruitment of the MSL complex to a CES is sufficient to induce a two-fold upregulation in transcription. This evidence has been provided by work in which the *roX1* and *roX2* gene fragments containing the CES were inserted separately into an insulated construct upstream of the *armadillo* promoter driving expression of the *lacZ* reporter gene, after which this entire construct was transformed into flies (Henry *et al.*, 2001). Males carrying either of these constructs typically exhibited elevated β-galactosidase activity in comparison to females, with the full length cDNAs inducing ratios close to the two-fold increase expected for dosage compensation, while the CES alone were less effective as either monomer or multimers. Some lines did nonetheless demonstrate male/female ratios of two, indicating that the autosomal environment in which the construct integrates may significantly affect the transcriptional enhancement by the MSL complex. Recent work in this lab has indicated that multimers of the 18D10 CES

are sufficient to induce dosage compensation of the *arm-lacZ* reporter gene, and that the *roX* CESs may be atypical MSL binding sites (Laverty & Weake, Massey University, 2005, unpublished data).

The effect of chromosome position on the action of the MSL complex was confirmed by analysis of transgenic flies carrying the *GMR-roX1-mini-white* transgene integrated at a repressive chromatin environment (Park *et al.*, 2002). In females, the *mini-white* transgene is expressed only in the dorsal eye, while in males, sporadic sectors of expression are observed in the ventral half of the eye. This suggests that the recruitment of the MSL complex to the *roX1* transgene in these males ameliorates the repressive local chromatin environment, increasing expression in the ventral sector of the eye. This work was extended by studies in which male-specific position effect variegation was observed in *ca.* 10% of *roX1* transgenic stocks, in which the transgene had inserted in autosomal euchromatin (Kelley and Kuroda, 2003). This effect may be mediated by MOF, which derepresses transcription of a chromatin template *in vitro* (Akhtar and Becker, 2000). This derepression of transcription provides strong evidence against the Birchler model for dosage compensation (Kelley and Kuroda, 2003).

Histone acetylation correlates with transcriptional upregulation of X-linked genes in males (Smith *et al.*, 2001). Chromatin immunoprecipitation was used to map the distribution of H4K16ac on X-linked genes known to be dosage compensated. This showed that H4K16ac is present along the entire length of dosage compensated genes such as *Pgd* and *Zw*, and is not restricted to the promoter regions (Smith *et al.*, 2001). Rather, the highest levels of H4K16ac are present in the middle and 3' UTR of the X-linked *Pgd* gene (Smith *et al.*, 2001). This is consistent with a model in which the two-fold increase in transcription associated with dosage compensation is facilitated by an enhancement of transcription elongation rather than promoter accessibility (Smith *et al.*, 2001). Significantly, the H4K16ac isoform is absent from a region of the X chromosome containing the *runt* gene known to be dosage compensated independently of the MSL complex (section 1.6) (Smith *et al.*, 2001).

1.6 Genes dosage compensated by an alternative mechanism to the MSL complex

The death of *Sxl* mutant females in an *msl* background supports the proposal that SXL may be required for the MSL-independent dosage compensation of a subset of X-linked genes in *Drosophila* (Bernstein and Cline, 1994). Supporting this, the X-linked gene *runt*, expressed early in development, is dosage compensated by a *Sxl*-dependent, *msl*-independent mechanism (Bernstein and Cline, 1994; Gergen, 1987). This mechanism may possibly act differently to the MSL-mediated dosage compensation by negatively regulating female expression of the gene (Kelley *et al.*, 1995). Significantly, a subset of X-linked genes has been identified using a bioinformatics approach that contains three or more copies of the poly-U stretch (AU_7 or U_8) SXL binding site (Kelley *et al.*, 1995). These genes include *Sxl* itself, *cut*, *runt*, *forked* and *bar*.

1.7 Non-dosage compensated genes

Alleles of some X-linked genes including *white* (*white eosin* and *white cherry*) have been shown to lack dosage compensation (Muller-Hermelink, 1932; Smith and Lucchesi, 1969). A temperature-sensitive *vermillion* allele has been isolated that lacks dosage compensation at the restrictive temperature (Camfield, 1974). *Sgs-4* is not dosage compensated in two wild-type stocks, Samarkand and Karsnäs, which have reduced levels of the SGS-4 polypeptide (Korge, 1981). The *yolk* protein genes (*Yp1-3*), which are expressed exclusively in females, can be induced in males by mutation of *Doublesex (dsx)*. However female X/X; *dsx/dsx* flies have twice the level of YP1-3 compared to X/Y; *dsx/dsx* flies, indicating that these are not dosage compensated (Ota *et al.*, 1981). These non-dosage compensated alleles were reviewed in (Lucchesi and Manning, 1987).

1.7.1 *Lsp-1α* is not dosage compensated

Larval serum protein-1 (LSP-1) constitutes one of the major components, together with LSP-2, of the haemolymph of *Drosophila* larvae just prior to pupariation (Roberts *et al.*, 1977). LSP-1 protein first appears in the young third instar larva, rising to a maximum of 8% of the total extractable protein by puparium formation (Roberts *et al.*, 1977). LSP-1 is composed of three polypeptides, the α-, β- and γ-chains, which are differentiable by molecular weight (Roberts *et al.*, 1977). The genes encoding these

three polypeptides have been mapped, with the β- and γ-chain genes mapping to autosomes and the α-chain to the X chromosome at the polytene interval 11A7 - 11B9 (Brock and Roberts, 1983; Roberts and Evans-Roberts, 1979b; Smith *et al.*, 1981). The genes encoding the subunits of LSP-1 have been cloned, with the α- and β-genes demonstrating considerable homology and differing from the γ-gene (Smith *et al.*, 1981). LSP-1 is not essential for survival, as homozygous *lsp1* null mutants[1] are viable, however the fecundity of the mutant is reduced, perhaps due to a reduction in the number and viability of eggs, and a rise in abnormal mating behavior (Roberts *et al.*, 1991b; Roberts *et al.*, 1991a).

Significantly, the *Lsp-1α* gene[2] is not dosage compensated, with twice as much of the α-chain polypeptide being produced in females in comparison to males, while the β- and γ-chain genes are expressed equivalently in both sexes (Roberts and Evans-Roberts, 1979a). Similarly, *Lsp-1α* RNA is not dosage compensated (Brock and Roberts, 1982). It is probable that there may be very little selective pressure for compensation on the α-chain, which has been proposed to be a recent translocation to the X chromosome from an autosome, as the α-polypeptide is very similar to both the β- and γ-chains (Gonzalez *et al.*, 2004; Roberts and Evans-Roberts, 1979a).

Relocation of the gene together with its 5' regulatory elements to 5 autosomal and 2 ectopic X chromosome sites results in its equivalent expression in males and females relative to the control genes, *rp49* and *Sgs-3* (Ghosh *et al.*, 1989). However, one of the X-linked insertions shows only partial dosage compensation with slightly increased activity being observed in females. These findings differ from the observations of this study, and possible reasons for these differences are given in the Discussion section (section 8). However, initially these findings suggested that *Lsp1α's* lack of dosage compensation was not caused by *cis*-acting regulatory elements within the 5' promoter or enhancer region. This suggests that *Lsp-1α* is capable of being dosage compensated, and that its location on the X chromosome, rather than regulatory elements within the gene itself precludes this possibility. A gene immediately adjacent to the 5' end of *Lsp-1α, L12/CG2560*, was also included in the construct (Ghosh *et al.*, 1989). Unexpectedly

[1] Carrying null mutations in all three of the LSP-1 α, β and γ subunits.

[2] *Lsp1α* is also referred to as *CG2559*

however, the *L12/CG2560* gene that ends approximately 1 kb from the 5' start point of *Lsp-1α* appears to be fully dosage compensated in wild-type *Drosophila* (Ghosh and Mukherjee, 1992). This gene encodes a *ca.* 1.4 kb transcript and is expressed maximally in second instar larvae[1] (Ghosh *et al.*, 1992).

1.7.2 The *Lsp1α* gene domain

Sequencing of the *Drosophila* genome led to the identification of three other genes close to *Lsp1α* (Adams *et al.*, 2000) (Figure 3; Table 2). Two genes, *CG15926* and *CG2560* (*L12*) are situated 13.4 kb and 1.3 kb respectively from the 5' end of *Lsp1α*. A third gene, *CG2556* is located 4.3 kb from the 3' end of *Lsp1α*. These four genes are all transcribed in the same direction from the upper strand of DNA. The fourth gene, *CG15730* is situated on the complementary strand, and is transcribed in the opposite direction from *Lsp1α*. This gene lies 1.5 kb from the 3' end of *Lsp1α*, and 0.6 kb from the 5' end of *CG2556*. This gene domain corresponds to the cytological location 11B. Significantly, CES for the MSL complex have been mapped to 11A and 11B-D, indicating that the dosage compensation status of *Lsp1α* cannot be attributed to the failure of this section of the X chromosome to recruit the MSL complex (Lyman *et al.*, 1997).

[1] Northern analysis of *CG2560* developmental expression indicates that it is maximally expressed in first instar larvae, rather than second instar larvae (section 3.2).

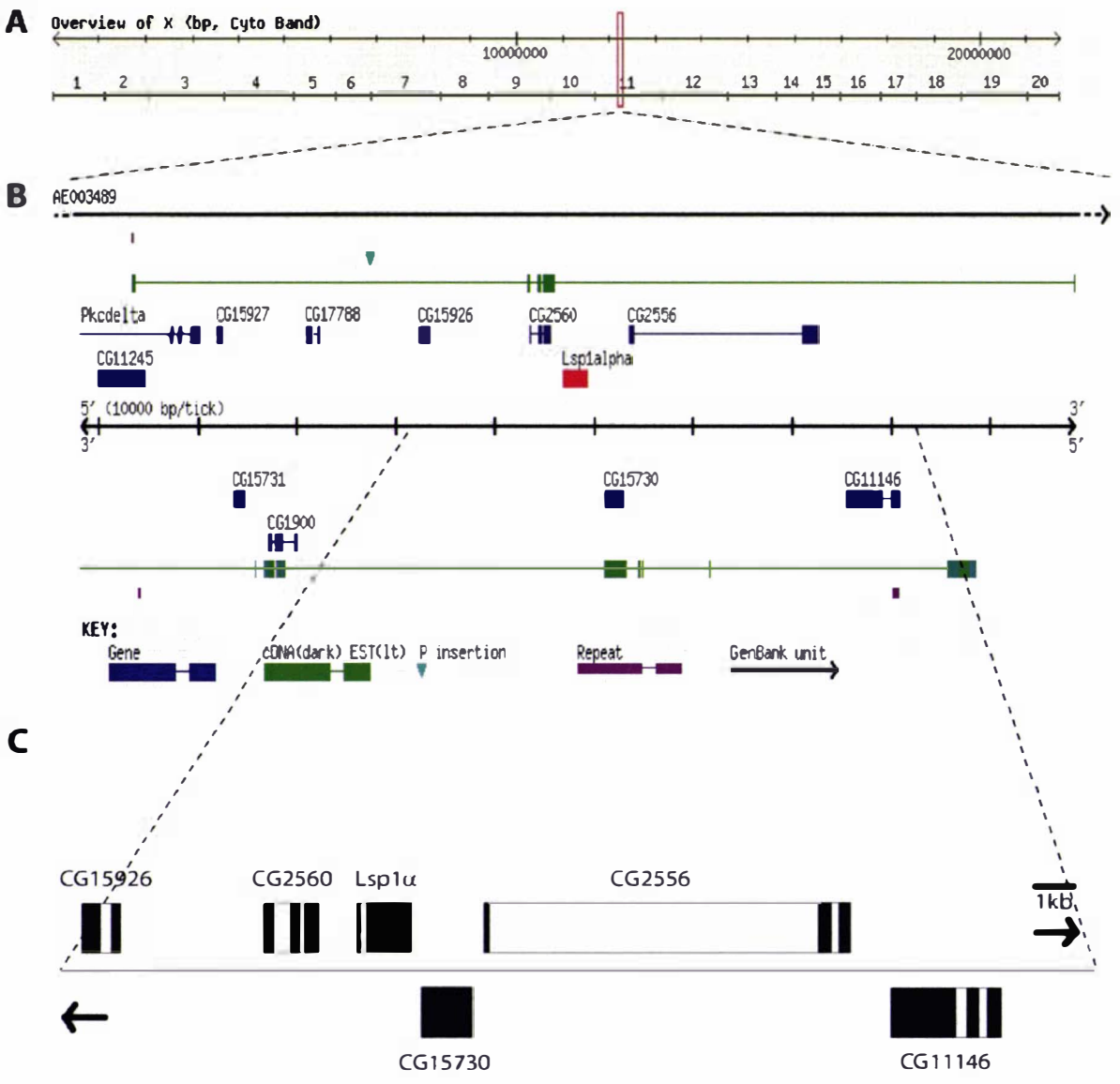


Figure 3. *Lsp1α* gene domain

Section of X chromosome (A) containing the *Lsp1α* gene (red) consisting of 100 kb (release 2) from positions 12,208,150 to 12,308,149 (B). Taken from the *Drosophila* Genome Project website (www.fruitfly.org). (C) The *Lsp1α* gene region is shown schematically in detail. Exons are indicated in black. The direction of transcription is indicated by the placement of genes above (forward strand) or below (reverse strand) the line.

Gene	Position	Evidence transcribed	Predicted gene product/domains
CG15926	X:12,242,347-12,243,492	1 cDNA 2 ESTs	-
CG2560 (L12)	X:12,253,510-12,255,555	1 cDNA 2 ESTs	insect cuticle protein
Lsp1α	X:12,256,891-12,259,408	(Roberts *et al.*, 1977)	larval serum protein hemocyanin
CG15730	X:12,260,971-12,263,070	1 cDNA	-
CG2556	X:12,263,713-12,282,568	1 cDNA 5 ESTs	-

Table 2. Predicted genes within the *Lsp1α* gene domain

Based on the information available on the *Drosophila* Genome Project website (www.fruitfly.org) (Adams *et al.*, 2000; Ashburner *et al.*, 2000).

1.8 Aim

Lsp1α appears to escape regulation by the MSL complex, exhibiting two-fold higher levels of expression in females compared to males. The results of Ghosh *et al.* (1989) indicate that *Lsp1α* is capable of regulation by the MSL complex when moved to other locations on the X chromosome. This suggested that the position of *Lsp1α* on the X chromosome was responsible for its lack of regulation by the MSL complex. It also indicated that the *Lsp1α* promoter was unlikely to be sex-specifically regulated. The initial aim of this work was therefore to confirm that the MSL complex does not regulate *Lsp1α* at its normal location.

The two models for the way in which the MSL complex recognises X-linked genes suggest two alternative explanations for how a gene might escape regulation by the MSL complex. If *Lsp1α* is not regulated by the MSL complex, these models lead to two potential hypotheses to explain this escape from dosage compensation. Firstly, that *Lsp1α* escapes regulation by the MSL complex because it lacks the sequences required to attract the MSL complex due to its evolutionary history. The alternative hypothesis states that if the MSL complex binds to a set of chromatin entry sites and spreads from these into actively transcribed genes on the X chromosome, it follows that some type of boundary element must prevent the MSL complex from accessing *Lsp1α*. Thus chromatin insulators may be involved in preventing the spread of the MSL complex into the domain containing *Lsp1α*, resulting in the failure of this gene to be dosage compensated. Hence the general aim of this study was to determine the boundaries of the non-dosage compensated region surrounding *Lsp1α*, and then to determine whether elements were present in the regions between *Lsp1α* and neighbouring dosage compensated genes that were able to block the spread of the MSL complex.

Objective 1: Determine the boundaries of the non-dosage compensated region surrounding *Lsp1α*

Firstly, to determine the developmental stage at which the genes surrounding *Lsp1α* are expressed using northern RNA hybridization analysis. Following this, to quantify the relative levels of transcripts for each gene using RNase protection in male and female *Drosophila.*

Objective 2: Test the regions between the flanking dosage compensated genes and *Lsp1α* for the ability to block the spread of the MSL complex

To clone the regions between *Lsp1α* and the flanking dosage compensated genes and insert either side of an X-linked reporter gene. Quantify level of reporter expression in male and female *Drosophila* in order to determine whether these regions are able to block the MSL complex from spreading into, and hypertranscribing the X-linked reporter construct.

Objective 3: Analyze *Lsp1α* for the presence of the histone modification associated with dosage compensation by the MSL complex

Use chromatin immunoprecipitation analysis to determine if *Lsp1α* is enriched for the histone acetylation mediated by the MSL complex. Include controls of X-linked genes expressed in the same tissue as *Lsp1α*.

Objective 4: Determine the evolutionary history of *Lsp1α* using phylogenetic analysis of *Drosophila* species

Use the genome sequence information from *Drosophila* species to identify LSP1 homologues. Compare these protein sequences to *D. melanogaster* LSP1α and conduct phylogenetic analysis to determine the relationship between LSP1 proteins in the different *Drosophila* species.

2 MATERIALS AND METHODS

2.1 Bacterial media

All bacterial media was prepared as described in (Sambrook *et al.*, 1989).

2.2 Antibiotics and media additives

All antibiotics and media additives were prepared as described in (Sambrook *et al.*, 1989). Ampicillin was used at a final concentration of 50 μg/ml, chloroamphenicol at 170 μg/ml, X-gal (5-bromo-4-chloro-3-indyol-β-*D*-galactosidase) at 25 μg/ml, and IPTG (Isopropylthio-β-*D*-galactosidase) at 25 μg/ml.

2.3 Growth of bacterial cultures

All bacterial cultures were grown in Luria Broth (LB) media as described in (Sambrook *et al.*, 1989). Long-term stocks of cultures were stored at -80°C as glycerol stocks (Sambrook *et al.*, 1989).

2.4 Stock solutions

All solutions were prepared as described in (Sambrook *et al.*, 1989) or (Ashburner, 1989b).

2.5 *Drosophila melanogaster*

All fly stocks were raised on standard cornmeal/molasses media sprinkled with yeast at either at 18°C, 22°C, 25°C or 28°C. Virgins were collected the morning following an overnight incubation at 18°C of a bottle that had been cleared of adults. Only those flies identified as virgins by examination under a dissecting microscope were selected for use in further crosses.

2.5.1 Description of stocks

A summary of the stocks used that were not generated in this study is provided in Table 3.

2.5.1.1 *y w*

The *y w* fly stock was used to generate all transgenic stocks, and also for RNase protection analysis, northern analysis and in chromatin immunoprecipitation assays.

2.5.1.2 Male-specific actin-GFP

This fly stock has the genotype FM7I, P{w^{+mC}=Act GFP}JMR3/c(1)DX,y^{1}f^{1}, and was provided by the Bloomington *Drosophila* Stock Centre from flies donated by Jean-Marc Reichhart (Stock 4559). The male flies carry an X chromosome expressing GFP under the control of the constitutive *actin 5C* promoter. The female flies have a compound X chromosome (X^X), hence cannot inherit the *actin-GFP* transgene and do not express GFP. This stock was used to separate male and female first instar larvae for RNase protection analysis.

2.5.1.3 L^{2}/CyO

This fly stock has the genotype y w; L^{2}/CyO, pr cn^{2} y^{+} and is used as a second chromosome marker for the genetic linkage analysis of w^{+} transformants.

2.5.1.4 GFP CyO

This fly stock has the genotype w[*]; ln(2LR)noc^{4L}Sco^{rv9R}, b^{1}/CyO, P{w[+mC]=ActGFP}JMR1, and was provided by the Bloomington *Drosophila* Stock Centre from flies donated by Michael Ashburner (Stock 4533).

2.5.1.5 Ser/Sb

This fly stock has the genotype y w; TM3, y^{+} Ser^{1}/Sb^{1} and is used as a third chromosome marker for the genetic linkage analysis of w^{+} transformants.

2.5.1.6 GFP Ser/Sb

This fly stock has the genotype w*; Sb^{1}/TM3, P{w+mC=ActGFP}JMR2, Ser^{1}, and was provided by the Bloomington *Drosophila* Stock Centre from flies donated by Michael Ashburner (Stock 4534).

2.5.1.7 CyO/Bc

This fly stock has the genotype Bc/CyO, pr cn^{2}. Bc is a larval marker.

2.5.1.8 CyO/Tft

This fly stock has the genotype y w; b el rd^S Tft cn/CyO, pr, cn^2.

2.5.1.9 Transposase stock

This fly stock has the genotype w*; $ry^{506}Sb^1P\{ry^{+7.2}=Delta2\text{-}3\}99B/TM6B,Tb^1$ and is a source of P transposase (Robertson *et al.*, 1988). This stock was provided by the Bloomington *Drosophila* Stock Centre (stock 1798).

2.5.1.10 Homologous recombination

This fly stock has the genotype y^1 w*; $P\{ry^{+t7.2}=70FLP\}11$ $P\{v^{+t1.8}=70I\text{-}SceI\}2B$ nocSco/CyO, S^2 and was provided by the Bloomington *Drosophila* Stock Centre from flies donated by Kent Golic (Stock 6934) (Gong and Golic, 2003). The balancer may be In(2L)CyO, In(2R)CyO, S^2 CyO^1. These flies express the FLP site-specific recombinase and I-SceI endonuclease under the control of the *hsp70* heat shock promoters. These transgenes are linked to the scutoid marker.

2.5.1.11 Constitutive FLP

This fly stock has the genotype w^{1118}; $P\{ry^{+t7.2}=70FLP\}10$ and was provided by the Bloomington *Drosophila* Stock Centre from flies donated by Kent Golic (Stock 6938). These flies express the FLP site-specific recombinase constitutively.

2.5.1.12 FLP test stock

This fly stock has the genotype w*; $P\{w^{+mW.hs}=FRT(w^{hs})\}2A$ and was provided by the Bloomington *Drosophila* Stock Centre from flies donated by Norbert Perrimon (Stock 1997). These flies carry the *mini-white* eye marker gene flanked by sites for the FLP site-specific recombinase.

2.5.1.13 hsp83-MSL2

This fly stock has the genotype y w; $msl1^{L60}$/CyO; $P\{w^+=hsp83MSL2\}$, and was provided by Rick Kelley and Mitzi Kuroda. It is used for the overexpression of MSL2 (Kelley *et al.*, 1995). The toxicity of the transgene in females is relieved by the removal of *msl1*. The L^{60} allele of *msl1* is a γ ray induced deletion that removes most of the coding region (*ca.* 2.5 kb).

2.5.1.14 hsp70-MSL1

This fly stock has the genotype y w; P{w^+=hsp70-MSL1}/ P{w^+=hsp70-MSL1}. It is used for the overexpression of MSL1 at 29°C (Scott *et al.*, 2000).

2.5.1.15 hsp70-MSL3

This fly stock has the genotype y w; P{w^+=hsp70-MSL3}/ P{w^+=hsp70-MSL3}. It is used for the overexpression of MSL3 at 29°C (Scott *et al.*, 2000).

2.5.1.16 hsp70-MLE

This fly stock has the genotype y w; P{w^+=hsp70-MLE}/ P{w^+=hsp70-MLE}. It is used for the overexpression of MLE at 29°C (Scott *et al.*, 2000).

2.5.1.17 hsp70-HA-MOF

This fly stock has the genotype y w; P{w^+=hsp70-HA-MOF}/ P{w^+= hsp70-HA-MOF}. It is used for the overexpression of MOF at 29°C. MOF is fused in frame to the HA tag at its amino terminal (Scott *et al.*, 2000).

Fly stock	Genotype	Reference/Source
y w	*y w*	
Male-specific actin-GFP	FM7I, P{w^{+mC}=Act GFP}JMR3/c(1)DX,y^{1}f^{1}	Jean-Marc Reichhart
L^{2}/CyO	y w; L^{2}/CyO, pr cn^{2} y^{+}	
GFP CyO	w[*]; ln(2LR)noc^{4L}Sco^{rv9R}, b^{1}/CyO, P{w[+mC]=ActGFP}JMR1	Michael Ashburner
Ser/Sb	y w; TM3, y+ Ser^{1}/Sb^{1}	
GFP Ser/Sb	w*; Sb^{1}/TM3, P{w+mC=ActGFP}JMR2, Ser^{1}	Michael Ashburner
CyO/Bc	Bc/CyO, pr cn^{2}	
CyO/Tft	y w; b el rd^{S} Tft cn/CyO, pr, cn^{2}	
Transposase	w*; ry^{506}Sb^{1}P{ry$^{+7.2}$=Delta2-3}99B/TM6B,Tb^{1}	(Robertson *et al.*, 1988)
Homologous recombination	y^{1} w*; P{ry$^{+t7.2}$=70FLP}11 P{v$^{+t1.8}$=70I-SceI}2B nocSco/CyO, S^{2}	(Gong and Golic, 2003)
Constitutive FLP	w^{1118}; P{ry$^{+t7.2}$=70FLP}10	Kent Golic
FLP test stock	w*; P{w$^{+mW.hs}$=FRT(w^{hs})}2A	Norbert Perrimon
hsp83-MSL2	y w; msl1^{L60}/CyO; P{w^{+}=hsp83MSL2}	(Kelley *et al.*, 1995)
hsp70-MSL1	y w; P{w^{+}=hsp70-MSL1}	(Scott *et al.*, 2000)
hsp70-MSL3	y w; P{w^{+}=hsp70-MSL3}	(Scott *et al.*, 2000)
hsp70-MLE	y w; P{w^{+}=hsp70-MLE}	(Scott *et al.*, 2000)
hsp70-HA-MOF	y w; P{w^{+}=hsp70-HA-MOF}	(Scott *et al.*, 2000)

Table 3. Fly stocks utilized in this study that were not generated in this project

2.5.2 Description of fly crosses

2.5.2.1 Linkage analysis

To determine the chromosome carrying the transgene insert, three separate crosses were conducted. Firstly, orange eyed male flies carrying the transgene were crossed to *y w* virgin females (section 2.5.1.1) at 25°C. If the transgene and marker are on the X chromosome, then all of the male offspring would have white eyes, and all the female offspring would have orange eyes. Secondly, orange-eyed male flies carrying the transgene on an autosome were crossed with L^2/CyO virgin females (section 2.5.1.3) at 25°C, and orange eyed, curly-winged male offspring were selected. These males were test crosses to *y w* virgin females at 25°C. If the transgene and marker are on the second chromosome then all the *CyO* offspring would have white eyes. Thirdly, orange-eyed male flies carrying the transgene on an autosome were crossed with TM3, *Ser/Sb* virgin females (section 2.5.1.5) at 25°C, and orange eyed, serate-winged male offspring were selected. These males were test crosses to *y w* virgin females at 25°C. If the transgene and marker are on the second chromosome then all the *Ser* offspring would have white eyes.

2.5.2.2 Scheme to move transgene inserts from the second chromosome to the X chromosome by mobilization of *P*-elements

This scheme was provided courtesy of Rick Kelley (Baylor College of Medicine, 2004, personal communication), and provides a means of moving transgene inserts from the second chromosome to another chromosome, including the X chromosome. The transposase stock $w^*;\ ry^{506}Sb^1P\{ry^{+7.2}=Delta2\text{-}3\}99B/TM6B,Tb^1$ (section 2.5.1.9) is crossed to the *CyO* stock $y\ w;\ L^2/CyO,\ pr\ cn^2\ y^+$ (section 2.5.1.3), and $y\ w;\ +/CyO;\ ry^{506}Sb^1P\{ry^{+7.2}=Delta2\text{-}3\}/+$ male offspring (stubble, curly) collected. These males are crossed to homozygous w^+ virgin females carrying the transgene construct on the second chromosome. Male $w^+/CyO;\ ry^{506}Sb^1P\{ry^{+7.2}=Delta2\text{-}3\}/+$ offspring (orange-eyed, stubble, curly) are collected and crossed to *y w* virgin females. Any orange-eyed, non-stubble, curly virgin female offspring are collected. For these flies to have both orange eyes and curly wings, the transgene insert and w^+ marker must have moved from the second chromosome. w^+ male offspring were not collected as the transgene would not be X-linked. The orange-eyed, non-stubble, curly virgin female offspring are crossed to *y w* male flies, and the following generation of orange males test crossed for X-linkage of the w^+ marker.

2.5.2.3 Homologous recombination scheme

This scheme is similar to that designed by (Gong and Golic, 2003). As homologous recombination is much more efficient in the female germline compared to the male germline, females are used for the second generation of the cross. Virgin females of the homologous recombination stock y^1 w^*; $P\{ry^{+t7.2}=70FLP\}11$ $P\{v^{+t1.8}=70I\text{-}SceI\}2B$ noc^{Sco}/CyO, S^2 (section 2.5.1.10; scutoid, curly) are crossed to w^+ *GFP* flies carrying the targeted construct (section 5.2). This cross is heat shocked on day 3 for 1 h at 38°C to increase expression of the FLP recombinase and I-SceI endonuclease, which are under the control of the heat shock promoter. As the targeted construct is on chromosome 2, w^+ *GFP*, scutoid, non-curly virgin female offspring are crossed to L^2/CyO males (section 2.5.1.3). Targeting of the construct to the X chromosome is detected by examining offspring of this cross for segregation of the w^+ and *GFP* markers (*i.e. w GFP* progeny). This cross is described schematically in Figure 26.

2.5.2.4 Homologous recombination test crosses

Males that constitutively express the FLP recombinase (section 2.5.1.11) were crossed to virgin females containing either the *mini-white* marker gene flanked by FRT sites (section 2.5.1.12) or the targeted construct (section 5.2) in which the *arm-EGFP* reporter is flanked by FRT sites. Adult progeny of the first cross were examined for the presence of the *mini-white* marker, and larval and adult progeny of the second cross were examined for presence of the *arm-EGFP* marker.

2.5.2.5 Overexpression of MSL2 in flies carrying reporter constructs

Virgin *GFP-Ser* females (section 2.5.1.6) were crossed with *CyO* males overexpressing MSL2 (section 2.5.1.13). *CyO, GFP-Ser* males that lacked the $msl1^{L60}$ mutation and carried one copy of *hsp83-MSL2* were crossed with virgin female flies homozygous for the reporter construct to be tested. All the progeny of this cross carry one copy of the reporter construct, and either the *GFP-Ser* balancer or the *hsp83-MSL2* chromosome.

2.5.2.6 Overexpression of MSL1 in flies carrying reporter constructs

Virgin GFP-Ser females (section 2.5.1.6) females were crossed to males overexpressing MSL1 (section 2.5.1.14). P{w+=hsp70-MSL1}/ TM3, P{w[+mC]=ActGFP}JMR2, Ser^1 male progeny were crossed to virgin females homozygous for the reporter transgene to

be tested at 29°C. All the progeny of this cross carry one copy of the reporter construct, and either the *GFP-Ser* balancer or the *hsp70-MSL1* chromosome.

2.5.2.7 Overexpression of MSL3 in flies carrying reporter constructs

Virgin GFP-CyO females (section 2.5.1.4) females were crossed to males overexpressing MSL3 (section 2.5.1.15). P{w+=hsp70-MSL3}/ CyO, P{w[+mC]=ActGFP}JMR1 male progeny were crossed to virgin females homozygous for the reporter transgene to be tested at 29°C. All the progeny of this cross carry one copy of the reporter construct, and either the *GFP-CyO* balancer or the *hsp70-MSL3* chromosome.

2.5.2.8 Overexpression of MLE in flies carrying reporter constructs

Virgin GFP-Ser females (section 2.5.1.6) females were crossed to males overexpressing MLE (section 2.5.1.16). P{w+=hsp70-MLE}/ TM3, P{w[+mC]=ActGFP}JMR2, Ser^1 male progeny were crossed to virgin females homozygous for the reporter transgene to be tested at 29°C. All the progeny of this cross carry one copy of the reporter construct, and either the *GFP-Ser* balancer or the *hsp70-MLE* chromosome.

2.5.2.9 Overexpression of MOF in flies carrying reporter constructs

Virgin GFP-CyO females (section 2.5.1.4) females were crossed to males overexpressing MOF (section 2.5.1.17). P{w+=hsp70-HA-MOF}/ CyO, P{w[+mC]=ActGFP}JMR1 male progeny were crossed to virgin females homozygous for the reporter transgene to be tested at 29°C. All the progeny of this cross carry one copy of the reporter construct, and either the *GFP-CyO* balancer or the *hsp70-HA-MOF* chromosome.

2.5.3 Generation of transgenic fly lines

2.5.3.1 Co-precipitation of plasmid DNA

Plasmid DNA prepared by CsCl banding (section 2.6.2.4) and the helper vector plasmid pUChsπΔ2,3 (7.3 kb; pUC18 containing 3.6 kb *P*-element coding sequence, *hsp70* promoter and *ry* 3' flanking region) (Rio and Rubin, 1985) were ethanol precipitated together and resuspended in 20 – 50 µl of injection buffer (0.1 mM NaH_2PO_4, 0.1 mM Na_2HPO_4, 5 M KCl) at 0.4 µg/µl and 0.1 µg/µl respectively. The solution was centrifuged at 13 000 *g* for 5 min prior to use to remove particulate matter that could

block the needle. Three µl of coprecipitated DNA was loaded into a Femtotip (Eppendorf) using a microloader (Eppendorf).

2.5.3.2 Microinjection of plasmid DNA into *Drosophila* embryos

Microinjection was conducted as described in (Spradling and Rubin, 1982) using a Transjector-5246 (Eppendorf) and micromanipulator (Leica) with a 0.5 µm Femtotip (Eppendorf). Manually dechorionated *y w* embryos were dehydrated for between 1 – 5 min in a chamber containing silica ChIPs and covered with halocarbon-700 oil (Sigma). Plasmid DNA was injected into the posterior end of the embryo. Embryos were placed in a high humidity environment at 18°C for 24 h, followed by 22°C for 24 h. Surviving first instar larvae were transferred to a vial containing standard cornmeal/molasses food without additional yeast. These flies are referred to as G_0.

2.5.3.3 Crossing of adult survivors and identification of transformants

G_0 adults were mated to *y w* flies individually at 25°C. The G_1 progeny resulting from this cross were examined for the marker (w^+). Transformant males and virgin females were crossed to *y w* flies individually at 25°C. The G_2 heterozygous progeny were crossed to establish a transformant stock.

2.5.4 Collection of developmental stages of *Drosophila*

2.5.4.1 Collection of *Drosophila* staged embryos, larvae, pupae and adult flies

Different developmental stages were collected by aging newly laid embryos collected from a large population cage for the appropriate period of time at 25°C (Ashburner, 1989a).

2.5.4.2 Collection of synchronous third instar larvae

Synchronous third instar larvae were collected using the blue food method (Maroni and Stamey, 1983). Bromophenol blue was added to standard cornmeal/molasses food at 0.05% (w/v). Climbing late third instar larvae raised on this food were collected. Larvae stop feeding late in third instar development, hence can be separated into 3 distinct classes according to gut colour: dark blue/purple; light blue; white. Light blue larvae were selected, as these are synchronous to within 4 h of development.

2.6 Molecular Biology methods

2.6.1 RNA

All solutions and equipment for use with RNA were prepared in an RNase-free manner. All solutions except those containing Tris, SDS, or solvents were treated with DEPC as described in (Sambrook *et al.*, 1989). All solutions containing RNA were treated with RNAsecure™ (Ambion) to permanently inactivate RNases at 60°C for 20 min.

2.6.1.1 Isolation of RNA from *Drosophila*

Eggs previously ground in liquid nitrogen, larvae or adult flies were homogenized in 1 ml of TRIzol® reagent (Invitrogen) in a Kontes Glass homogenizer, and transferred to a 1.5 ml microcentrifuge tube and incubated for 5 min at room temperature. Two hundred µl of chloroform was added, mixed vigorously, and incubated for 3 min at room temperature, followed by centrifugation at 12 000 *g* for 10 min. The upper aqueous phase was transferred to a fresh tube, to which was added 500 µl of isopropanol. After 10 min incubation at room temperature, the RNA was pelleted by centrifugation at 12 000 *g* for 10 min. The pellet was washed in 70% (v/v) ethanol, recentrifuged at 7500 *g* for 10 min, and air-dried for 10 min at room temperature. RNA was resuspended in 96 µl of DEPC-treated water and 4 µl of RNA*secure*™ (Ambion). RNases were inactivated permanently by incubation of this solution at 60°C for 20 min.

2.6.1.2 DNase treatment of RNA for RNase protection analysis

Following RNA isolation (section 2.6.1.1), 1 µl of RNase-free DNase (Roche) was incubated with the RNA solution at 37°C for 15 min. One hundred µl of TE Buffer was added, and the RNA extracted using 1 volume of phenol:chloroform:isoamylalcohol (25:24:1). This solution was vortexed for 1 min then centrifuged to separate phases at 12 000 *g* for 2 min. The upper aqueous phase was transferred to a fresh tube, and any remaining phenol removed using 1 volume of chloform:isoamylalcohol (24:1). This solution was also vortexed for 1 min then centrifuged to separate the phases at 12 000 *g* for 2 min. The aqueous phase was transferred to a fresh tube, and the RNA precipitated using 0.1 volume 3 M NaOAc and 2.5 volumes of ice-cold absolute ethanol at -20°C for 30 min. The RNA was pelleted at 12 000 *g* for 20 min, then washed in 70% (v/v) ethanol and air dried for 10 min. The RNA was resuspended in 96 µl of DEPC-water and 4 µl of RNA*secure*™ (Ambion). Any RNases present in the solution were

inactivated by heating to 60°C for 20 min. RNA was stored for short-term use at -20°C, and for longer-term use in 2.5 volumes of absolute ethanol at -80°C.

2.6.1.3 DNase treatment of RNA for real-time RT-PCR analysis

Following RNA isolation (section 2.6.1.1), 2 µl of Turbo™ DNase (Ambion) and 11 µl of 10 X DNase Buffer (Ambion) were incubated with 100 µl of RNA at 37°C for 30 min. One µl of Turbo™ DNase (Ambion) was added after this time, and incubated at 37°C for a further 30 min. Eleven µl of the DNase Inactivation Reagent (Ambion) was added, mixed, and incubated at room temperature for 2 min. The reaction was centrifuged at 10 000 *g* for 1.5 min, and the RNA solution transferred to a fresh tube.

2.6.1.4 Quantification of RNA with Ribogreen kit

Three separate 100 fold dilutions of each RNA sample were quantified using the Ribogreen dye (Molecular Probes) according to the manufacturer's instructions. Fluorescence intensity was measured on the Fluostar Galaxy (BMG Labtechnologies).

2.6.1.5 Purification of poly(A)$^+$ RNA from total RNA preparations

Poly(A)$^+$ RNA was isolated from total RNA preparations using oligo (dT) cellulose (Roche) according to the manufacturer's instructions and the protocol of Sambrook *et al.* (1989; 7.26-7.29). All fractions and final poly(A)$^+$ RNA concentrations were determined using the RiboGreen kit (section 2.6.1.4).

2.6.1.6 Northern hybridization analysis

Northern hybridization analysis was conducted based on the formaldehyde-containing gel method of Sambrook *et al.* (1989; 7.43). Samples were electrophoresed along side the 0.24 - 9.5 kb RNA ladder (Invitrogen). RNA was transferred to Hybond™-XL nylon membrane (Amersham Pharmacia Biotech) in 20 X SSC for 18 h at room temperature, and UV-crosslinked using the Ultraviolet Crosslinker (Ultra-LUM Inc.). Probe was prepared using method B of the Ready-to-go™ beads kit (-dCTP) (Amersham Pharmacia Biotech). Two hundred µl of gel-purified DNA fragment (200 ng) was incubated with 5 µl of ^{32}P-[α]-dCTP (PerkinElmer™) using this kit at 37°C for 1 h. Unincorporated nucleotides were removed from the labeled probe using the ProbeQuant™ G-50 Micro column according to the manufacturer's instructions

(Amersham Pharmacia Biotech). DNA was denatured by boiling for 1 min then 100 ng of denatured probe with specific activity > 2 x 10^8 cpm/μg was hybridized to membrane at 65°C for 18 h in 6 x SSC, 2 x Denhardt's reagent, 0.1% (w/v) SDS, 100 μg/ml salmon sperm DNA. Membranes were washed in 1 x SSC, 0.1% (w/v) SDS at room temperature for 20 min, followed by 3 washes in 0.2 x SSC, 0.1% (w/v) SDS at 68°C for 20 min each. All products were autoradiographed on BioMax™ MS film (Kodak Scientific Imaging) at -70°C for 1 – 18 h using an intensifying screen. Membranes were stripped where required in boiling 0.1% (w/v) SDS for 6 – 18 h with shaking.

2.6.1.7 *In vitro* transcription of RNA probes for RNase protection analysis

In vitro transcription of RNA was conducted based on the method specified by the manufacturers of the polymerases (Roche). Plasmid template was prepared using a CsCl gradient (section 2.6.2.4), linearized with a restriction endonuclease that did not generate a 3' overhang, and gel-purified using the QIAquick® Gel Extraction kit (Qiagen). Half a μg of this template DNA was added to 0.5 μl each of 10 mM ATP, GTP and UTP, 1 μl of 10 x transcription buffer (Roche), 0.4 μl of RNAsecure™ (Ambion), 0.5 – 1 μl of 0.1 or 1 mM CTP to generate high or low specific activity probes respectively (Table 4), and enough DEPC-treated water to make up a final volume of 10 μl. These components were heated to 60°C for 20 min to inactivate RNases and cooled. To this, 1.3 – 2.6 μl of 800 Ci/mmol ^{32}P-[α]-CTP (20 mCi/ml) (PerkinElmer™) and 0.5 μl (20 units) of the appropriate polymerase was added (Roche) and incubated at 37°C for 1 h. Following this the plasmid template was removed with 1 μl of RNase-free DNase (Roche) at 37°C for 15 min. After inactivation of the DNase with 1 μl of 0.5 M EDTA, unincorporated radionucleotide was removed using the NucAway™ spin column (Ambion). Integrity of the probe was confirmed by electrophoresis on a 5% (v/v) polyacrylamide/8 M urea gel. Only two probes, those for *CG15926* and *CG2556* required gel purification as described in the RPA III™ kit (Ambion).

Template	Probe length	Protected fragment(s)	Polymerase	Unlabelled CTP added	^{32}P-[α]-CTP added
CG15926	400 nt	325 nt	T3	1 µl 0.1 mM	2.6 µl
CG2560	449 nt	366 nt	T7	0.5 µl 0.1 mM	1.3 µl
Lsp1α	565 nt	391 nt + 294 nt	SP6	0.5 µl 1 mM	1.3 µl
CG2556	343 nt	268 nt	T3	1 µl 0.1 mM	2.6 µl
Pgd	431 nt	43 nt + 171 nt	T7	1 µl 0.1 mM	2.6 µl
rp49	375 nt	312 nt	T3	1 µl 1 mM	1.3 µl

Table 4. Summary of the templates used for the *in vitro* transcription of antisense ssRNA probes for RNase protection analysis

2.6.1.8 RNase protection analysis

RNase protection analysis was conducted using the RPA III™ kit (Ambion) according to the manufacturer's instructions. Two minor modifications were made to this method. Samples were precipitated with pellet paint® co-precipitant (Novagen) to assist with visualization of the pellet in the initial co-precipitation of probe and RNA. Samples were resuspended in hybridization buffer to which RNA*secure*™ (Ambion) had been added, and RNases inactivated at 60°C for 20 min. All RNA samples used in RNase protection were treated with DNase (section 2.6.1.2). Probes produced by *in vitro* transcription (section 2.6.1.7) were added in 3 – 10 fold molar excess over target mRNA as specified by the manufacturer's instructions. All products were electrophoresed using the Mini PROTEAN® 3 Cell (BioRad) on a 5% (v/v) polyacrylamide/8M urea gel at 140 V for 1 – 1.5 h. Product size was determined by comparison to an RNA ladder generated by *in vitro* transcription of the Century™ Marker Plus template set (Ambion).

2.6.1.9 Reverse transcription

Reverse transcription was carried out according the method specified by the manufacturer of the expand reverse transcriptase (Roche). One µg of total RNA or 100 ng of poly(A)$^+$ RNA was converted to cDNA using random hexamer primers (Roche).

2.6.2 DNA

2.6.2.1 Molecular size markers

Lambda DNA digested with *Hind*III and *Sac*II was used as molecular size markers for most applications. This consists of 20.3 kb, 9.4 kb, 3.7 kb, 2.8 kb, 2.3 kb, 2.0 kb, 1.5 kb, 1.1 kb and 0.5 kb DNA fragments. The 100 bp molecular size markers (GC-105-004; GeneCraft, Germany) consist of 1.6 kb, 1 kb, 0.9 kb, 0.8 kb, 0.7 kb, 0.6 kb, 0.5 kb, 0.4 kb, 0.3 kb 0.2 kb and 0.1 kb DNA fragments.

2.6.2.2 Isolation of DNA from adult *Drosophila*

Two different methods were used for the isolation of DNA. High quality DNA was isolated from adult *Drosophila* as described in protocol 47 (Ashburner, 1989b). DNA for inverse PCR mapping of transgene inserts was isolated as described in the "Inverse PCR and cycle sequencing of *P*-element insertions for STS generation" protocol provided by the BDGP resources (Rehm, 2004).

2.6.2.3 Southern hybridization analysis

Six µg of high quality genomic DNA (section 2.6.2.2) was digested with 12 units of either *Eco*RV or *Cla*I in 100 µl volume containing the appropriate buffer for 1 h at 37°C. Following this, 12 further units of enzyme was added and incubated for 14 h at 37°C. The digested DNA was purified by phenol/chloroform extraction and ethanol precipitation and resuspended in TE. Southern hybridization was conducted as described in Ausubel *et al.* (1997). Probe was prepared using method B of the Ready-to-go™ beads kit (-dCTP) (Amersham Pharmacia Biotech). Two hundred µl of gel-purified DNA fragment (200 ng) was incubated with 5 µl of ^{32}P-[α]-dCTP (PerkinElmer™) using this kit at 37°C for 1 h. Unincorporated nucleotides were removed from the labeled probe using the ProbeQuant™ G-50 Micro column according to the manufacturer's instructions (Amersham Pharmacia Biotech). Membranes were prehybridized in 0.4 M Na_2HPO_4, 50 mM NaH_2PO_4, 7% SDS, 1% BSA, 1 mM EDTA (pH 8.0) for 2 h at 65°C, followed by hybridization in the same buffer at 65 - 68°C for 18 h with 100 ng of denatured probe. Membranes were washed 4 times in 0.1 x SSC, 0.1% SDS at 65°C for 30 – 60 min. All products were autoradiographed on BioMax™ MS film (Kodak Scientific Imaging) at -70°C for 1 – 18 h using an intensifying screen,

and also analysed using the FLA-5000 phosphorimager (Fujifilm). Membranes were stripped where required in boiling 0.1% (w/v) SDS for 6 – 18 h with shaking.

2.6.2.4 Plasmid DNA

Small-scale alkaline lysis preparation of plasmid DNA for analysis of cloning was performed as described in Unit 1.25 of Sambrook *et al.* (1989). Medium-scale preparation of plasmid DNA for cloning was performed using the Qiagen® Plasmid Midi Kit (Qiagen). Large scale alkaline lysis preparation of plasmid DNA was prepared as described in Unit 1.7.1 of Current Protocols in Molecular Biology (Ausubel *et al.*, 1997). CsCl-ethidium bromide gradient purification of plasmid DNA for use as a template for *in vitro* transcription (section 2.6.1.7) or for microinjection (section 2.5.3.2) was performed according to Unit 1.42 of Sambrook *et al.* (1989). Ultracentrifugation was conducted in a Sorvall Combi ultracentrifuge using a Sorvall TV-865 rotor at 55 000 rpm for 18 h at room temperature (Sambrook *et al.*, 1989).

2.6.2.5 Restriction endonuclease digestion

Most restriction endonucleases were supplied by New England Biolabs (NEB). However, the following enzymes were supplied by alternative companies: *Asp*718 (Roche), *Kpn*I (Invitrogen). Restriction endonuclease digestion was conducted according to the manufacturer's instructions. Typically, preparative digests were conducted on 5 μg of plasmid DNA with 2 μl of the appropriate enzyme at 37°C for 3 – 5 h. Diagnostic digests of alkaline lysis minipreps were conducted on 5 μl of plasmid DNA using 0.3 μl of the appropriate enzyme at 37°C for 1 h.

2.6.2.6 Dephosphorylation of 5' ends

Following restriction endonuclease digestion, 15 units (1.5 μl) of calf intestinal alkaline phosphatase (NEB) was added to the digested plasmid DNA, and the reaction incubated at 37°C for 1 h. Heat inactivation was not generally conducted, as plasmid DNA was typically recovered using gel purification.

2.6.2.7 Polynucleotide kinase treatment of linker DNA

Annealed linkers were treated with 0.5 µl of T4 polynucleotide kinase (NEB) to add phosphate groups to the 5' end of the DNA, with 1 mM ATP in buffer supplied by the manufacturer at 37°C for 1 h, followed by heat inactivation at 65°C for 20 min.

2.6.2.8 Blunt end production

Klenow fragment of *E. coli* DNA polymerase (5'→3' exo$^-$) (NEB) or T4 DNA polymerase (NEB) were used according to the manufacturer's instructions to form blunt ends on plasmid DNA previously digested using restriction endonucleases. Typically, 9 units (3 µl) of T4 DNA polymerase was added to 5 µg of digested plasmid DNA (in one of the 4 standard NEB buffers immediately following digestion), supplemented with dNTPs at 100 µM final concentration and incubated at 12°C for 20 min. Reactions were heat inactivated at 75°C for 10 min.

2.6.2.9 Ligation

Prepared plasmid and insert DNA were ligated together using 1 µl (400 units) of T4 DNA ligase (NEB) in a final reaction volume of 10 µl with the appropriate buffer at 16°C for *ca.* 16 h according to the manufacturer's instructions

2.6.2.10 Gel purification

Amplified products from PCR or restriction fragments were subjected to gel electrophoresis in 0.8 – 1.5% (w/v) agarose, excised, and gel purified using the Qiaquick® gel extraction kit (Qiagen) according to the manufacturer's instructions.

2.6.3 PCR

PCR was carried out according the method specified by the manufacturer of the polymerase (Roche). Fragments <3 kb were amplified using *Taq* DNA polymerase (Roche). One hundred ng of genomic DNA or 5 µl cDNA was used as a template in a 50 µl reaction volume with 5 µl of 10X reaction buffer (Roche), 0.4 µM of each specific primer and 0.2 mM of dNTPs (Roche). A typical reaction (method 1) consisted of 1 cycle at 94°C for 2 min, followed by 30 - 35 cycles of: 94°C for 30 s; annealing temperature for 30s; extension at 72°C for 45 s – 2 min depending on the size of the

amplified product. This was followed by 7 min at 72°C to ensure that amplified products were full-length. To increase the yield of the amplified product, conditions were modified in certain reactions (method 2). The first 10 cycles were conducted as described previously, followed by 20 - 25 further cycles in which the extension at 72°C was increased by 5 s in each subsequent cycle.

Long range PCR of fragments > 3 kb was conducted using the expand long template PCR system (Roche). One hundred ng of genomic DNA was used as a template in a 50 µl reaction volume with 5 µl of the specified 10X reaction buffer (Roche), 0.3 µM of each specific primer and 0.4 mM of dNTPs. A typical reaction (method 3) consisted of 1 cycle at 92°C for 2 min, followed by 10 cycles of: 92°C for 10 s; annealing temperature for 30 s; extension at 68°C for 4 - 8 min depending on the size of the amplified product. This was followed by 25 cycles of: 92°C for 10 s; annealing temperature for 30 s; extension at 68°C for 4 - 8 min plus 20 s each subsequent cycle. A final extension at 7 min at 68°C followed these cycles to ensure that amplified products were full-length. The primer sequences used in PCR are described in appendix 10.1.

2.6.3.1 Inverse PCR determination of the site of transgene insertion

Inverse PCR using primers specific for *P*-element vectors and genomic DNA from transformed lines was conducted according to Rehm (2004). The primer sequences used for inverse PCR and sequencing of the resulting inverse PCR products are described in appendix 10.2. *Hin*P1I proved the most efficient enzyme for producing an inverse PCR product suitable for sequencing.

2.6.4 Transformation of *E. coli*

DH5α *E. coli* was used for all cloning procedures unless otherwise stated. High-efficiency competent cells were prepared (Inoue *et al.*, 1990). Competent *E. coli* were transformed with plasmid (10 µl of ligation reaction or 1 µl of plasmid solution) using a heat shock protocol (Inoue *et al.*, 1990). Max Efficiency® STBL2™ competent cells (GibcoBRL) were used, according to the manufacturer's instructions, for cloning of plasmid DNA that had a tendency to rearrange, for example the pW30 plasmid.

2.6.5 Sequencing

All sequencing was conducted by the Allan Wilson Centre Sequencing Facility (Massey University). The primer sequences used for sequencing are described in appendix 10.3. Plasmid DNA was purified using the QIAquick® Spin Miniprep kit (Qiagen), and PCR products purified using the QIAquick® PCR purification kit (Qiagen), and quantified using agarose gel electrophoresis prior to sequencing.

2.6.6 Chromatin immunoprecipitation

The chromatin immunoprecipitation method was based on (Miura *et al.*, 1999) and (Deryckere and Gannon, 1994) (fat body isolation); (Sachs and Shi, 2000) (crosslinking of protein to DNA); Tim Parnell & Pamela Geyer, (immunoprecipitation; University of Iowa, 2004, personal communications); with additional modifications suggested by Edwin Smith (Rockefeller University, 2005, personal communications). All steps were performed on ice using pre-chilled containers and solutions, and all centrifugation steps were performed using a refrigerated centrifuge.

Flies of the appropriate genotype were reared on bromophenol blue food in bottles at 25°C. Two hundred male climbing third instar larvae were collected and washed in Ringers solution, following which their fat bodies were dissected manually and frozen immediately in liquid nitrogen. Fat bodies were ground in liquid nitrogen using a mortar and pestle, followed by homogenization in a Dounce homogenizer with a loose pestle in 5 ml of Buffer A (10 mM Hepes, pH 7.6; 1 mM EDTA; 150 mM NaCl; 0.6% (v/v) Triton X-100 (BDH); 4 mM DTT; 10 mM sodium butyrate; 1X protease inhibitor cocktail (Roche)). The homogenate was centrifuged at 500 *g* for 30 s at 4°C in a 15 ml Corex tube to remove tissue debris. The supernatant was transferred to a fresh Corex tube and stored on ice for 5 min, followed by centrifugation at 1500 *g* for 10 min at 4°C. The pelleted nuclei were resuspended in 360 µl of ice-cold nucleus isolation buffer (0.25 M sucrose; 10 mM Tris-HCl, pH 7.4; 3 mM $CaCl_2$; 10 mM sodium butyrate; 1X protease inhibitor cocktail (Roche)) in a microtube.

Formaldehyde was incubated with the nuclei resuspension to a final concentration of 1% (v/v) for 10 min at room temperature with shaking (10 µl of 37% formaldehyde per 360 µl of nuclei resuspension). The nuclei were pelleted at 1500 *g* for 10 min at 4°C,

resuspended in 360 µl of nucleus isolation buffer, and pelleted again at 1500 x *g* for 10 min at 4°C. The pelleted nuclei were resuspended in 200 µl of lysis buffer (1% (w/v) SDS; 50 mM Tris-HCl, pH 8.1; 10 mM EDTA; 10 mM sodium butyrate; 1X protease inhibitor cocktail (Roche)) for 10 min on ice.

The nuclei lysis solution was sonicated in a 0.5 ml tube on ice in the presence of *ca.* 50 µl volume of 425-600 micron acid-washed glass beads (Sigma) for 6 x 30 s pulses at power level 1.5 using a VirSonic sonicator. The lysate was separated from the beads by punching a small hole in the base of the 0.5 ml tube and pulse spinning its contents into a 1.5 ml microtube. The sonicated lysate was diluted with 1.8 ml of chromatin immunoprecipitation buffer (25 mM Tris-HCl, pH 8.0; 137 mM NaCl; 2.7 mM KCl; 1% (v/v) Triton X-100 (BDH); 1 mM EDTA; 10 mM sodium butyrate) and centrifuged at 14,000 *g* for 10 min at 4°C. Two hundred µl of the supernatant was reserved for input DNA purification. The remaining 1.8 ml of chromatin supernatant was reserved for immunoprecipitation.

The 200 µl of input DNA was incubated with 10 µl of 10 mg/ml RNase for 10 min at 37°C, followed by the addition of 20 µl of 10 mg/ml proteinase K for 6 h at 37°C. The crosslinks were reversed by incubation for 6 h at 65°C. The DNA was purified using the QIAquick® PCR purification kit (Qiagen) and eluted in 50 µl of the elution buffer.

The 1.8 ml of chromatin supernatant was pre-incubated with 60 µl of a 50% (v/v) resuspension of Protein A Sepharose Beads (Sigma) in chromatin immunoprecipitation buffer, in the presence of 1X protease inhibitor cocktail (Roche) and 9 µl of 10 mg/ml high quality salmon sperm DNA (100 µg/ml final concentration) for 1 h with shaking at 4°C. Simultaneously, 6 µl of the rabbit anti histone H4 (Ac16) antibody (AHP417; Serotec) was pre-bound to 120 µl of the 50% (v/v) resuspension of Protein A Sepharose Beads (Sigma) in chromatin immunoprecipitation buffer, in the presence of 1X protease inhibitor cocktail (Roche), 10 µl of 10 mg/ml high quality salmon sperm DNA (100 µg/ml final concentration), and 20 µl of 10 mg/ml BSA (NEB) in 1 ml of chromatin immunoprecipitation buffer for 1 h with shaking at 4°C.

The antibody-bound beads were pelleted at 1000 rpm for 30 s at 4°C, washed in chromatin immunoprecipitation buffer, and pelleted at 1000 rpm for 30 s at 4°C. The chromatin was separated from the beads at 1000 rpm for 30 s at 4°C, and the supernatant incubated with the antibody-bound beads for 3 h at 4°C with shaking. Following this, the chromatin-antibody-bead solution was pelleted at 1000 rpm for 30 s at 4°C into a 0.5 ml tube, washed and pelleted several times in chromatin immunoprecipitation buffer, and pelleted at 1000 rpm for 30 s at 4°C. The beads were then washed successively for 5 min at 4°C with shaking, followed by centrifugation at 1000 rpm for 30 s at 4°C to pellet the beads in: Low salt wash buffer (150 mM NaCl; 20 mM Tris-HCl, pH 8.0; 2 mM EDTA; 1% (v/v) Triton X-100 (BDH); 0.1% (w/v) SDS; 10 mM sodium butyrate); high salt wash buffer (500 mM NaCl; 20 mM Tris-HCl, pH 8.0; 2 mM EDTA; 1% (v/v) Triton X-100 (BDH); 0.1% (w/v) SDS; 10 mM sodium butyrate); LiCl wash buffer (250 mM LiCl; 10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 1% (w/v) sodium deoxycholate; 1% (v/v) IGEPAL CA-630 (Sigma); 10 mM sodium butyrate); TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA). The beads were rinsed in TE Buffer, pelleted at 1000 rpm for 30 s at 4°C and resuspended in 100 µl of TE Buffer.

The immunoprecipitated DNA was purified from the 100 µl of resuspended beads by the addition of 1 µl of 10 mg/ml RNase for 10 min at 37°C, followed by incubation with 5 µl of 10% (w/v) SDS and 2 µl of 10 mg/ml proteinase K for 6 h at 37°C. Crosslinking was reversed by incubation for 6 h at 65°C. The DNA was purified using the QIAquick® PCR purification kit (Qiagen) and eluted in 50 µl of the elution buffer.

2.6.7 Quantitative real-time PCR

Quantitative real-time PCR was conducted using the LightCycler FastStart DNA Master^{PLUS} SYBR Green I reaction mix (Roche) in a LightCycler Instrument (Roche) as per the manufacturer's instructions. Two µl of template, 5 µl of PCR grade water, 0.5 µl of each primer (10 µM) and 2 µl of the LightCycler FastStart DNA Master^{PLUS} SYBR Green I reaction mix (Roche) were added to each capillary. For chromatin immunoprecipitation assays the template consisted of either undiluted immunoprecipitated DNA or 100-fold diluted input DNA. For RT-PCR analyses, the template consisted of 10-fold or 100-fold diluted cDNA. Each PCR reaction was conducted in triplicate. The LightCycler protocol used is described in Table 5. The

channel setting was F1 and the crossing point automatically determined by the LightCycler software (Roche).

Program:	**Denaturation**			**Cycles: 1**
Segment number	*Temperature (°C)*	*Hold time (s)*	*Slope (°C/s)*	*Acquisition mode*
1	95	600	20	None
Program:	**Amplification**			**Cycles: 45**
1	95	10	20	None
2	55	10	20	None
3	72	12	20	Single
Program:	**Melting Curve**			**Cycles: 1**
1	95	0	20	None
2	65	15	20	None
3	95	0	0.1	Continuous
Program:	**Cooling**			**Cycles: 1**
1	40	30	20	None

Table 5. The LightCycler protocol used in quantitative real-time PCR

2.7 Enzyme assays

2.7.1 β-galactosidase assays

β-galactosidase activity was measured as described in (Fitzsimons *et al.*, 1999). Third instar larvae of all transgenic lines were staged using the bromophenol blue method (section 2.5.4.2) and sexed conventionally. Larvae were rinsed twice in Ringers solution, and once in water prior to collection in microcentrifuge tubes for assays. β-galactosidase activity and protein levels were both measured immediately after the collection of larvae. Due to the difference in promoter activity between the *Lsp1α*, *Pgd* and *armadillo* promoters used, varying amounts of homogenate were used in the β-galactosidase assays (Table 6). Values were normalized to protein levels, but not to wet weight as this was not consistent due to washing of larvae. The β-galactosidase activity and protein levels of each fly line were determined for at least 3 separate collections of larvae emerging from ≥3 separate vials of bromophenol blue food. For each of these 3 separate larvae samples, β-galactosidase activity and protein levels were measured 3

times. β-galactosidase activity was measured in hemisected adults exactly as described in (Fitzsimons *et al.*, 1999).

Promoter	Harvested larvae	Volume homogenate assayed	
		homozygous line	heterozygous lines
I or *Lsp1α*	5	5 – 15 µl	15 - 25 µl
Pgd	10	200 µl	-
armadillo	5	25 µl	50 µl

Table 6. Volume of homogenate assayed for β-galactosidase activity

2.7.2 Protein concentration determination

The protein concentration was measured as described in (Fitzsimons *et al.*, 1999) using the Bio-Rad protein assay (BioRad).

3 STAGE AND DOSAGE COMPENSATION STATUS DETERMINATION

3.1 cDNA and genomic clones utilized in northern and RNase protection analyses

3.1.1 Acquired cDNA and genomic clones

Clones for the genes *CG15926*, *CG2560*, *CG15730* and *CG2556* were obtained from ResGen™ (Invitrogen) as part of the Berkeley *Drosophila* Genome Project (Rubin *et al.*, 2000) (Table 7). These clones consist of oligo-dT primed cDNA clones integrated into either the pFLC-1 (section 10.5.13) or pOTB7 (section 10.5.14) vectors in the sense (pFLC-1) or antisense (pOTB7) direction from the T7 polymerase binding site. The identity of these clones was confirmed using diagnostic restriction endonuclease digests and sequencing. All the clones with the exception of *CG15730* were correctly identified. Restriction endonuclease digestion of the *CG15730* clone did not produce the expected fragments, and the sequence of the clone could not be obtained using the M13 forward or reverse primers. This clone was consequently not used for any subsequent work.

A genomic clone for *Pgd* consisting of the first 1461 bp of the gene including the 5' UTR, first exon and intron, and partial second exon was obtained from Dr. Max Scott (Scott and Lucchesi, 1991). This clone is integrated in the *Cla*I and *Eco0109*I restriction endonuclease sites of the pBluescript II KS vector in the antisense orientation from the T7 polymerase binding site. The identity of this clone was confirmed by diagnostic restriction endonuclease digestion and sequencing.

A genomic clone for *rp49* consisting of the first 640 bp including the first exon and intron, and part of the second exon was obtained from Dr. Max Scott (O'Connell and Rosbash, 1984). This clone is integrated in the *Eco*RI and *Hind*III restriction endonuclease sites of the pBluescript II KS vector in the sense orientation from the T7 polymerase binding site.

3.1.2 Experimentally obtained cDNA and genomic clones

3.1.2.1 *CG15730* in pBluescript II KS

The possible incorrect identity of the *CG15730* clone obtained from ResGen™ (Invitrogen) necessitated cloning of *CG15730*. A 1478 bp region of *CG15730* was amplified from male adult cDNA with the AT20771F and AT20771R primers (section 10.1) using two rounds of PCR (method 2: annealing temperature of 54°C: 30X). Male adult cDNA was used as the template for PCR of *CG15730* because the only clone of *CG15730* in the *Drosophila* genome resource had been obtained from an adult testes cDNA library. Further studies conducted after this cloning, indicate that the real template for this PCR reaction was probably contaminating genomic DNA present in the cDNA (section 3.2.1). The 1478 bp fragment of *CG15730* was digested using *Bam*HI and *Hind*III generating an 1174 bp fragment. This fragment was gel purified and ligated into pBluescript II KS that had previously been digested with both of these enzymes, and the resulting products transformed into competent *DH5α E. coli* cells which were plated on selective LB-IPTG/X-gal/amp plates and grown at 37°C overnight. Clones consisting of pBluescript with the 1174 bp *CG15730* insert were identified using restriction endonuclease digestion and confirmed using sequencing.

3.1.2.2 *Lsp1α* into pBluescript II KS

A clone containing *Lsp1α* could not readily be obtained, thus requiring the cloning of this gene during the initial phase of the project. A 2344 bp region of *Lsp1α* was amplified from third instar larval cDNA with the Forw-Lsp1α and Rev-Lsp1α primers (section 10.1) using two rounds of PCR (method 2: annealing temperature of 55°C: 30X). Previous evidence had demonstrated that *Lsp1α* is maximally expressed during the third instar larval phase of development (Roberts *et al.*, 1977). This 2344 bp fragment was digested using *Pst*I and *Sac*I to produce an 853 bp fragment which was then gel purified and ligated into pBluescript II KS that had been previously digested with both of these restriction endonuclease enzymes. The ligation products were transformed into competent *DH5α E. coli* cells, which were plated on selective LB-IPTG/X-gal/amp plates and grown at 37°C overnight. Clones consisting of pBluescript II KS with the 853 bp *Lsp1α Pst*I/*Sac*I insert obtained by restriction endonuclease digestions following RT-PCR of third instar larval cDNA were identified using

restriction endonuclease and confirmed using sequencing. This clone was used for subsequent northern analyses.

3.1.2.3 Cloning of *Lsp1α* 3' UTR into pGEM®-T Easy

The 3' UTR and surrounding region of *Lsp1α* was amplified by PCR of genomic *Drosophila* DNA in order to provide a template that differed from *Lsp1β* and was suitable for RNase protection analysis. The 3' UTR and surrounding region of *Lsp1α* was amplified from genomic *Drosophila* DNA by PCR with the Lsp1αfor3'reg and Lsp1αrev3'reg primers (section 10.1) (method 1: annealing temperature of 45°C: 30X). The 575 bp product obtained from the *y w* strain of *Drosophila* was significantly smaller than the 691 bp product expected from the *y2; cn bw sp* strain used in generating the genomic BDGP sequence. This 575 bp PCR product was gel purified, ligated directly into the pGEM®-T Easy vector (Promega; section 10.5.9), and transformed into *DH5α E. coli* cells. Restriction endonuclease digestion of the clone obtained indicated that the sequence of this region differed significantly from that expected from the BDGP genomic sequence. Sequencing of the 575 bp insert confirmed this. The sequence is highly homologous to the BDGP genomic sequence in the coding sequence and probable 3' UTR, however has a 118 bp deletion following the putative poly(A)$^+$ site of *Lsp1α* (section 10.6.1). There are also a large number of SNPs in this intragenic region.

3.1.2.4 Subcloning of *CG2560* fragment into pBluescript II KS

The *CG2560* clone did not provide a good template for *in vitro* transcription using T3 polymerase. Hence a 366 bp *Cla*I/*Pst*I fragment of the gene was subcloned into pBluescript II KS to provide an optimal template for *in vitro* transcription. The restriction endonucleases *Cla*I and *Pst*I were used to excise the 366 bp fragment of *CG2560* from the *CG2560*/pFLC-1 clone. This fragment was gel purified and ligated into pBluescript II KS in the antisense direction from the T7 polymerase binding site, and transformed into *DH5α E. coli* cells. The identity of the clone was confirmed using diagnostic restriction endonuclease digests and sequencing.

Gene	cDNA/ Genomic	vector	Antibiotic-resistance	Orientation from T7 site
CG15926	RE29468	pFLC-1	ampicillin	Sense
CG2560	RE57452	pFLC-1	ampicillin	Sense
CG15730[1]	AT20771	pOTB7	chloroamphenicol	Antisense
CG2556	RE03921	pFLC-1	ampicillin	Sense
Pgd	Genomic	pBluescript II KS	ampicillin	Antisense
rp49	Genomic	pBluescript II KS	ampicillin	Sense

Table 7. cDNA and genomic clones acquired from ResGen or Dr. Scott

Gene	insert	vector	Antibiotic-resistance	Orientation from T7 site
CG15730	cDNA	pBluescript II KS	ampicillin	Sense
Lsp1α	cDNA	pBluescript II KS	ampicillin	Antisense
Lsp1α	3' UTR	pGEM®-T Easy	ampicillin	Sense
CG2560	cDNA	pBluescript II KS	ampicillin	Antisense

Table 8. cDNA and genomic clones constructed for northern and RNase protection analyses

[1] This clone was not used for northern RNA hybridization or RNase protection analysis.

3.2 Stage-specific expression of genes within the *Lsp1α* gene domain

The expression of the genes within the *Lsp1α* gene domain was examined with regard to developmental stage in order to ascertain the optimal developmental stage in which to determine the dosage compensation status of each gene. The following developmental stages were examined in *y w* flies: embryos harvested 0 – 2 h after laying, embryos harvested 12 h after laying, first instar larvae, second instar larvae, third instar larvae, pupae, male adults and female adults (section 2.5.4.1). Between 1 – 2 mg of RNA was extracted from multiple collections of each developmental stage. The *Lsp1α* and *CG2560* mRNAs are relatively abundant and are detectable by northern RNA hybridization analysis with 10 μg of total RNA (Figure 4, Figure 6). The *CG15926* and *CG2556* mRNAs could not be readily detected with total RNA and thus poly(A)$^+$ RNA was purified from 0.5 – 1.0 mg of total RNA using oligodT cellulose (section 2.6.1.5). The *rp49* mRNA was used as a loading control for this northern RNA hybridization analysis, and the developmental expression pattern observed for this gene is consistent with previously published northern analysis despite the overexposure of the film in these experiments (O'Connell and Rosbash, 1984). *CG15730* was not detected with 2 μg of poly(A)$^+$ RNA by northern RNA hybridization analysis (data not shown). However, *CG15730* can be detected in both male (not shown) and female adult RNA if larger amounts of poly(A)$^+$ RNA (15 – 20 μg) are used in northern RNA hybridization analysis (Figure 5). This product is larger than that expected for the gene, and may result from contaminating genomic DNA, or from non-specific cross-hybridization. All RNA used in northern RNA hybridization analysis was of high quality as determined by agarose gel electrophoresis and ethidium bromide staining (not shown).

CG15926 encodes three transcripts of approximately 2 kb, 3 kb and 4 kb (Figure 4). These transcripts are expressed differentially throughout development, with maximal expression of the more abundant 2 kb and 3 kb transcripts in embryos 0 – 2 h after laying, first and second instar larvae, and both male and female adults. The 4 kb transcript is much less abundant than the two smaller transcripts, and exhibits a slightly different expression pattern, peaking in embryos 12 h after laying, and again in adults. This gene was shown to be upregulated in male adults (section 3.5), and hence the dosage compensation status of this gene could not be determined using RNase protection analysis.

CG2560 encodes a single abundant transcript of approximately 1.4 kb, which is expressed in first, second and third instar larvae (Figure 4). The levels of the *CG2560* transcript reach a maximum in first instar larvae, and decline in second and third instar larvae. Due to the peak of expression of *CG2560* in first instar larvae, the dosage compensation status of this gene has been examined in first instar larvae (section 3.5).

CG2556 encodes four transcripts of approximately 2 kb, 4.3 kb, 4.5 kb and 4.7 kb (Figure 4). These transcripts are expressed differentially throughout development, with the largest 4.7 kb transcript being expressed at much lower levels than the smaller three transcripts. The 2 kb transcript is the sole transcript detected in 0 – 2 h embryos, and is probably of maternal origin. This transcript is expressed again in second instar larvae, then peaks in adult females. All three of the more abundant transcripts are expressed at a higher level in females relative to males, although this may be due to loading effects (Figure 4). The 4.3 kb and 4.5 kb transcripts are expressed at all stages after the first instar larval stage. The dosage compensation status of *CG2556* has been examined in third instar larvae (section 3.7), as its transcript levels proved to be both too variable and low in hemisected adult flies for RNase protection analysis (data not shown).

The *CG15730* probe recognized a transcript at extremely low levels, requiring 15 – 20 μg of poly(A)$^+$ RNA for detection using northern analysis. Due to this, it was not possible to examine the stage-specific expression of this gene using northern RNA hybridization analysis because of the large amounts of sample tissue required. However, the expression of *CG15730* has been analysed in male (not shown) and female adults using northern RNA hybridization analysis (Figure 5). *CG15730* is present at very low levels in both male and female adult flies, as three principle transcripts of 4.3 kb, 4.5 kb and 8 kb. This is considerable larger than the *AT20771* cDNA of 2.233 kb, indicating that either this cDNA does not represent the complete *CG15730* gene, that these large products represent contaminating genomic DNA present in the poly(A)$^+$ RNA preparation. It is also possible that these bands represent weak cross-hybridization to abundant transcripts.

The stage-specific expression of *Lsp1α* was also confirmed by northern RNA hybridization analysis. As is consistent with previous results, a 2.85 kb transcript of

Lsp1α is expressed at high levels almost exclusively in third instar larvae (Roberts *et al.*, 1977; Smith *et al.*, 1981) (Figure 6). Although the third instar larval, pupal, and adult male RNA samples in this gel are underloaded with respect to *rp49*, it is apparent that *Lsp1α* transcripts are only present in the third instar larval stage. Furthermore, a decrease in the level of *rp49* expression is visible in published northern analysis during late larval and early pupal stages relative to the other stages of development (O'Connell and Rosbash, 1984). The dosage compensation status of this gene has been examined in third instar larvae (section 3.7).

3.2.1 The CG15730 transcript can not be detected in adult poly(A)+ mRNA using RT-PCR

In order to confirm whether the detection of *CG15730* with 15 – 20 μg of poly(A)$^+$ RNA using northern RNA hybridization analysis in male and female adults was due to the presence of contaminating genomic DNA, RT-PCR of *CG15730* was performed on mixed adult poly(A)$^+$ mRNA. Reverse transcription of 100 ng of DNase-treated mixed adult poly(A)$^+$ mRNA was performed in the presence of reverse transcriptase and RNA template, and in the absence of either of these (section 2.6.1.9). PCR was performed on the cDNA, the reactions performed in the absence of reverse transcriptase or RNA template, a positive genomic control, and a negative water control using primers specific for *CG15730* (Figure 7A) and *CG2556* (Figure 7B). PCR was performed on *CG15730* for 40 cycles using the AT20771F and AT20771R primers, and on *CG2556* for 35 cycles using the CG2556E2F and CG2556E4R primers (section 10.1). The primers for *CG2556* span an intron, hence produce a genomic product of 820 bp compared to a cDNA product of 636 bp. *CG15730* does not contain an intron, hence the genomic and cDNA products are both expected to be 1478 bp (Figure 7A). No PCR products were detected in any of the negative control lanes (lanes 2, 4 and 6), and PCR products of the correct size were detected in the genomic template lane (lane 5). However, while a PCR product of the correct size was detected for *CG2556* in the cDNA lane (lane 2B), no such product was detected in the corresponding PCR reaction of *CG15730* (lane 2A). *CG2556* is not abundantly expressed in adult *Drosophila*, hence the detection of this gene but not *CG15730*, indicates that *CG15730* is not expressed at detectable levels in entire adult *Drosophila.* There is a high level of sequence conservation between *Drosophila* species in the putative open reading frame of *CG15730* (appendix 10.7). 194 of the 784 residues (24%) in the putative CG15730

protein are conserved (identical) between 7 *Drosophila* species (excluding *D. simulans* due to the incomplete nature of the protein sequence). Due to the existence of a cDNA corresponding to this gene, and this high level of sequence conservation, it appears likely that *CG15730* is a real gene, but may be expressed in a very small set of cells, or only under certain conditions. Due to the difficulty in detecting expression of *CG15730*, and the inconclusive nature of the gene, RNase protection of this gene was not attempted.

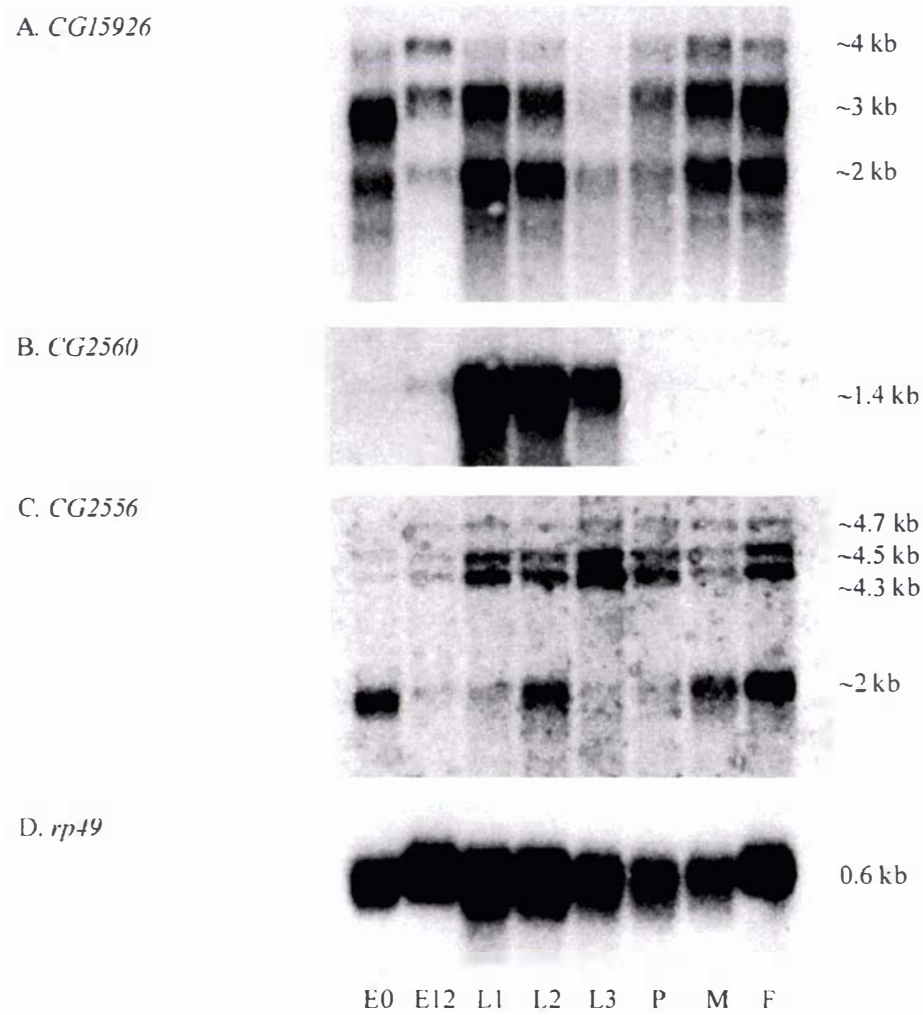


Figure 4. Stage-specific expression of *CG15926*, *CG2560* and *CG2556*

Northern analysis of 2 μg of poly(A)$^+$ RNA from embryos 0 – 2 h after laying (E0), embryos 12 h after laying (E12), first instar larvae (L1), second instar larvae (L2), third instar larvae (L3), pupae (P), adult males (M), and adult females (F). All embryo, larval and pupae samples consist of mixed male and female RNA. (A) Northern probed with *CG15926* cDNA and exposed for 18 h with BioMax film (Kodak) and intensifying screen at -80°C. (B) Northern probed with *CG2560* cDNA and exposed for 10 h with BioMax film (Kodak) and intensifying screen at -80°C. (C) Northern probed with *CG2556* cDNA and exposed for 18 h with BioMax film (Kodak) and intensifying screen at -80°C. (D) Northern probed with *rp49* cDNA as a loading control and exposed for 1 h with BioMax film (Kodak) and intensifying screen at -80°C.

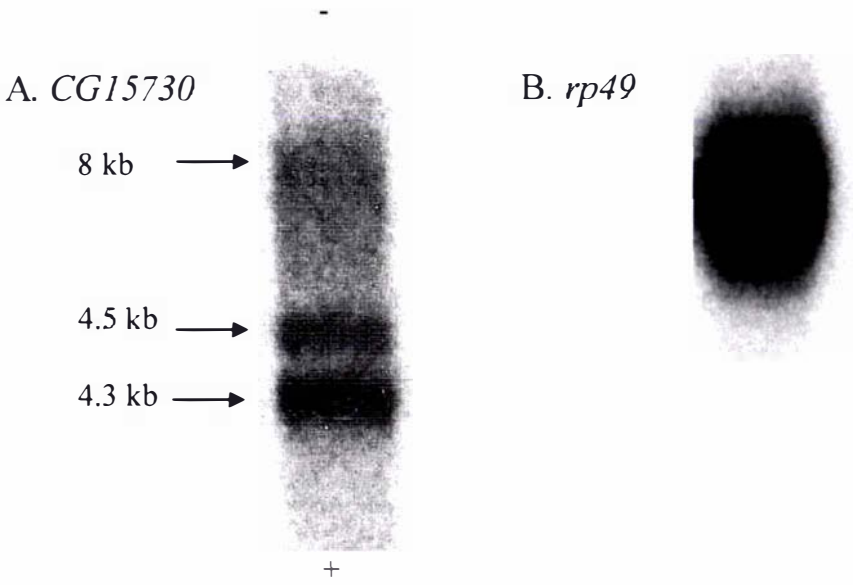


Figure 5. Northern analysis of *CG15730*

Northern analysis of 20 μg of poly(A)$^+$ RNA from adult female *Drosophila melanogaster*. (A) Northern probed with *CG15730* cDNA and exposed to BioMax film for 18 h (Kodak) and intensifying screen at -80°C. (B) Northern probed with *rp49* cDNA as a loading control and exposed to BioMax film for 1 h (Kodak) and intensifying screen at -80°C. Putative transcripts of 4.3 kb, 4.5 kb and 8 kb were detected that were considerably larger than the 2.2 kb transcript expected.

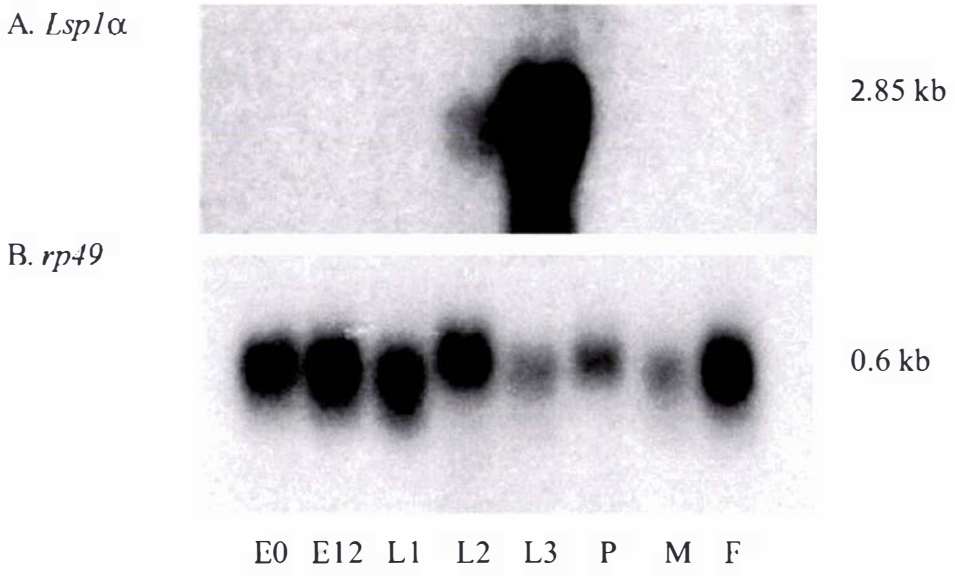


Figure 6. Stage-specific expression of *Lsp1α*

Northern analysis of 10 μg of total RNA from embryos 0 – 2 h after laying (E0), embryos 12 h after laying (E12), first instar larvae (L1), second instar larvae (L2), third instar larvae (L3), pupae (P), adult males (M), and adult females (F). All embryo, larval and pupae samples consist of mixed male and female RNA. (A) Northern probed with *Lsp1α* cDNA and exposed for 18 h with BioMax film (Kodak) and intensifying screen at -80°C. (B) Northern probed with *rp49* cDNA and exposed for 18 h with BioMax film (Kodak) at -80°C.

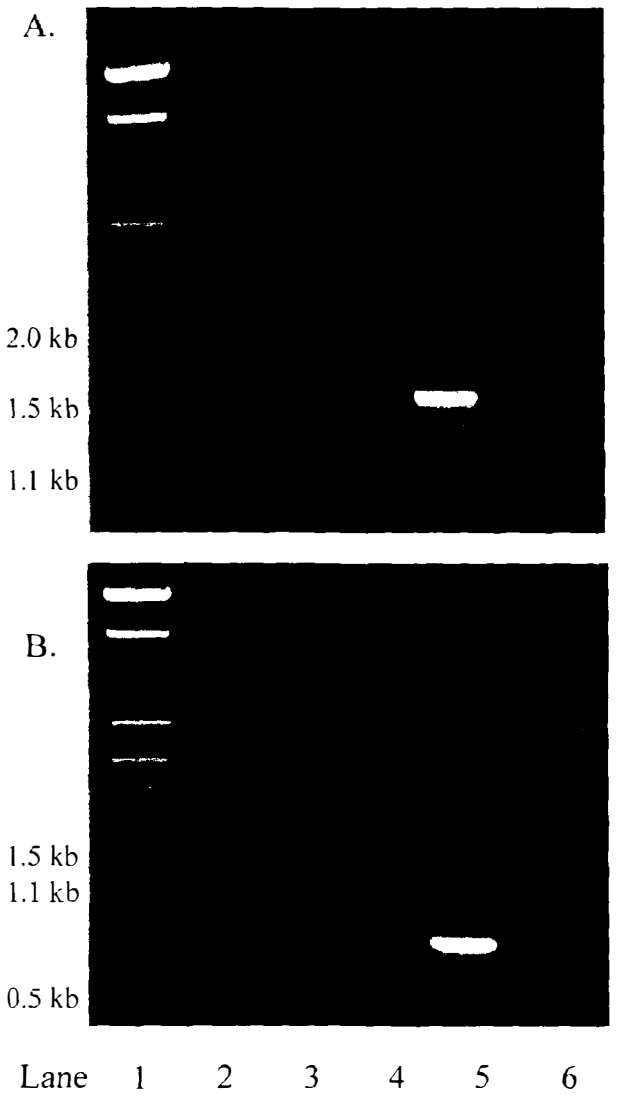


Figure 7. *CG15730* is not detected in mixed adult poly(A)$^+$ mRNA relative to *CG2556* using RT-PCR

RT-PCR of *CG15730* (A) and *CG2556* (B) performed as described in 3.2.1. 10% of each PCR reaction was electrophoresed on 1% (w/v) agarose gel. Lane 1 consists of molecular size markers (section 2.6.2.1). The templates for the PCR consist of: 5 µl of cDNA generated from 100 ng of mixed male and female *y w Drosophila melanogaster* adult poly(A)$^+$ mRNA (lane 2); 1 µl of reverse transcription reaction containing no RNA template (lane 3); 1 µl of reverse transcription reaction containing no reverse transcriptase (lane 4); 100 ng of genomic DNA (lane 5); 1 µl of sterile water (lane 6 negative control).

3.3 Production of probes for *in vitro* transcription

Templates for the generation of antisense single stranded RNA probes for RNase protection analysis were generated by restriction endonuclease digestion of plasmid clones, followed by gel purification.

A template for *in vitro* transcription of *CG15926* was prepared by digesting the *CG15926* clone (section 3.1.1) with *Dra*I. This generates an *in vitro* transcription template that produces an antisense ssRNA probe of 400 nt using T3 polymerase, consisting of 325 nt of the 3' end of the *CG15926* clone and 75 nt of the pFLC-1 MCS. This probe protects two discrete fragments of approximately 100 nt and 225 nt when the hybridized probe and mRNA are digested with RNaseA/T1 (Figure 8). This is probably due to nicking of an AT rich region that lies 100 nt from the 3' end of the *CG15926* mRNA by RNaseA, as digestion with RNaseT1 alone results in the production of a single protected fragment of the expected 325 nt size. This fragment is more diffuse than the two separate fragments, therefore A/T1 digestion was used rather than T1 alone in all quantitative analyses.

The *CG2560* subclone (section 3.1.2.4) was digested with *Eco0109*I, which cuts within the MCS of pBluescript at the 5' end of the *CG2560* fragment. This template produces an antisense 449 nt ssRNA probe using T7 polymerase that consists of the entire 366 nt of the *CG2560* fragment together with the pBluescript MCS regions either side (Figure 8). This probe protects a single discrete 366 nt fragment of the *CG2560* mRNA.

The *Lsp1α* 3' UTR clone in pGEM®-T Easy (section 3.1.2.3) was digested with *Eco*RV, which cuts 90 bp from the 5' end of the 3' UTR clone. This template produces an antisense 565 nt ssRNA using SP6 polymerase that protects two fragments, a major product of 294 nt and a minor product of 391 nt, corresponding to two polyadenylation sites present in the 3' UTR fragment (Figure 8, section 10.6.1). The minor product at 391 nt is represented by a very faint band, and is unlikely to be of biological significance.

The restriction enzyme *Ase*I was used to digest the *CG2556* clone (section 3.1.1). A 343 nt antisense ssRNA probe is transcribed from this template with T3 polymerase

consisting of 268 nt of the 3' end of the *CG2556* clone and 75 nt of the pFLC-1 MCS. This probe protects a discrete 268 nt fragment of the *CG2556* mRNA (Figure 8).

The *Pgd* clone (section 3.1.1) was digested with *Nru*I, which cuts inside the first exon. The antisense ssRNA probe transcribed from this template with T7 polymerase covers the first and second exons of *Pgd*, and includes the intervening intron. Hence, this 431 nt probe protects two discrete fragments of 43 nt and 171 nt corresponding to exons 1 and 2 respectively (Figure 8). Only the 171 nt fragment is visible in Figure 8.

The *rp49* clone (section 3.1.1) was digested with *Afl*III, which cuts inside the first intron. The antisense ssRNA probe generated from this template using T3 polymerase is 375 nt and consists of 312 nt of the second exon, 8 nt of the intron, and 55 nt of the pBluescript MCS. This probe protects a discrete 312 nt fragment of the *rp49* mRNA (Figure 8).

Template	Polymerase	Probe length	Protected fragment(s)
CG15926	T3	400 nt	325 nt
CG2560	T7	449 nt	366 nt
Lsp1α	SP6	565 nt	391 nt + 294 nt
CG2556	T3	343 nt	268 nt
Pgd	T7	431 nt	43 nt + 171 nt
rp49	T3	375 nt	312 nt

Table 9. Summary of the templates used for the *in vitro* transcription of antisense ssRNA probes for RNase protection analysis

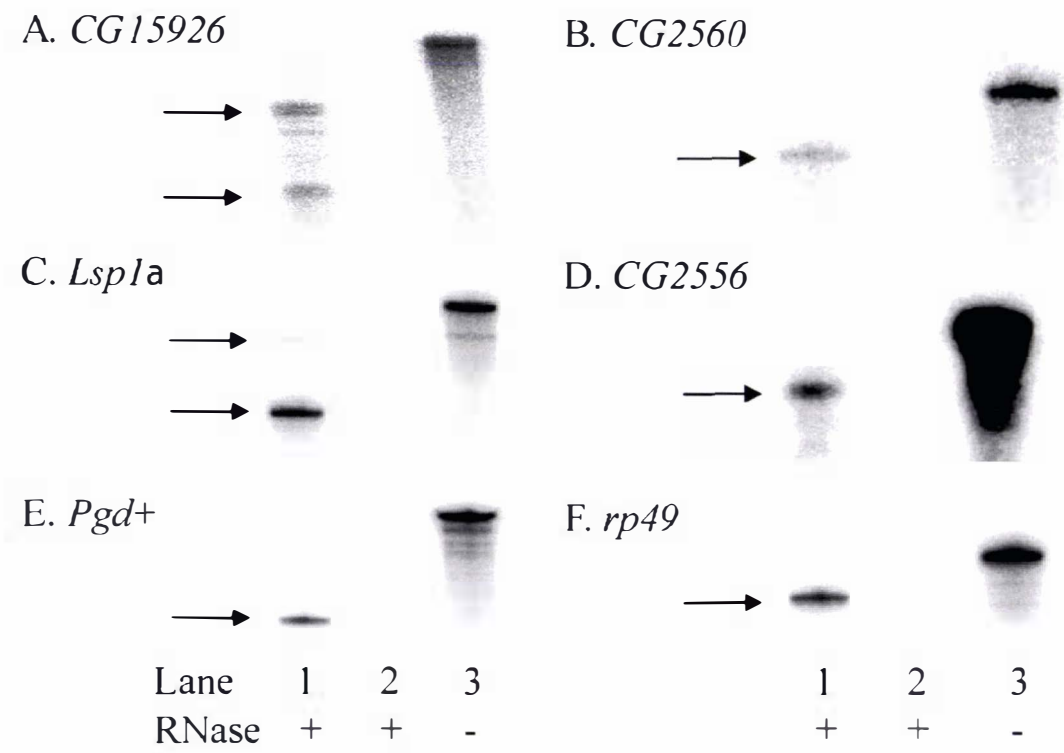


Figure 8. RNase protection analysis of *CG15926, CG2560, Lsp1α, CG2556, Pgd* and *rp49*

Lane 1 consists of 20 μg (A, D) or 4 μg (B, C, E, F) of RNA from the stage at which the gene is maximally expressed hybridized to the appropriate antisense ssRNA probe and digested with RNase. Lane 2 consists of 20 μg (A, D) or 10 μg (B. C, E, F) of yeast RNA hybridized to the appropriate antisense ssRNA probe and digested with RNase. Lane 3 consists of 10 μg of yeast RNA hybridized to the appropriate antisense ssRNA probe, not digested with RNase, representing intact probe. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1 – 1.5 h and either visualized on a FLA-5000 phosphorimager (Fujifilm) or by exposure to BioMax film (Kodak). Protected transcripts are indicated by the arrows and are discussed in the text and described in Table 9.

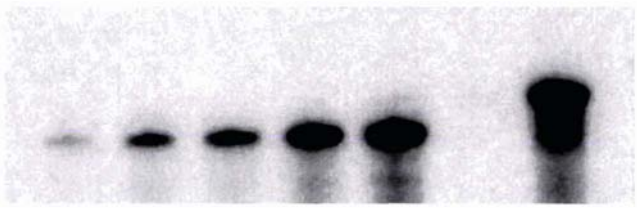
3.4 Determination of the linear range of the RNase protection assays

Prior to the determination of the dosage compensation status of genes within the *Lsp1α* domain using RNase protection, it was necessary to establish a range of RNA concentrations at which the assay remained linear for each gene. RNA harvested from the developmental stage at which each gene is maximally expressed as determined by northern analysis (section 3.2) was extracted. A range of amounts of this RNA was assayed using RNase protection with specific antisense ssRNA probes for each of the genes within the *Lsp1α* gene domain, as well as the two control genes, *rp49* and *Pgd*. Initial studies demonstrated that assays using RNA extracted with Trizol that had not undergone DNase treatment coupled with a secondary phenol/chloroform extraction were not linear at higher RNA concentrations (data not shown). Hence, all assays were performed using RNA that had undergone a secondary phenol/chloroform extraction following DNase treatment.

As demonstrated in Figure 9, Figure 10, Figure 11 and Figure 12, RNase protection analysis is linear over the range from 2 – 16 μg of total RNA for *rp49*, *Pgd*, *CG2560* and *Lsp1α*. The specific activity of the *rp49* and *Lsp1α* probes were 20- and 10-fold lower than that of *CG2560*, while the *Pgd*, *CG2556* and *CG15926* probes had the highest significant activity due to the relative low abundance of these transcripts (Table 4). One μg of total RNA produces levels nearly equivalent with that of the background, and hence is too low for use in dosage compensation determination. Four μg of total RNA was used for the dosage compensation determination of each of these genes. The poorer fit of the *CG2560* data points to the line of best fit at the lower end of the range is probably caused by incorrect loading, and this gene was still considered to be within the linear range at 4 μg of total RNA.

RNase protection analysis of *CG2556* is linear over the range from 7.5 – 30 μg of total RNA (Figure 13). The outlying value at 15 μg is due to incomplete loading of that sample in the gel. Twenty μg of total RNA was used for the dosage compensation determination of this gene in first instar larvae. RNase protection analysis of *CG15926* is linear over the range from 25 – 100 μg of total RNA (Figure 14). Forty μg of total RNA was used for the dosage compensation determination of this gene in hemisected adults.

A.



Lane	1	2	3	4	5	6	7
RNA (µg)	1	2	4	8	16	Y	Y
RNase	+	+	+	+	+	+	-

B.

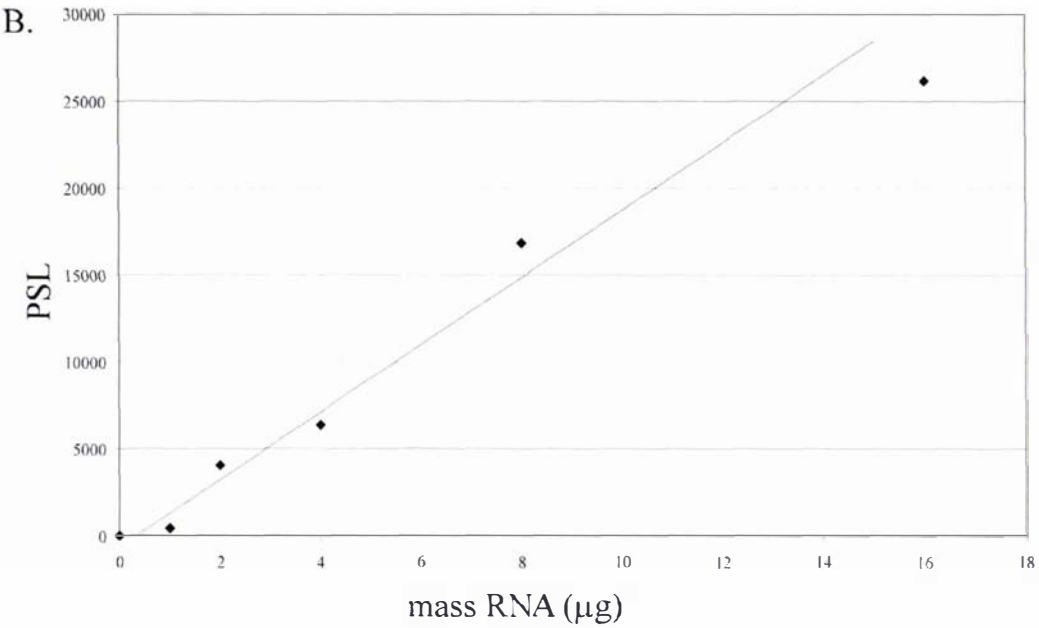


Figure 9. Linear range of *rp49* RNase protection analysis

RNA extracted from bromophenol-blue staged male third instar larvae subjected to RNase protection with an antisense *rp49* ssRNA of 375 nt (A). The samples run in lanes 1 – 5 consist of 1, 2, 4, 8 and 16 µg of total RNA respectively (312 nt). Lane 6 corresponds to a negative control of 16 µg of yeast RNA (Y) hybridized to probe and degraded by RNase. Lane 7 consists of intact *rp49* probe (375 nt) not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1 h, and quantified using FLA-5000 phosphorimager (Fujifilm). The quantified results are shown graphically (B) as PSL (photo stimuli level) versus mass of RNA in the assay, with lane 6 (negative control) as the background value of zero. PSL values do not have units.

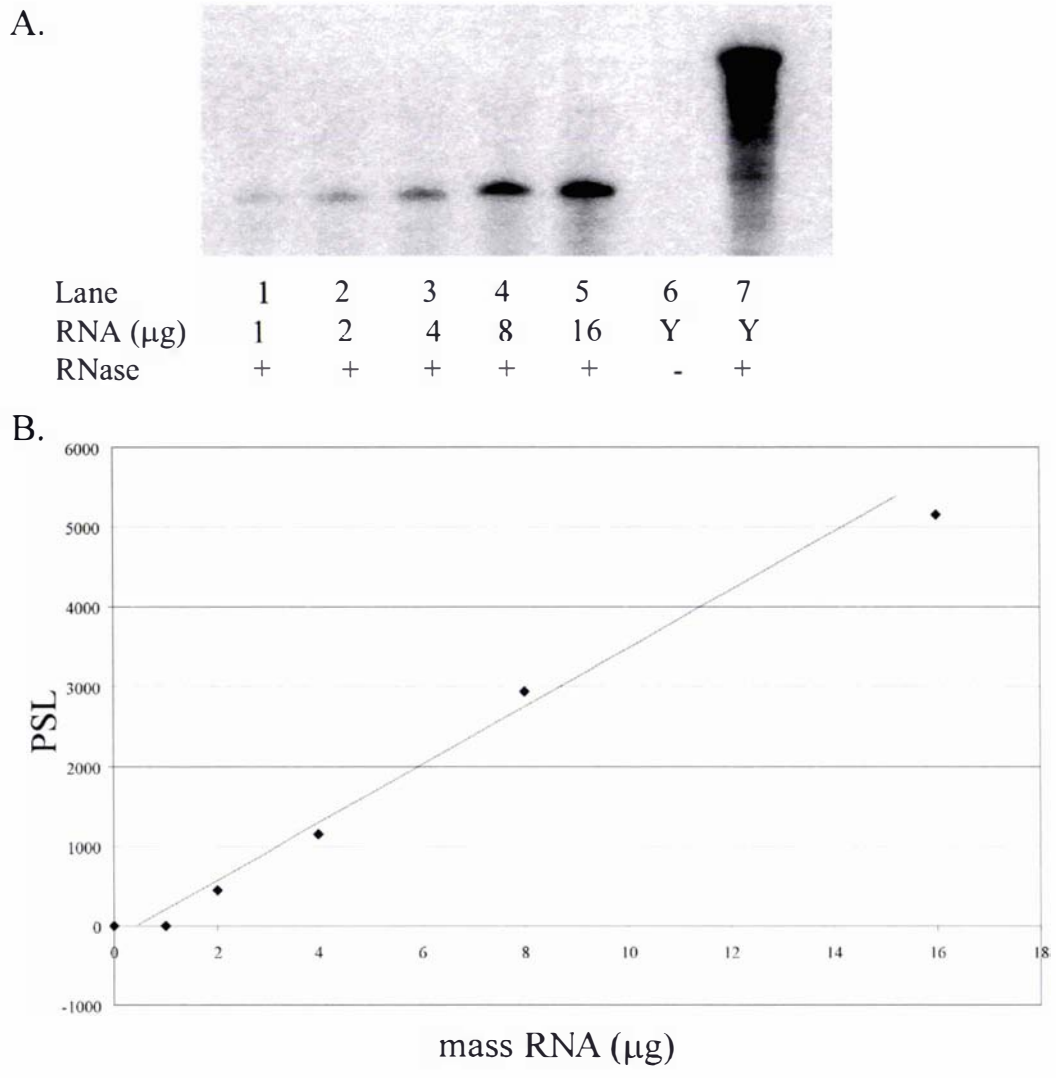
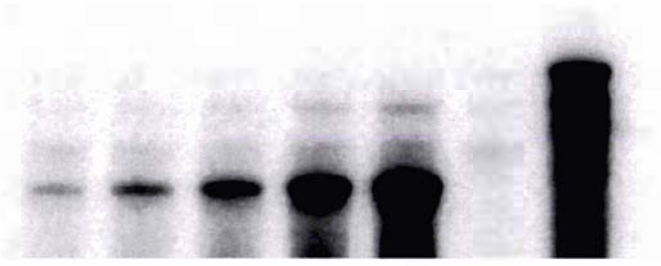


Figure 10. Linear range of *Pgd* RNase protection analysis

RNA extracted from bromophenol blue-staged male third instar larvae subjected to RNase protection with an antisense *Pgd* ssRNA of 431 nt (A). The samples run in lanes 1 – 5 consist of 1, 2, 4, 8 and 16 μg of total RNA respectively (43 nt + 171 nt). Lane 6 corresponds to a negative control of 20 μg of yeast RNA (Y) hybridized to probe and degraded by RNase. Lane 7 consists of intact *Pgd* probe (431 nt) not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1.5 h, and quantified using a FLA-5000 phosphorimager (Fujifilm). The quantified results are shown graphically (B) as PSL versus mass of RNA in the assay, with lane 6 (negative control) as the background value of zero.

A.



Lane	1	2	3	4	5	6	7
RNA (μg)	1	2	4	8	16	Y	Y
RNase	+	+	+	+	+	+	-

B.

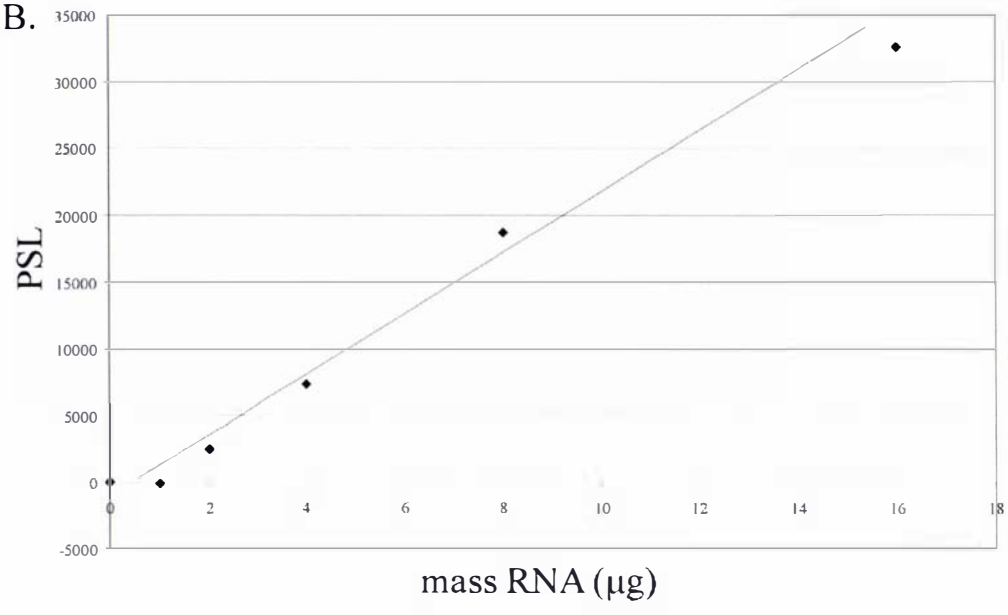
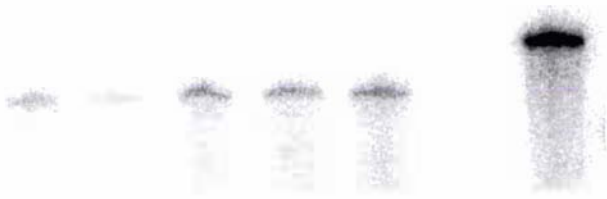


Figure 11. Linear range of *Lsp1α* RNase protection analysis

RNA extracted from bromophenol blue-staged male third instar larvae subjected to RNase protection with an antisense *Lsp1α* ssRNA of 565 nt (A). The samples run in lanes 1 – 5 consist of 1, 2, 4, 8 and 16 μg of total RNA respectively (294 nt + 391 nt). Lane 6 corresponds to a negative control of 20 μg of yeast RNA (Y) hybridized to probe and degraded by RNase. Lane 7 consists of intact *Lsp1α* probe (565 nt) not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1 h, and quantified using a FLA-5000 phosphorimager (Fujifilm). The quantified results are shown graphically (B) as PSL versus mass of RNA in the assay, with lane 6 (negative control) as the background value of zero.

A.



Lane	1	2	3	4	5	6	7
RNA (μg)	4	8	12	16	24	Y	Y
RNase	+	+	+	+	+	+	-

B.

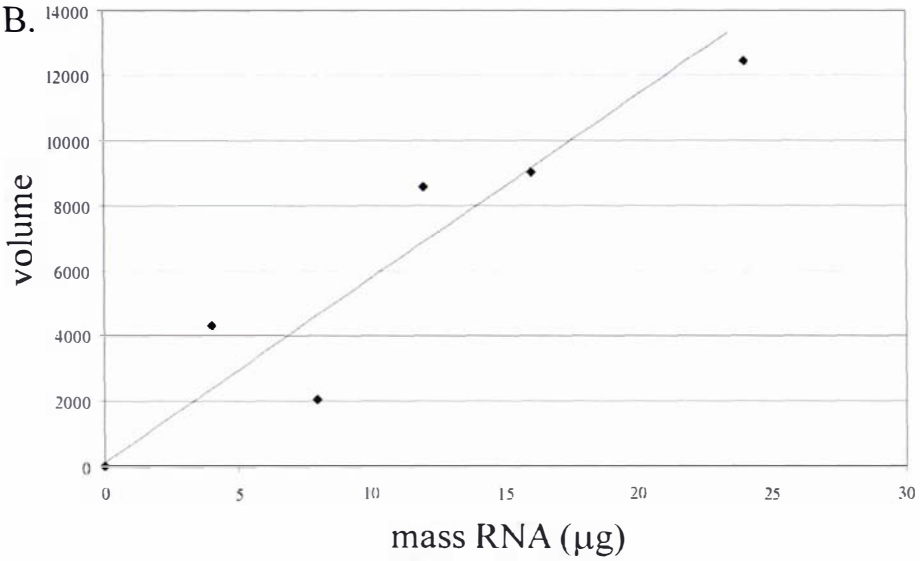


Figure 12. Linear range of *CG2560* RNase protection analysis

RNA extracted from mixed first instar larvae subjected to RNase protection with antisense ssRNA probe corresponding to *CG2560* (A). The samples run in lanes 1 – 5 consist of 4, 8, 12, 16 and 24 μ g of total RNA (366 nt). Lane 6 corresponds to a negative control of 24 μg of yeast RNA (Y) hybridized to probe and degraded by RNase. Lane 7 consists of intact probe (449 nt) not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1 h, and quantified using a Storm 860 phosphorimager (Molecular Dynamics). The quantified results are shown graphically (B) as volume versus amount of RNA in the assay, with lane 6 (negative control) as the background value of zero.

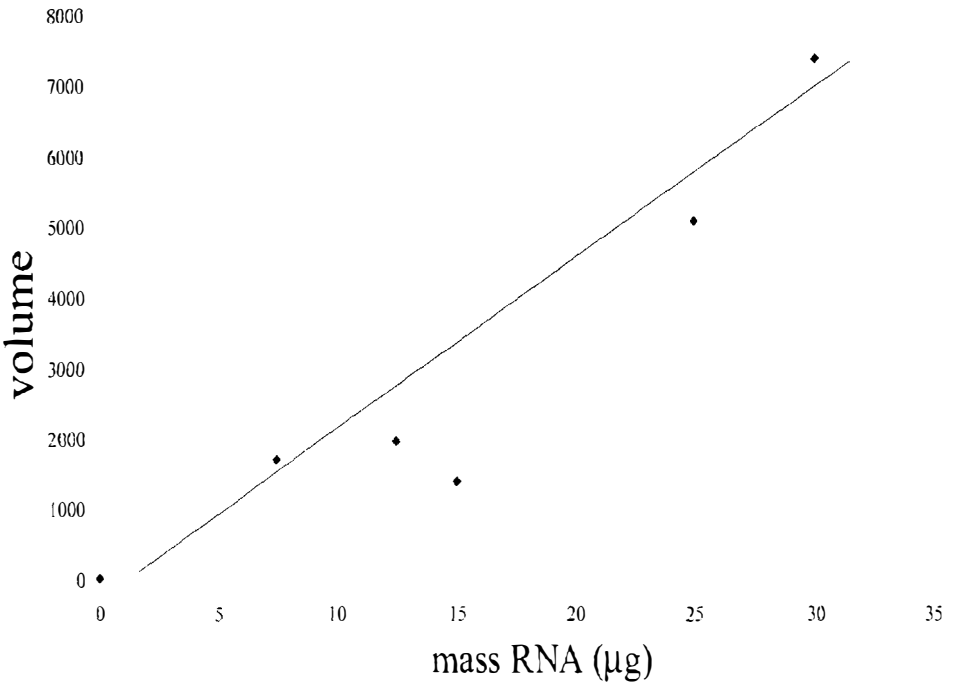


Figure 13. Linear range of *CG2556* RNase protection analysis

RNA extracted from mixed adults subjected to RNase protection with antisense ssRNA probes corresponding to *CG2556*. The samples consist 7.5, 12.5, 15, 25 and 30 µg of total RNA. A negative control of 30 µg of yeast RNA (Y) hybridized to probe and degraded by RNase was included. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1.5 h, and quantified using a Storm 860 phosphorimager (Molecular Dynamics). The quantified results are shown graphically as cpm versus amount of RNA in the assay, with the negative control as the background value of zero.

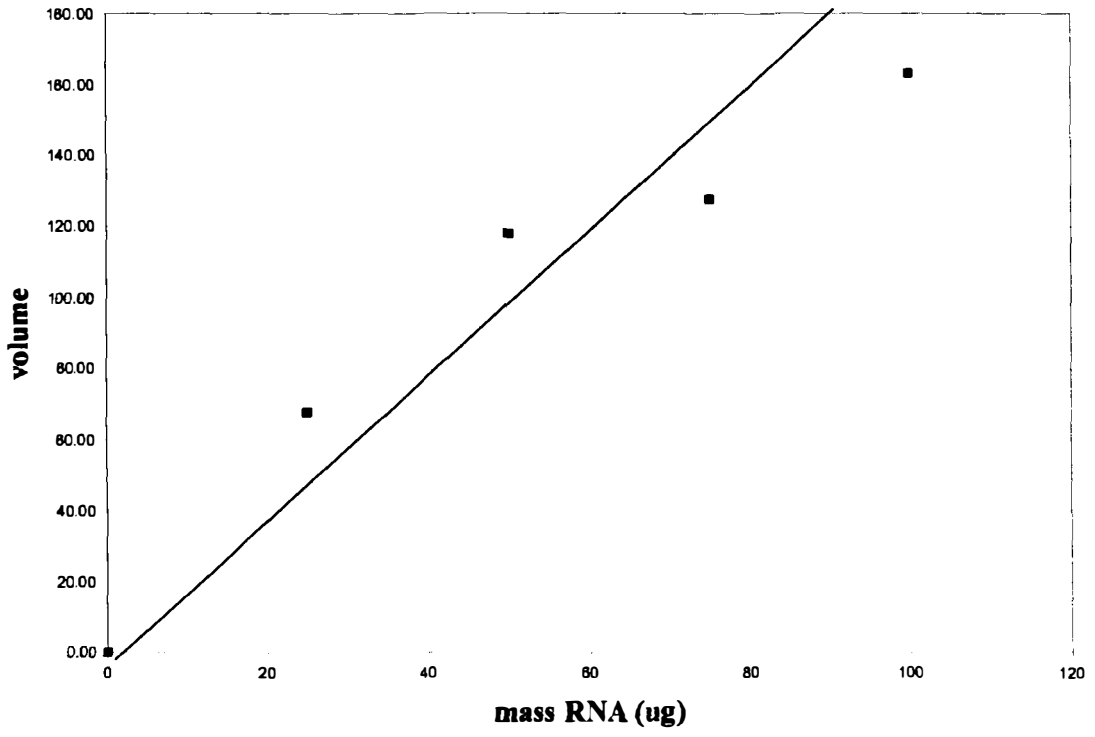


Figure 14. Linear range of *CG15926* RNase protection analysis

RNA extracted from mixed adults subjected to RNase protection with antisense ssRNA probes corresponding to *CG15926*. The samples consist 25, 50, 75 and 100 μg of total RNA. A negative control of 100 μg of yeast RNA (Y) hybridized to probe and degraded by RNase was included. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1.5 h, and quantified using a Storm 860 phosphorimager (Molecular Dynamics). The quantified results are shown graphically as cpm versus amount of RNA in the assay, with the negative control as the background value of zero.

3.5 *CG15926* is upregulated in male adults

The dosage compensation status of *CG15926* was examined in hemisected *y w* adults as northern RNA hybridization analysis indicated that this gene was expressed in both male and female adult flies. Hemisected adults were used to minimize the presence of transcripts from the reproductive organs. Four separate simultaneous harvests of female and male hemisected *y w* adult flies were conducted, and the RNA from these flies extracted, DNase treated, and subjected to a secondary phenol/chloroform extraction. These RNA samples were assayed using antisense ssRNA probes for *CG15926*, *rp49* and *Pgd*. Forty μg of RNA was assayed for *CG15926*, and 8 μg of RNA for *Pgd* and *rp49*. The protected fragments were separated on 5% (v/v) polyacrylamide/8M urea gels, which were dried and quantified using a FLA-5000 phosphorimager (Figure 15, Fujifilm). The value (photo stimuli level; PSL) obtained for each band was corrected for the background value of the lane in which yeast RNA had been hybridized to the probes and completely degraded by RNase treatment (Table 10). In order to control for uneven levels of RNA in the samples, the values obtained for *CG15926* and *Pgd* were further corrected by dividing by the values obtained in each lane for *rp49*. Following this, the female to male ratios of both genes were calculated for the 4 matched samples (Table 11). A female to male ratio of one indicates that a gene is dosage compensated, whereas a female to male ratio of two suggests that a gene is not dosage compensated. A female to male ratio of less than one indicates that a gene is upregulated in males. *Pgd* has a female to male transcript ratio of 0.90 ± 0.13, indicating that it is dosage compensated as expected. *CG15926*, however, has a female to male transcript ratio of 0.60 ± 0.12 indicating that it is upregulated in males. This increased male expression indicates a possible additional role for CG15926 in males, although the function of this protein is unknown, and it does not share significant sequence identity with any proteins of known function (Table 2). However, the dosage compensation status of *CG15926* could not be examined using RNase protection analysis due to this increased expression in male adults.

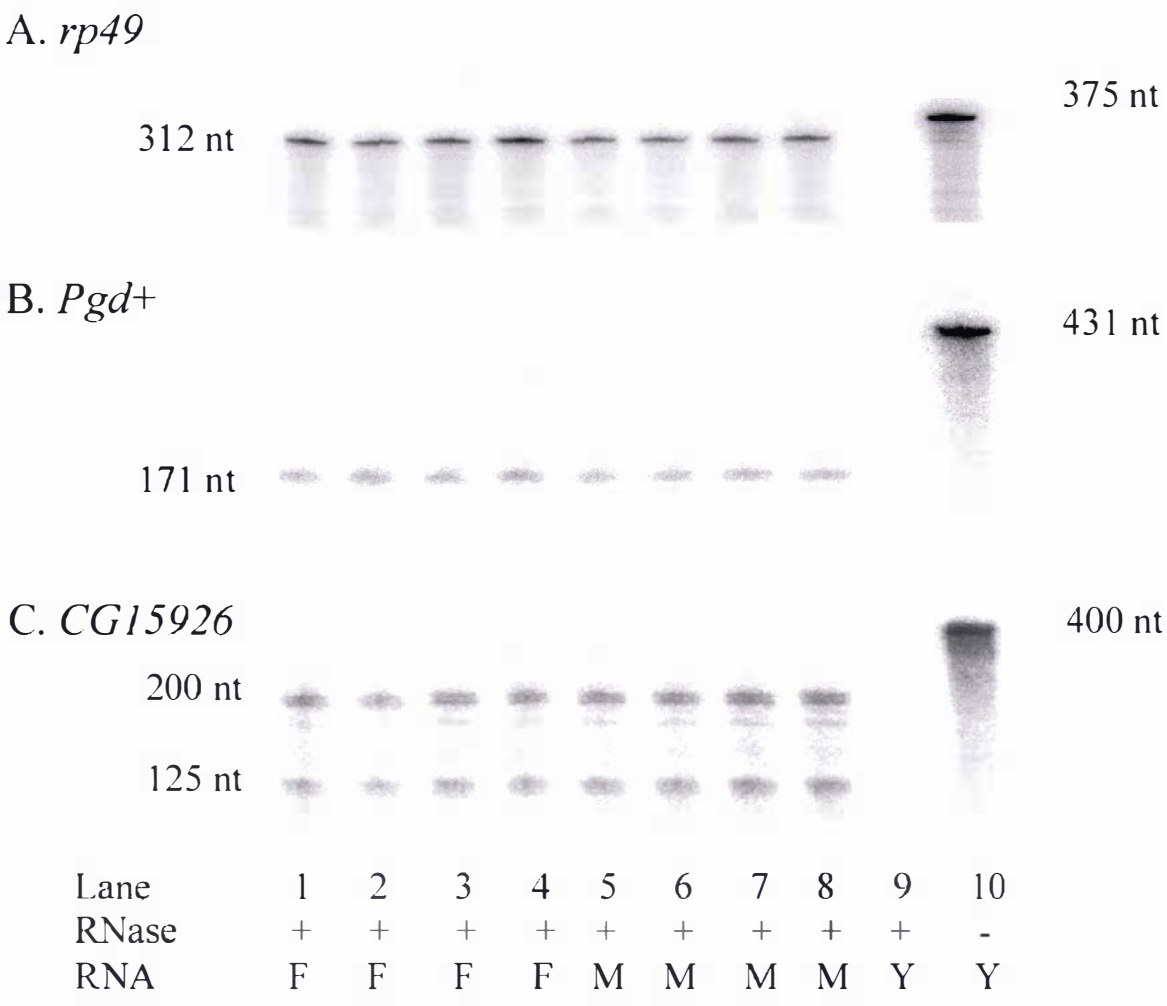


Figure 15. RNase protection analysis of *rp49*, *Pgd* and *CG15926* transcripts in male and female hemisected adults

RNA extracted from female (lanes 1 – 4) and male (lanes 5 – 8) hemisected *y w* adults was subjected to RNase protection with antisense ssRNA probes corresponding to *rp49*, *Pgd* and *CG15926*. The samples run in lanes 1 – 8 consist of 40 μg (*CG15926*) or 8 μg (*rp49*, *Pgd*) of total RNA, lane 9 corresponds to a negative control of 10 (A, B) or 40 (C) μg of yeast RNA hybridized to the probe and degraded by RNase. Lane 10 consists of intact probe not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1 h and quantified using a FLA-5000 phosphorimager (Fujifilm).

Gene	Sex	Volume[1]	Volume - background	(Volume-background)/*rp49*
CG15926	female	6180.09	5991.32	1.02
CG15926	female	4396.81	4208.04	0.88
CG15926	female	6313.45	6124.68	1.07
CG15926	female	5754.76	5565.99	0.67
CG15926	male	6864.9	6676.13	1.50
CG15926	male	7006	6817.23	1.63
CG15926	male	8520.36	8331.59	1.47
CG15926	male	8809.9	8621.13	1.49
CG15926	Y+R	188.77	0.00	
Pgd	female	1092.81	1018.43	0.17
Pgd	female	1389.77	1315.39	0.27
Pgd	female	1177.29	1102.91	0.19
Pgd	female	1565.84	1491.46	0.18
Pgd	male	1017.13	942.75	0.21
Pgd	male	1125.71	1051.33	0.25
Pgd	male	1391.16	1316.78	0.23
Pgd	male	1291.54	1217.16	0.21
Pgd	Y+R	74.38	0	
rp49	female	5973.17	5874.30	
rp49	female	4888.03	4789.16	
rp49	female	5836.79	5737.92	
rp49	female	8425.14	8326.27	
rp49	male	4563.87	4465.00	
rp49	male	4292.71	4193.84	
rp49	male	5753.77	5654.90	
rp49	male	5889.47	5790.60	
rp49	Y+R	98.87	0.00	

Table 10. Levels of *CG15926, Pgd* and *rp49* transcripts in male and female hemisected *y w* adult flies determined by RNase protection analysis

All female and male transcript levels (PSL) corrected for the background level of the yeast + probe + RNase sample (negative control). *CG15926* and *Pgd* levels are normalized to *rp49* to correct for loading differences.

[1] PSL values do not have units.

	CG15926		*Pgd*	
	F/M Ratios	Mean	F/M Ratios	Mean
Lane 1/Lane 5	0.68	0.60 ± 0.12	0.82	0.90 ± 0.13
Lane 2/Lane 6	0.54		1.10	
Lane 3/Lane 7	0.72		0.83	
Lane 4/Lane 8	0.45		0.85	

Table 11. Female: male ratios of *CG15926* and *Pgd* transcript levels

Female to male ratios of *CG15926* and *Pgd* transcript levels normalized to *rp49* taken from Table 10 and Figure 15. Mean female to male ratios and 95% confidence intervals are indicated for each set of ratios.

3.6 Determination of the dosage compensation status of *CG2560*

Northern RNA hybridization analysis showed that the *CG2560* transcript was most abundant in first instar larvae. There is no conventional method of distinguishing gender in first instar larvae, hence a strain in which only the male larvae express GFP was used to determine the dosage compensation status of *CG2560* (section 2.5.1.2).

Four separate simultaneous harvests of female and male first instar larvae were conducted, and the RNA from these larvae extracted, DNase treated, and subjected to a secondary phenol/chloroform extraction. These RNA samples were assayed in duplicate using multiple antisense ssRNA probes for *CG2560*, *rp49* and *Pgd*. The protected fragments produced by these 3 probes were separated on a 5% (v/v) polyacrylamide/8M urea gel, which was dried and quantified using a Storm 860 phosphorimager (Molecular Dynamics) (Figure 16). The two similarly sized products of equal intensity representing *CG2560* were both quantified. The value (volume) obtained for each band was corrected for the background value of the lane in which yeast RNA had been hybridized to the probes and completely degraded by RNase treatment. In order to control for uneven loading, the values obtained for *CG2560* and *Pgd* were further corrected by dividing by the values obtained in each lane for *rp49*. Following this, the female to male ratios of both genes were calculated for the 4 matched samples (Table 12). A female to male ratio of one indicates that a gene is dosage compensated, whereas a female to male ratio of two suggests that a gene is not dosage compensated. These ratios strongly indicate that *CG2560* is dosage compensated, with the ratios obtained in gel A ranging from 0.94 to 1.14, while *Pgd*, which is known to be dosage compensated, ranges from 0.54 to 1.01. The ratio of 0.54 obtained for *Pgd* is quite low, and may be due to discrepancies in the gel or in the storage of the phosphor plate. The ratios in gel B have a wider range (0.93 to 1.21 for *CG2560*; 0.72 to 1.62 for *Pgd*), and are also consistent with dosage compensation of *CG2560*. The ratios obtained from lanes 3 and 7 can be discounted due to overloading of lane 3 in gel B.

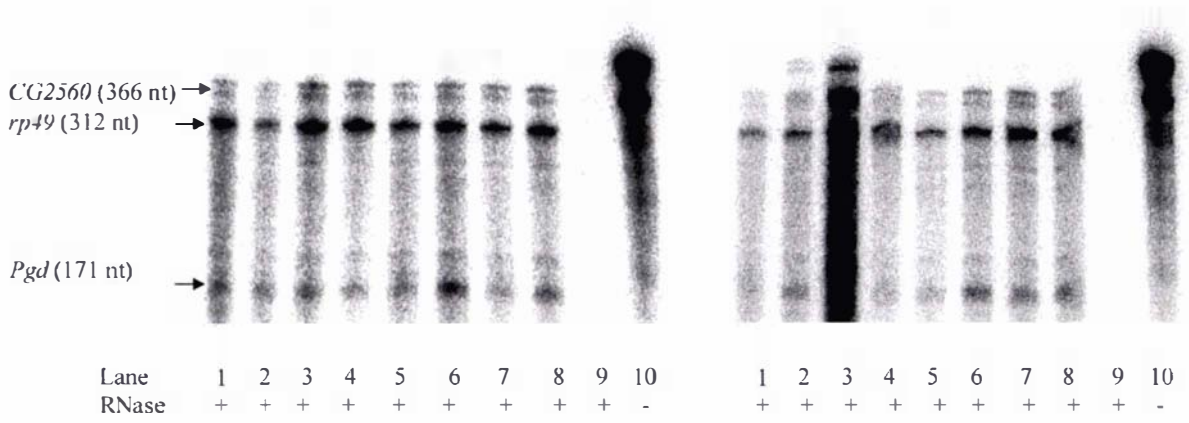


Figure 16. *CG2560* dosage compensation assay

Duplicate gels (A) and (B) consisting of RNA extracted from female (lanes 1 – 4) and male (lanes 5 – 8) first instar larvae subjected to RNase protection with multiple antisense ssRNA probes corresponding to *CG2560*, *rp49* and *Pgd*. The samples run in lanes 1 – 8 consist of 4 μg of total RNA, with lanes 1 and 4, 2 and 5, 3 and 6, and 7 and 8 consisting of matched samples. Lane 9 corresponds to a negative control of 10 μg of yeast RNA hybridized to the probes and degraded by RNase. Lane 10 consists of intact probes not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1.5 h and quantified using a Storm 860 phosphorimager (Molecular Dynamics).

Gene	Sex	Gel A			Gel B		
		Volume[1]	Volume - background	(Volume-background)/ *rp49*	Volume	Volume - background	(Volume-background/ *rp49*
CG2560	female	7528.27	6287.10	0.27	4079.22	2458.60	0.22
CG2560	female	4620.28	3379.11	0.31	9903.01	8282.39	0.42
CG2560	female	13522.34	12281.17	0.39	73325.08	71704.46	0.74
CG2560	female	9444.25	8203.08	0.30	9726.86	8106.24	0.31
CG2560	male	5683.53	4442.36	0.26	4978.76	3358.14	0.23
CG2560	male	9861.46	8620.29	0.33	11512.94	9892.32	0.35
CG2560	male	7211.98	5970.81	0.34	16099.80	14479.18	0.35
CG2560	male	8701.46	7460.29	0.31	13106.12	11485.50	0.31
CG2560	Y+R	1241.17	0.00		1620.62	0.00	
Pgd	female	13004.40	12441.74	0.53	8848.67	8235.41	0.73
Pgd	female	8344.15	7781.49	0.72	14843.51	14230.25	0.73
Pgd	female	11445.38	10882.72	0.34	80849.07	80235.81	0.83
Pgd	female	6981.21	6418.55	0.24	8443.87	7830.61	0.30
Pgd	male	10474.45	9911.79	0.58	7086.16	6472.90	0.45
Pgd	male	19025.07	18462.41	0.71	17141.07	16527.81	0.59
Pgd	male	7363.85	6801.19	0.39	15670.04	15056.78	0.36
Pgd	male	11243.88	10681.22	0.44	16045.73	15432.47	0.42
Pgd	Y+R	562.66	0.00		613.26	0.00	
rp49	female	24936.59	23538.02		12450.38	11257.89	
rp49	female	12181.04	10782.47		20775.89	19583.40	
rp49	female	33034.24	31635.67		98236.31	97043.82	
rp49	female	28472.96	27074.39		27153.74	25961.25	
rp49	male	18370.59	16972.02		15492.79	14300.30	
rp49	male	27308.94	25910.37		29442.87	28250.38	
rp49	male	19001.41	17602.84		42909.56	41717.07	
rp49	male	25709.28	24310.71		38271.05	37078.56	
rp49	Y+R	1398.57	0.00		1192.49	0.00	

Table 12. Levels of *CG2560*, *Pgd* and *rp49* transcripts in male and female first instar larvae determined by multi-probe RNase protection analysis

All female and male transcript levels (volume) corrected for the background level of the yeast + probe + RNase sample (Y+R; negative control). *CG2560* and *Pgd* levels are normalized to *rp49* to correct for loading differences. The underlined values from lane 3 in Gel B are discarded due to overloading.

[1] PSL values do not have units.

Female: Male ratio	Gel A		Gel B	
	CG2560	*Pgd*	*CG2560*	*Pgd*
Lane 1/Lane 5	1.02	0.91	0.93	1.62
Lane 2/Lane 6	0.94	1.01	1.21	1.24
Lane 3/Lane 7	1.14	0.89	2.13*	2.29*
Lane 4/Lane 8	0.99	0.54	1.01	0.72

Table 13. Female: male ratios of *CG2560* and *Pgd* transcript levels

Female to male ratios of *CG2560* and *Pgd* transcript levels normalized to *rp49*, and with outlying values removed. Indicated values (*) discarded due to overloading of lane 3, Figure 16.

	Gene	Mean	95% confidence interval
Gel A	*CG2560*	1.02	0.08
	Pgd	0.84	0.20
Gel B	*CG2560*	1.05	0.16
	Pgd	1.19	0.51

Table 14. Mean female: male ratios of *CG2560* and *Pgd*

Mean female to male ratios and 95% confidence intervals calculated using ratio values from Table 13 excluding the lane 3/lane 7 ratio from gel B (Figure 16).

3.7 Determination of the dosage compensation status of *Lsp1α* and *CG2556*

Third instar larvae of the *y w* genotype were used rather than the GFP strain used for first instar analysis (section 3.5). *Pgd* and *Lsp1α* transcripts were not equivalent in male and female third instar larvae from the GFP strain due to retarded male development possibly caused by the presence of a balancer chromosome. This data is not shown in detail, but female to male transcript ratios normalized to *rp49* ranged for *Pgd* from 0.62 to 1.08, and for *Lsp1α* from 0.42 to 1.73. The GFP fly stock was therefore not suitable for the analysis of *Pgd* and *Lsp1α* transcript levels in third instar larvae.

Two separate methods were used for collecting third instar larvae of the same developmental stage. The first method involved timing the development of the larvae. Newly hatched first instar larvae were collected after 36 h and grown on standard cornmeal/molasses media at 22°C for a further 36 h. Following this third instar larvae of the same age and approximately same size were collected and separated according to sex. The second approach used the blue-food method (Maroni and Stamey, 1983). This method utilizes the fact that larvae stop feeding late in the third instar, shortly prior to pupariation. Newly hatched first instar larvae were collected and grown on standard cornmeal/molasses media containing 0.05% (w/w) bromophenol blue at 22°C for *ca.* 36 h. Larvae fed on this medium show a dark blue alimentary canal, the colour of which persists for several hours after feeding has ceased. After this time (3 –4 h), the colour becomes significantly lighter, and larvae become completely white just before pupariation. Third instar larvae collected that have a light blue alimentary canal will be developmentally synchronous to within 2 – 4 h.

Four separate collections of 10 conventionally sexed *y w* third instar larvae were made using each method of staging. RNA was extracted from these larvae and DNase treated (section 2.6.1.2). Due to the similar sizes of some of the protected fragments produced by rp49 and *Lsp1α*, and the reduced expression levels of *CG2556* compared to these genes, multi-probe RNase protection was not attempted for these genes. RNase protection analyses were conducted on 4 μg of RNA from each of the 4 female and 4 male RNA samples using antisense ssRNAs corresponding to *rp49*, *Pgd* and *Lsp1α*. RNase protection was also conducted on 20 μg of RNA from each of the female and

male samples using an antisense ssRNA corresponding to *CG2556*. The protected fragments resulting from these assays were separated on 5% (v/v) polyacrylamide/8M urea gels, which were dried and quantified using a FLA-5000 phosphorimager (Fujifilm) (Figure 17). The value (photo stimuli level; PSL) obtained for each band was corrected for the background value of the lane in which yeast RNA had been hybridized to the probes and completely degraded by RNase treatment. As *rp49* was not analysed in the same gel as the other genes, no corrections were made for loading differences. Instead, several values were discarded due to misloading, for example the absent sample in lane 6 of Pgd gel (Figure 17), or due to suspected sample mix-ups. No more than 1 value was discarded from each of the 4 female or 4 male values for each gene, and each mean is based on at least 3 values. All the raw values used are shown in Table 15 and Table 16.

Following this, the mean female and male PSL values for each gene were calculated, together with ratios of these averages (Table 15, Table 17, Figure 17). The ratios were also calculated using an alternative method, in which female to male ratios of all 4 genes were calculated for the 3 to 4 matched samples (Table 16, Table 17, Table 18). This second method enabled 95% confidence intervals to be calculated for the mean female to male ratio. As the ratios calculated using both these methods are very similar, only ratios calculated using the latter method will be discussed below. A female to male ratio of one indicates that a gene is dosage compensated, whereas a female to male ratio of two suggests that a gene is not dosage compensated.

As expected, *rp49*, an autosomal gene, has a female to male transcript ratio of 0.79 ± 0.27 in the bromophenol blue-staged larvae, and 0.95 ± 0.22 in the age/size-staged larvae. This shows that the amount of RNA loaded in the female and male samples is approximately equivalent. *Pgd*, however has a female to male transcript ratio of 1.84 ± 0.51 in the age/size-staged larvae. This was not expected for an X-linked gene that is known to be dosage compensated, however may have occurred due to the developmental asynchrony of the third instar larvae collected. 6-PGD enzyme activity levels vary throughout third instar larval development (Williamson *et al.*, 1980), and third instar larvae that develop from first instar larvae collected within a 1 h period from hatching can reach pupariation with an asynchrony of 10 – 15 h (Maroni and Stamey, 1983). Using the alternative bromophenol blue-staged larvae, *Pgd* has the expected

female to male ratio of 0.91 ± 0.42. Thus, *Pgd* is dosage compensated in third instar larvae.

Lsp1α has a female to male transcript ratio of 1.81 ± 0.14 in the bromophenol blue-staged larvae, and of 2.32 ± 0.27 in the age/size-staged larvae. Hence, in concordance with the results of previous groups (section 1.7.1), *Lsp1α* is not dosage compensated in third instar larvae. The novel gene, *CG2556*, has a female to male transcript ratio of 0.99 ± 0.15 in the bromophenol blue-staged larvae, and of 0.98 ● 0.27 in the age/size-staged larvae. Thus *CG2556* is dosage compensated in third instar larvae.

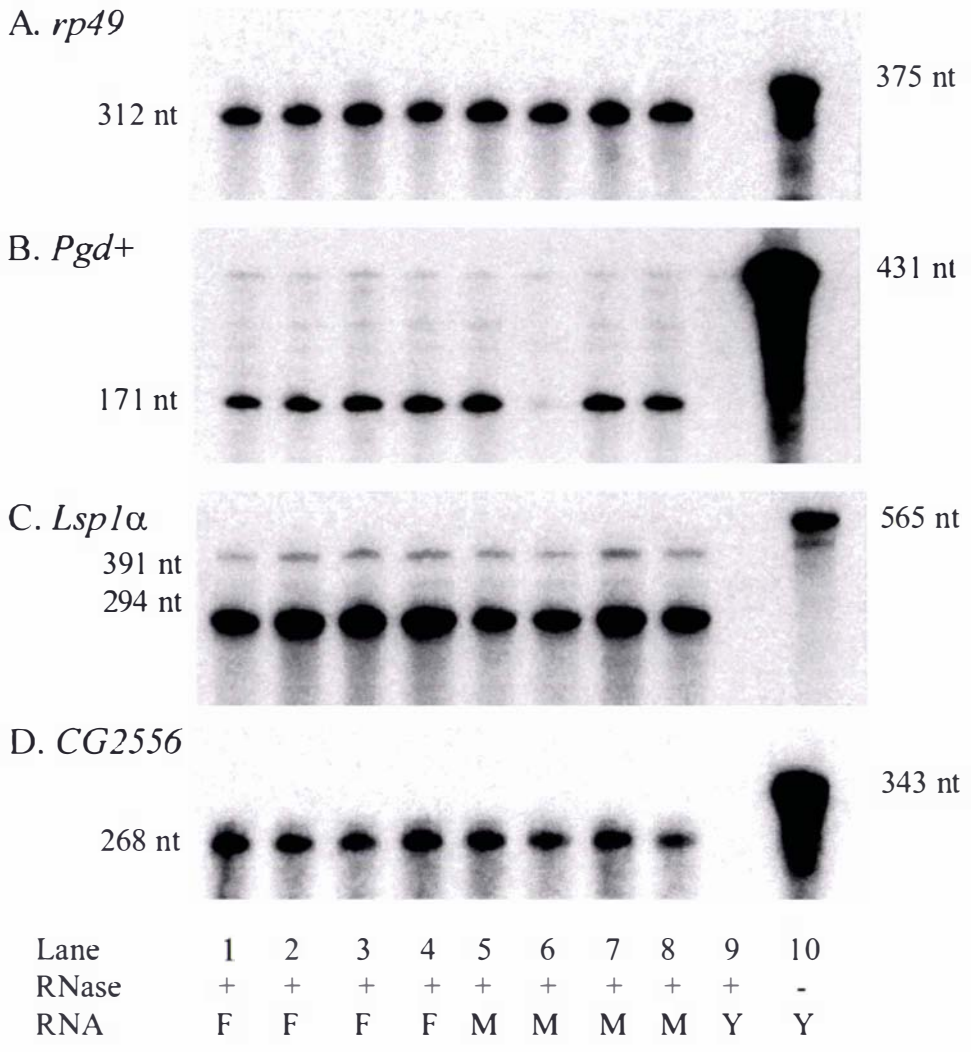


Figure 17. RNase protection analysis of *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcripts in male and female third instar larvae

RNA extracted from female (lanes 1 – 4) and male (lanes 5 – 8) third instar larvae was subjected to RNase protection with antisense ssRNA probes corresponding to *rp49*, *Pgd*, *Lsp1α* or *CG2556*. The samples run in lanes 1 – 8 consist of 4 μg of total RNA, lane 9 corresponds to a negative control of 10 μg of yeast RNA hybridized to the probe and degraded by RNase. Lane 10 consists of intact probe not subjected to RNase treatment. Samples were electrophoresed on a 5% (v/v) polyacrylamide/8 M urea gel at 140 V for 1 h (*rp49*, *Lsp1α* and *CG2556*) or 1.5 h (*Pgd*) and quantified using a FLA-5000 phosphorimager (Fujifilm). The sample gels shown for each probe were exposed to BioMax film (Kodak) and intensifying screen at -80°C for 3 – 18 h.

Gene	Sex	PSL[1]	PSL-background
rp49	female	7113.66	5748.16
rp49	female	8610.26	7244.76
rp49	female	6930.12	5564.62
rp49	female	5953.21	4587.71
rp49	male	9256.26	7890.76
rp49	male	7701.82	6336.32
rp49	male	8146.27	6780.77
rp49	male	10785.67	9420.17
rp49	yeast + RNase	1365.5	0
rp49	5 – 10% yeast - RNase	30290.69	28925.19
Pgd	female	2203	2079.87
Pgd	female	2745.73	2622.6
Pgd	female	2913.76	2790.63
Pgd	female	1884.24	1761.11
Pgd	male	2780.8	2657.67
Pgd	male	2233.21	2110.08
Pgd	male	2984.1	2860.97
Pgd	yeast + RNase	123.13	0
Pgd	5 – 10% yeast - RNase	82969.12	82845.99
Lsp1α	female	25476.35	24901.9
Lsp1α	female	24822.88	24248.43
Lsp1α	female	23958.51	23384.06
Lsp1α	male	14473.61	13899.16
Lsp1α	male	13059.91	12485.46
Lsp1α	male	14400.01	13825.56
Lsp1α	yeast + RNase	574.45	0
Lsp1α	5 – 10% yeast - RNase	12897.64	12323.19
CG2556	female	1137.72	826.21
CG2556	female	1116.71	805.2
CG2556	female	1429.97	1118.46
CG2556	female	1437.22	1125.71
CG2556	male	1308.08	996.57
CG2556	male	1065.35	753.84
CG2556	male	1373.18	1061.67
CG2556	yeast + RNase	311.51	0
CG2556	5 – 10% yeast - RNase	7330.38	7018.87

Table 15. Levels of *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcripts in male and female third instar larvae staged by the bromophenol blue-method

All female and male transcript levels (PSL) corrected for the background level of the yeast + probe + RNase sample (negative control).

[1] PSL values do not have units.

Gene	Sex	PSL[1]	PSL-background
rp49	female	5952.53	5750.35
rp49	female	5339.27	5137.09
rp49	female	5895.41	5693.23
rp49	female	6200.24	5998.06
rp49	male	5297.93	5095.75
rp49	male	7202.71	7000.53
rp49	male	7517.22	7315.04
rp49	male	5340.38	5138.20
rp49	yeast + RNase	202.18	0.00
rp49	5 – 10% yeast - RNase	24052.05	23849.87
Pgd	female	8391.46	8316.06
Pgd	female	7676.55	7601.15
Pgd	female	8094.43	8019.03
Pgd	female	8493.04	8417.64
Pgd	male	4702.52	4627.12
Pgd	male	4640.34	4564.94
Pgd	male	6052.17	5976.77
Pgd	male	3357.30	3281.90
Pgd	yeast + RNase	75.40	0.00
Pgd	5 – 10% yeast - RNase	29619.63	29544.23
Lsp1α	female	41410.77	41162.55
Lsp1α	female	41386.72	41138.50
Lsp1α	female	36235.98	35987.76
Lsp1α	male	20543.48	20295.26
Lsp1α	male	18009.31	17761.09
Lsp1α	male	13980.19	13731.97
Lsp1α	yeast + RNase	248.22	0.00
Lsp1α	5 – 10% yeast - RNase	18961.24	18713.02
CG2556	female	6678.91	6020.71
CG2556	female	5186.82	4528.62
CG2556	female	6211.13	5552.93
CG2556	female	6811.97	6153.77
CG2556	male	5056.83	4398.63
CG2556	male	6725.98	6067.78
CG2556	male	7364.61	6706.41
CG2556	male	6876.01	6217.81
CG2556	yeast + RNase	658.20	0.00
CG2556	5 – 10% yeast - RNase	107643.75	106985.55

Table 16. Levels of *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcripts in male and female third instar larvae staged by age/size

All female and male transcript levels (PSL) corrected for the background level of the yeast + probe + RNase sample (negative control).

[1] PSL values do not have units.

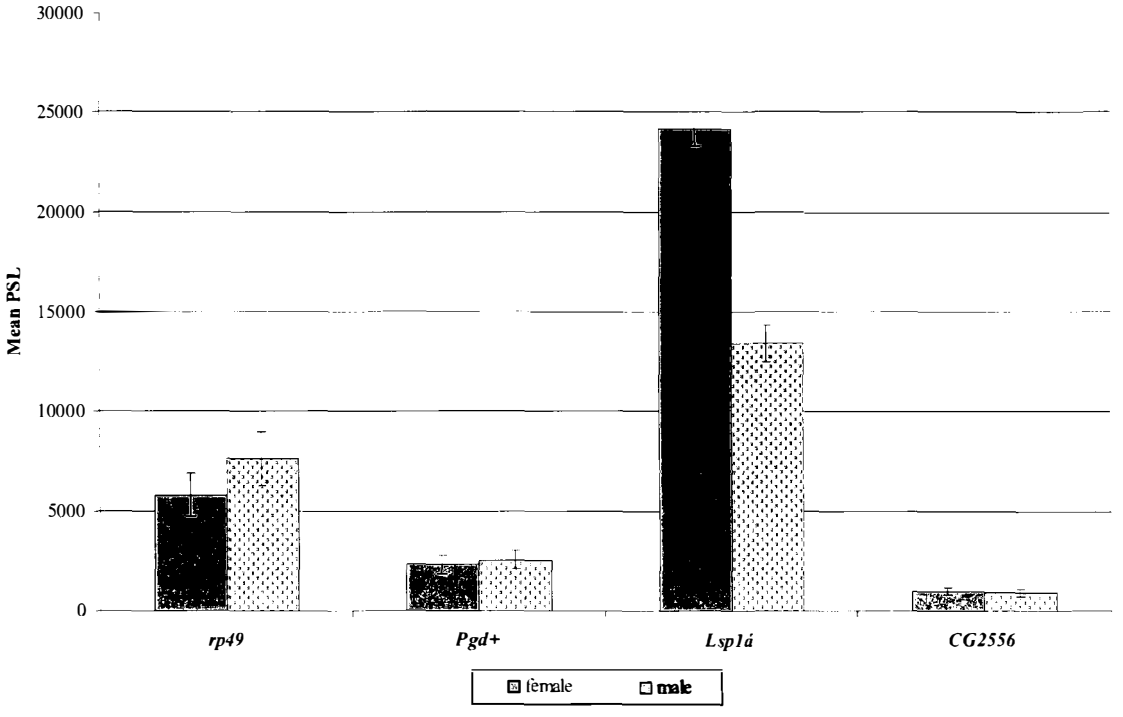


Figure 18. Mean *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcript levels in male and female third instar larvae staged by bromophenol blue-method as determined by RNase protection analysis

RNA extracted from female and male third instar larvae staged using the blue food method was subjected to RNase protection with antisense ssRNA probes corresponding to either *rp49*, *Pgd*, *Lsp1α* or *CG2556*. Samples were quantified using a FLA-5000 phosphorimager (Fujifilm) following gel electrophoresis. 95% confidence intervals are indicated for the mean PSL values shown below in Table 17.

Gene	Female		Male		Female: Male ratio
	Mean PSL	95% CI	Mean PSL	95% CI	of means
rp49	5786.31	1075.65	7607.01	1346.70	0.76
Pgd	2313.55	467.56	2542.91	439.48	0.91
Lsp1α	24178.13	861.54	13403.39	900.52	1.80
CG2556	968.90	173.58	937.36	183.58	1.03

Table 17. Mean *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcript levels in male and female third instar larvae staged by bromophenol blue-method as determined by RNase protection analysis

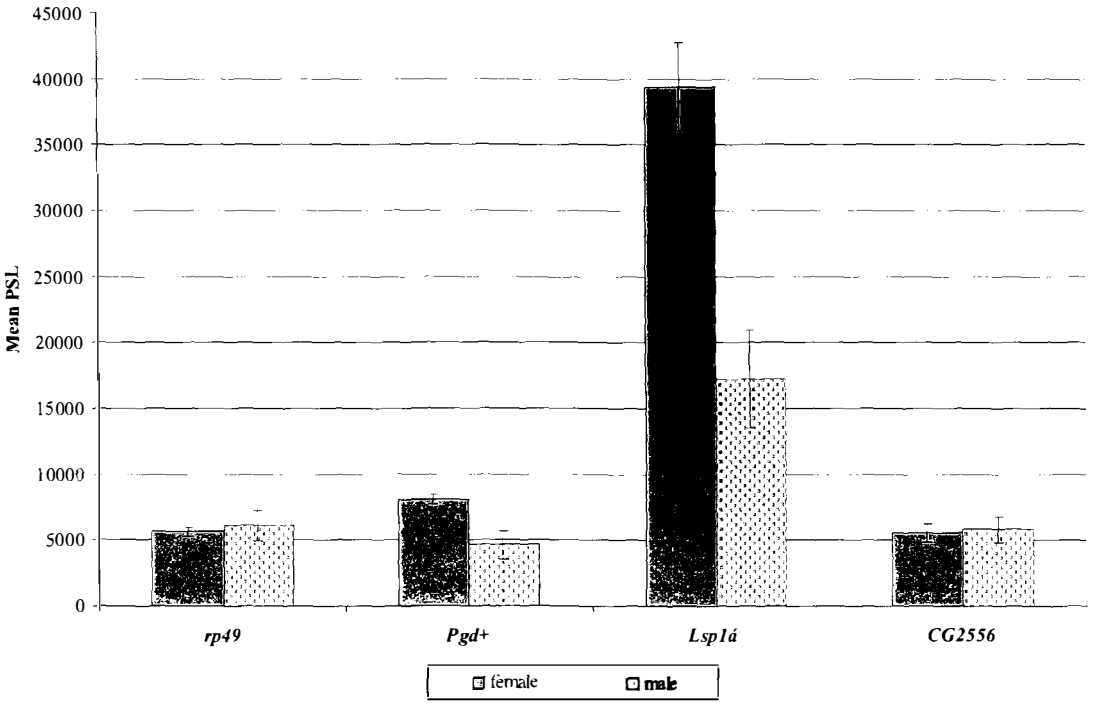


Figure 19. Mean *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcript levels in male and female third instar larvae staged by age/size as determined by RNase protection analysis

RNA extracted from female and male third instar larvae staged by age/size was subjected to RNase protection with antisense ssRNA probes corresponding to either *rp49*, *Pgd*, *Lsp1α* or *CG2556*. Samples were quantified using a FLA-5000 phosphorimager (Fujifilm) following gel electrophoresis. 95% confidence intervals are indicated for the mean PSL values shown below in Table 18.

Gene	Female		Male		Female: Male ratio
	Mean PSL	95% CI	Mean PSL	95% CI	of means
rp49	5644.68	356.07	6137.38	1161.63	0.92
Pgd	8088.47	358.93	4612.68	1078.60	1.75
Lsp1α	39429.60	3372.97	17262.77	3745.43	2.28
CG2556	5564.01	722.03	5847.66	983.67	0.95

Table 18. Mean *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcript levels in male and female third instar larvae staged by age/size as determined by RNase protection analysis

Gene	Female: Male ratios of pairs	Mean female: male ratio	95% CI
rp49	0.73 1.14 0.82 0.49	0.79	0.27
Pgd	0.78 1.32 0.62	0.91	0.42
Lsp1α	1.79 1.94 1.69	1.81	0.14
CG2556	0.83 1.07 1.06	0.99	0.15

Table 19. Female: male ratios of pairs, and mean ratios of *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcript levels in third instar larvae staged by the bromophenol blue-method

Gene	Female: Male ratios of pairs	Mean female: male ratio	95% CI
rp49	1.13 0.73 0.78 1.17	0.95	0.22
Pgd	1.80 1.67 1.34 2.56	1.84	0.51
Lsp1α	2.03 2.32 2.62	2.32	0.34
CG2556	1.37 0.75 0.83 0.99	0.98	0.27

Table 20. Female: male ratios of pairs, and mean ratios of *rp49*, *Pgd*, *CG2556* and *Lsp1α* transcript levels in third instar larvae staged by age/size

3.8 Expression of *CG2556* and *CG2560* in the fat body of third instar larvae

Lsp1α is exclusively expressed in the fat body of third instar larvae (Davies *et al.*, 1986). Both of the genes that flank *Lsp1α*, *CG2560* and *CG2556*, are expressed in third instar larvae (section 3.2, Figure 4). Real-time RT-PCR was conducted on fat body and whole larval cDNA to determine whether these flanking genes are expressed in the fat body tissue.

RNA was isolated from 10 whole male *y w* third instar larvae, and from fat bodies dissected from 10 male *y w* third instar larvae (section 2.6.1.1). RNA was DNase treated using the Turbo DNase (Ambion; section 2.6.1.3) to remove all contaminating genomic DNA. Treatment with other commercially available DNases did not remove residual DNA, and products were observed when real-time PCR was conducted on the RNA samples (data not shown). Quantitative real-time PCR was conducted in triplicate on cDNA generated from 1 μg of the DNase-treated RNA (sections 2.6.1.9, 2.6.7). PCR was also conducted on RNA, which had not been subjected to reverse transcription, and on water. These reactions produced products with crossing points >8 cycles after the cDNA reaction products, that were composed entirely of primer dimers as determined by melting curve analysis.

Quantitative real-time PCR was conducted on 100-fold dilutions of fat body and whole larval cDNA using the rp49-596F and rp49-720R; Lsp1a-1884F and Lsp1a-2085R; Lsd2-1490F and Lsd2-1674R; CG2560-696F and CG2560-875R; GpdhF and GpdhR primers (section 10.1). Quantitative real-time PCR was also conducted on 10-fold dilutions of fat body and whole larval cDNA using the CG2556-1129F and CG2556-1229R primers (section 10.1). The crossing point was automatically determined by the LightCycler software for each PCR reaction in triplicate. The mean crossing point was used in all further calculations, as the standard deviation was generally <0.2 cycles (section 10.8). Any outlying PCR values were discarded. The fold enrichment of a particular cDNA in fat body compared to whole larvae was determined using the equation: 2^(crossing point whole larvae – crossing point fat body).

Two genes expressed in the fat body of third instar larvae, *Lsp1α* and *Lsd-2*, are enriched 2.24 and 1.82-fold respectively in fat body tissue compared to whole larvae (Figure 20, Table 21). This high level of relative expression is consistent with their expression in fat body tissue. Two constitutively expressed genes, *rp49* and *Gpdh*, are enriched 0.79 and 0.78-fold respectively in fat body tissue compared to whole larvae. This level of relative expression of close to one is consistent with their expression throughout the larva. The low level of enrichment for *CG2560* indicates that this gene is not expressed in fat body tissue, and is consistent with the putative function of *CG2560* as a larval cuticle protein (11/04) (Ashburner *et al.*, 2000). *CG2556* is enriched 0.92-fold in fat body tissue compared to whole larvae. This level of relative expression is similar to that of the constitutively expressed genes *rp49* and *Gpdh*, and indicates that *CG2556* is expressed both in the fat body, and in other parts of the larva.

Gene	Fold enrichment
Lsp1α	2.24
Lsd-2	1.82
Gpdh	0.78
rp49	0.79
CG2556	0.92
CG2560	0.03

Table 21. Fold enrichment of specific cDNAs in fat body tissue compared to whole third instar larvae

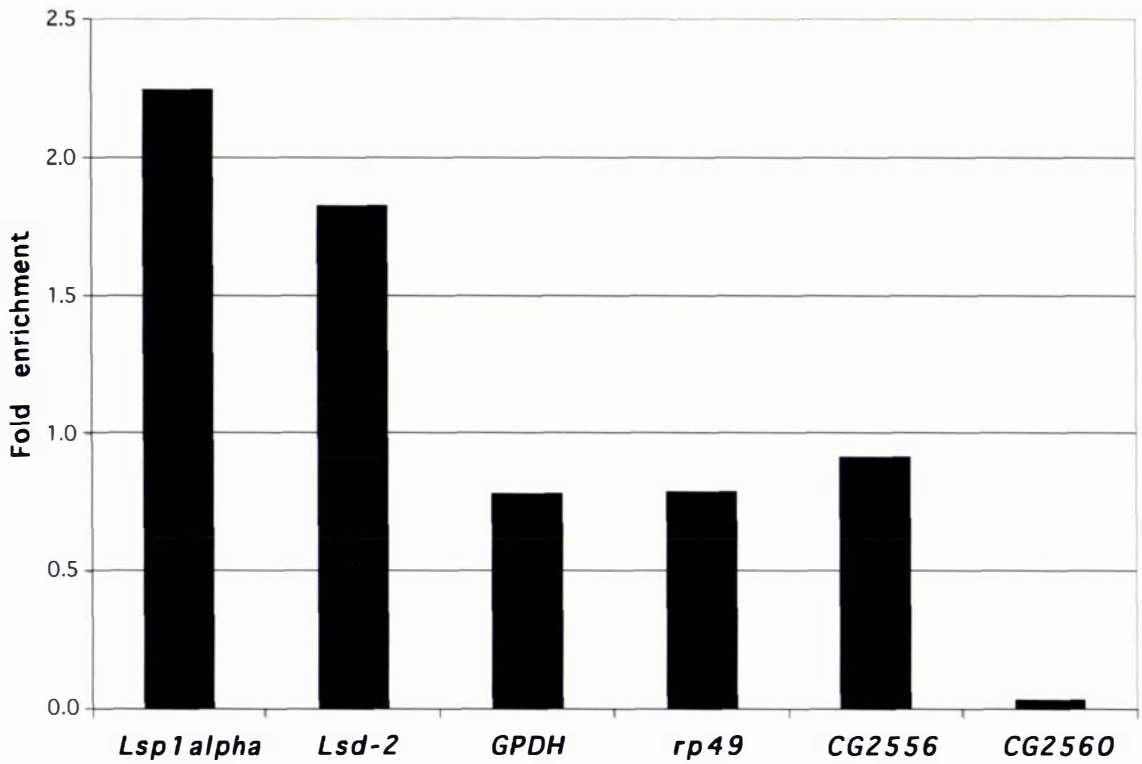


Figure 20. Fold enrichment of the genes flanking *Lsp1α* in fat body tissue compared to whole third instar larva

Real-time RT-PCR of *Lsp1α*, *Lsd-2*, *Gpdh*, *rp49*, *CG2556* and *CG2560* in male fat body and whole third instar male larvae cDNA. The fold enrichment of each transcript in fat body compared to whole larvae cDNA is shown.

4 MEASUREMENT OF PROMOTER ACTIVITY

Lsp1α is flanked by two dosage compensated genes, yet is expressed at two-fold higher levels in female larvae in comparison to male larvae. As *Lsp1α* is not enriched for the histone modification associated with dosage compensation by the MSL complex (chapter 6), it is unlikely that this gene is regulated by the MSL complex. If the MSL complex binds initially at the chromatin entry sites, then moves from these sites into actively transcribed genes on the X chromosome, then it might be supposed that *Lsp1α* escapes regulation by the MSL complex due to elements that lie between it and the flanking dosage compensated genes that are able to block the spread of the MSL complex. In order to test this hypothesis, constructs were generated in which the regions between *Lsp1α* and the flanking dosage compensated genes were inserted either side of an X-linked reporter gene (Figure 21B construct c). It was hypothesized that the reporter gene would escape dosage compensation by the MSL complex if these flanking regions contained elements able to block the spread of the MSL complex. Due to the previously published study by Ghosh *et al.* (1989), it was initially supposed at the beginning of this study that the *Lsp1α* promoter was unlikely to show any sex-specific regulation. However, in order to control for this possibility, constructs were made in which the *Lsp1α* promoter drives expression of the *lacZ* reporter gene (Figure 21B constructs b, d). A control construct consisting of the constitutive *armadillo* promoter driving *lacZ* expression (Figure 21B construct a) used in previous studies (Henry *et al.*, 2001) was included as a control.

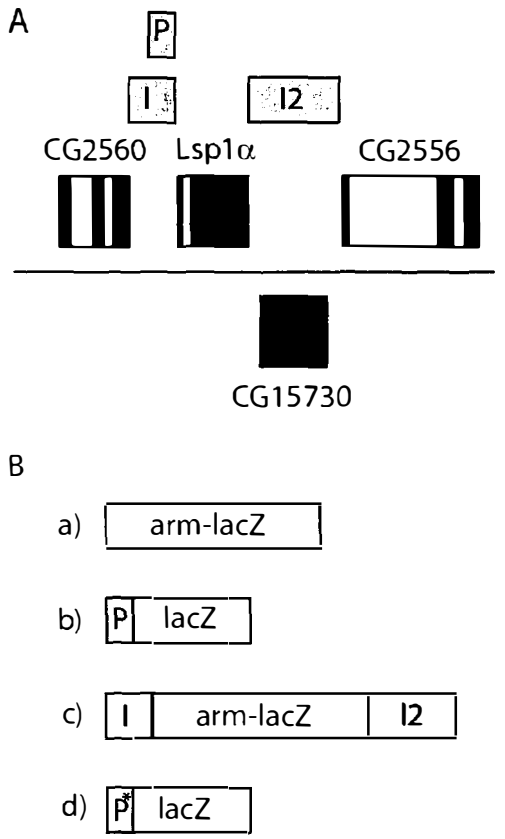


Figure 21. Schematic of constructs

(A) The regions used in the generation of constructs: the 593 bp *Lsp1α* promoter (P), 883 bp extended *Lsp1α* promoter (I) and 4596 bp region between *Lsp1α* and *CG2556* (I2) Exons (black) and introns (white) are indicated for each gene in the region. (B) The constructs used to generate transgenic fly lines: *arm-lacZ* (a), *P-lacZ* (b), *I-arm-lacZ-I2* (c), *I-lacZ-I2* (d) and *P*-lacZ* (e). *P*-lacZ* is identical to the *P-lacZ* construct, except for a mutation in a putative Doublesex (DSX) binding site in the promoter sequence.

4.1 Generation of constructs

4.1.1 Isolation of *Lsp1α* genomic regions

4.1.1.1 *Lsp1α* promoter into pGEM®-T Easy

The promoter region of *Lsp1α* required for full correct temporal and spatial expression was identified by deletion mapping as being a 570 bp region 5' of the translation start site (Delaney *et al.*, 1987). This region was amplified by PCR with the Lsp1pForw and Lsp1pRev primers (section 10.1) from *Drosophila y w* genomic DNA (method 1: annealing temperature of 45°C: 30X). The 593 bp expected product was gel purified, ligated directly into the pGEM®-T Easy vector (Promega), and transformed into *DH5α E. coli* cells. A 593 bp clone consisting of the promoter (-573 bp to +20 bp) was identified using restriction endonuclease digestion and confirmed by sequencing. There are only a few SNPs between the *y w* strain used for this cloning and the *Drosophila* genomic sequence (*y2; cn bw sp*), however these include two new *Eco*RI sites (section 10.6.2). This fragment is referred to as *Lsp1α* or (P).

4.1.1.2 Mutant *Lsp1α* promoter into pGEM®-T Easy

A putative DSX binding site in the *Lsp1α* promoter was mutated using PCR mutagenesis from TGTACAACGTGG to TGTTTCCCGTGG. The Lsp1pForw and Lsp1pdsxmutRev primers (section 10.1) were used to amplify a 297 bp product from *Drosophila y w* genomic DNA (method 1: annealing temperature of 50°C: 30X). The Lsp1pRev and Lsp1pdsxmutForw primers (section 10.1) were used to amplify a 309 bp product from *Drosophila y w* genomic DNA (method 1: annealing temperature of 50°C: 30X). The 297 bp and 309 bp amplification products were gel purified together, and used as a template for PCR with the Lsp1pForw and Lsp1pRev primers (section 10.1) (method 1: annealing temperature of 50°C: 30X). The 593 bp resulting product containing the ACAA to TTCC mutation in the putative DSX binding site was gel purified, ligated directly into the pGEM®-T Easy vector (Promega), and transformed into *DH5α E. coli* cells. This mutation was confirmed by sequencing (section 10.6.2). This fragment is referred to as *Lsp1αmut* or (P*).

4.1.1.3 Genomic region between *CG2560* and *Lsp1α* into pGEM®-T Easy

The 883 bp genomic region between the end of the 3' UTR of *CG2560* and the beginning of the *Lsp1α* ORF includes the *Lsp1α* promoter region. This fragment was

amplified by PCR with the InsulForw and Lsp1pRev primers (section 10.1) from *Drosophila y w* genomic DNA (method 1: annealing temperature of 45°C: 30X). The 883 bp product was gel purified, ligated directly into the pGEM®-T Easy vector (Promega), and transformed into *DH5α E. coli* cells. A clone consisting of this genomic region (-863 bp to +20 bp) was identified by restriction endonuclease digestion and confirmed by sequencing. There are only a few SNPs, and a single 10 bp insertion, between the *y w* strain used for this cloning and the *Drosophila* genomic sequence (*y2; cn bw sp*) (section 10.6.2). This fragment is referred to as (I).

4.1.1.4 Genomic region between *Lsp1α* and *CG2556* into pGEM®-T Easy

The genomic region between the end of the 3' UTR of *Lsp1α* and the beginning of the 5' UTR of *CG2556* was amplified by long range PCR with the Lsp1aCG2556F and Lsp1aCG2556R#2 primers (section 10.1) from *Drosophila y w* genomic DNA (method 3: annealing temperature of 45°C: 35X). The 4596 bp product obtained was gel purified, ligated directly into the pGEM®-T Easy vector (Promega), and transformed into *DH5α E. coli* cells. Its identity was confirmed by restriction endonuclease digestion and sequencing of ends. This fragment is referred to as (I2).

4.1.2 Construction of the pCaSpeR-Lsp1α-βgal and pCaSpeR-I-βgal reporter constructs

The *armadillo* promoter of pCaSpeR-arm-βgal (*arm-lacZ*) (Fitzsimons, 1998; Thummel *et al.*, 1988; Vincent *et al.*, 1994) was excised with *Eco*RI and *Asp*718, and the resulting plasmid CIP treated and gel purified. The *Eco*RI - *Asp*718 linker (section 10.4) containing the *Bam*HI, *Nhe*I and *Stu*I restriction endonuclease sites between the *Eco*RI and *Asp*718 cohesive ends was treated with polynucleotide kinase and ligated into the cut vector. This ligation reaction was transformed into methylation deficient *E. coli* (*dam-13::Tn9 dcm-6 hsdM hsdR2 recF143 mcrA mcrB*; (MacNeil *et al.*, 1992) cells. This strain of *E. coli* was used because the *Stu*I site within the linker was subject to dcm methylation. This replacement was confirmed by restriction endonuclease digestion and sequencing with the arm-lacZseq sequencing primer (section 10.3). This plasmid was named pCaSpeR-EA-βgal (section 10.5.3).

Either the 593 bp *Lsp1α* promoter, or the 883 bp genomic region between *CG2560* and *Lsp1α* was inserted in the sense orientation 5' of the *lacZ* reporter gene in pCaSpeR-EA-βgal. The 593 bp (P) or 883 bp (I) fragments were excised from the pGEM®-T Easy vector (Promega) using *Not*I and blunt-ended using Klenow polymerase followed by gel purification. These fragments were ligated into pCaSpeR-EA-βgal that had been digested with *Stu*I, CIP treated, and gel purified. These were then transformed into *DH5α E. coli* cells. These insertions were confirmed using restriction endonuclease digestion and sequencing with the arm-lacZseq sequencing primer (section 10.3). The resulting plasmids were named pCaSpeR-Lsp1α-βgal (*P-lacZ*) (section 10.5.4) and pCaSpeR-I-βgal (section 10.5.6).

4.1.3 Construction of the pCaSpeR-Lsp1αmut-βgal plasmid containing a mutation in the putative DSX binding site of the *Lsp1α* promoter

The 593 bp *Lsp1α* promoter containing the mutation in the putative DSX binding site (section 4.1.1.2) was inserted 5' of *lacZ* in pCaSpeR-EA-βgal. The mutant promoter was excised from the pGEM®-T Easy vector (Promega) using *Not*I and blunt-ended using Klenow fragment of *E. coli* DNA polymerase followed by gel purification. This fragment was ligated into pCaSpeR-EA-βgal that had been digested with *Stu*I, CIP treated, and gel purified. This was then transformed into *DH5α E. coli* cells. This replacement was confirmed by restriction endonuclease digestion and sequencing with the *arm-lacZ*seq sequencing primer (section 10.3). The resulting plasmid was named pCaSpeR-Lsp1αmut-βgal (*P*-lacZ*) (section 10.5.5).

4.1.4 Insertion of the putative insulator region between *Lsp1α* and *CG2556*, 3' of *lacZ* under the control of the extended *Lsp1α* promoter

A suitable restriction site was required for the insertion of putative insulator elements from pGEM®-T Easy into pCaSpeR-I-βgal. Due to the lack of *Pst*I sites in the MCS of pGEM®-T Easy, and the presence of flanking *Not*I sites in the MCS of pGEM®-T Easy, as well as the lack of *Not*I sites in pCaSpeR-I-βgal, a *Pst*I - *Not*I linker (section 10.4) was inserted in the *Pst*I site 3' of the *lacZ-SV40* reporter gene construct. The pCaSpeR-I-βgal vector was digested with *Pst*I, CIP treated and gel purified. The *Pst*I - *Not*I linker (section 10.4) that had been treated with polynucleotide kinase was ligated into the cut vector, and transformed into *DH5α E. coli* cells. Clones carrying the

plasmid with the *Pst*I - *Not*I linker were identified using restriction endonuclease digestion.

The 4596 bp genomic region between the end of the 3' UTR of *Lsp1α* and the beginning of the 5' UTR of *CG2556* was inserted 3' of *lacZ* in the sense orientation in pCaSpeR-I-βgal. The (I2) fragment was excised from the pGEM®-T Easy vector using *Not*I and gel purified. This 4596 bp *Not*I fragment was ligated into pCaSpeR-I-βgal containing the *Pst*I - *Not*I linker that had been digested with *Not*I, CIP treated and gel purified, and transformed into *DH5α E. coli* cells. Clones of this construct were identified using restriction endonuclease digestion, and confirmed by sequencing with the armlacZNotIseq sequencing primer (section 10.3). This plasmid was named pCaSpeR-I-βgal-I2 (*I-lacZ-I2*) (section 10.5.7).

4.1.5 Insertion of putative insulator regions either side of *arm-lacZ* in pCaSpeR-arm-βgal

A suitable restriction site was required for the insertion of putative insulator elements from pGEM®-T Easy into pCaSpeR-arm-βgal. Similarly to pCaSpeR-Lsp1α-βgal or pCaSpeR-I-βgal (section 4.1.4), as there are no *Not*I sites in pCaSpeR-arm-βgal, a *Pst*I - *Not*I linker (section 10.4) was inserted in the *Pst*I site 3' of the *arm-lacZ-SV40* reporter gene construct in the pCaSpeR-arm-βgal plasmid. The pCaSpeR-arm-βgal vector was digested with *Pst*I, CIP treated and gel purified. The *Pst*I - *Not*I linker that had been treated with polynucleotide kinase was ligated into the cut vector, and transformed into *DH5α E. coli* cells. Clones carrying pCaSpeR-arm-βgal with the *Pst*I - *Not*I linker were identified using restriction endonuclease digestion.

Due to the presence of an internal *Eco*RI site in the *CG2560-Lsp1α* genomic region (section 10.6.2), an alternative restriction endonuclease site was required for the insertion of putative insulator fragments from pGEM®-T Easy. There are no *Stu*I sites in pCaSpeR-arm-βgal, so the *Eco*RI - *Stu*I linker (section 10.4) was inserted in the *Eco*RI site 5' of the *arm-lacZ-SV40* reporter gene construct in the pCaSpeR-arm-βgal plasmid. The pCaSpeR-arm-βgal vector from the previous step was digested with *Eco*RI, CIP treated and gel purified. *Eco*RI - *Stu*I linker (section 10.4) that had been treated with polynucleotide kinase was ligated into the cut vector, and transformed into

DH5α E. coli cells. This enabled blunt end cloning of putative insulator fragments excised from pGEM®-T Easy with *Not*I. Clones carrying pCaSpeR-arm-βgal with the *Eco*RI - *Stu*I linker, as well as the *Pst*I - *Not*I linker were identified using restriction endonuclease digestion. The resulting plasmid was named pCaSpeR-arm-βgal *Stu*I/*Not*I (section 10.5.2).

The pCaSpeR-arm-βgal *Stu*I/*Not*I vector was digested with *Stu*I, CIP-treated, and gel-purified. The 883 bp (I) fragment was excised from pGEM®-T Easy with *Not*I. This *Not*I fragment was blunt-ended using Klenow polymerase followed by gel purification, and ligated into the *Stu*I site of the digested pCaSpeR-arm-βgal *Stu*I/*Not*I plasmid, 5' of the *arm-lacZ* reporter construct. This ligation reaction was transformed into *DH5α E. coli* cells. Clones carrying pCaSpeR-arm-βgal with the (I) region 5' of *arm-lacZ* in the sense orientation were identified using restriction endonuclease digestion, and confirmed by sequencing with the *arm-lacZ*seq sequencing primer, generating pCaSpeR-I-arm-βgal (section 10.3).

Following this, pCaSpeR-I-arm-βgal was digested with *Not*I, CIP treated, and gel purified. The 4596 bp (I2) region was excised from the pGEM®-T Easy vector using *Not*I and gel purified. This 4596 bp *Not*I fragment was ligated into the cut pCaSpeR-arm-βgal *Stu*I/*Not*I vector, and transformed into *DH5α E. coli* cells. Clones carrying pCaSpeR-I-arm-βgal with the 4596 bp (I2) region 3' of *arm-lacZ* in the sense orientation were identified using restriction endonuclease digestion, and confirmed by sequencing with the armlacZNotIseq sequencing primer (section 10.3). This plasmid was named pCaSpeR-I-arm-βgal-I2 (*I-arm-lacZ-I2*) (section 10.5.8).

4.2 Generation of transgenic fly lines carrying reporter constructs

The plasmids constructed in (section 4.1) were used to generate transgenic fly lines by *P*-element mediated transformation (section 2.5.3). A *y w* stock was used as the recipient strain, and transgenic lines were detected by the presence of a *mini-white* marker gene. All fly lines used in this study were characterized by linkage analysis (section 2.5.2.1), inverse PCR (section 2.6.3.1) and Southern hybridization analysis (section 2.6.2.3). The copy number of the insert as determined by Southern

hybridization analysis, and the chromosomal position of the insert as determined by inverse PCR are described in Table 22.

4.2.1 Characterization of transgenic *Drosophila* lines

The copy number of the transgene was determined for each transgenic *Drosophila* line using Southern hybridization analysis. Genomic DNA was extracted from each of the fly lines and subjected to digestion with *Eco*RV or *Cla*I. Digested DNA was probed with a 4.7 kb *Hind*III/*Eco*RI fragment from pCaSpeR-arm-βgal consisting of the 5' *P*-element and *mini-white* marker gene (Figure 22).

Both *Eco*RV and *Cla*I cut in the middle of the *mini-white* gene, hence two fragments would be expected from each transgene and two fragments from the endogenous *white* gene for each transgenic fly line. *Cla*I digestion of *y w* genomic DNA produces a strongly hybridizing band at *ca.* 2 kb and a weaker product at *ca.* 10 kb representing the endogenous *white* gene. The weaker band at 10 kb is not always observed in these Southerns due to the high level of background. *Eco*RV digestion produces two strong hybridization products that migrate at *ca.* 2 kb and *ca.* 5 kb.

*Cla*I digestion of the transgene produces two fragments, the size of which depends on the construct, as *Cla*I cuts at various places within the *armadillo,* (P) or (I) sequences, and the genomic region in which the transgene has inserted. Consistent products of 5.502 kb, 4.374 kb, 6.407 kb and 4.664 kb are expected for each of the *arm-lacZ*, *P-lacZ* or *P*-lacZ*, *I-arm-lacZ-I2* and *I-lacZ-I2* transgenes respectively. The correct products are present in each of these lines (Table 22), with the exception of lane 21, which was incorrectly loaded. Only one extra band corresponding to the remainder of the 5' *P-mini-white* transgene plus genomic sequence is expected for single copy transgenes. Only a single band is present for all *arm-lacZ* lines with the exception of *arm-lacZ:68D3* (lane 9), which appears to have two transgene inserts. An extra band may also be present in *P*-lacZ:75D4* (lane 22), indicating more than one transgene insertion, although the weaker product of *ca.* 10 kb expected from the endogenous *white* gene is not present in this sample. It is possible that this line only contains a single copy of the transgene, and that a new *Cla*I site has been generated in the endogenous *white* gene.

*Eco*RV has an extra unmarked recognition site near the *Eco*RI site at the start of the *mini-white* gene[1], hence a 2.987 kb band is present in all transgenic lines regardless of the type of construct. A weaker band that hybridizes to the remainder of the *ca.* 1.7 kb 5' *P-mini-white* gene and genomic sequence is present in all transgenic lines, but is absent in *y w* flies. The presence of only one extra band indicates that there is only a single copy of the transgene present in each line. As observed with *Cla*I digestion, an extra weakly hybridizing band may be present in *arm-lacZ:68D3* (lane 9), indicating the presence of two transgenes. No extra band in present in lane 21, indicating that this line probably represents a single copy transgene insertion.

It is unclear whether the X-linked *I-lacZ-I2* line (lane 24) has one or two copies of the transgene by Southern hybridization analysis, due to the poor transfer in the *Cla*I gel and high background on that lane in the *Eco*RV gel. Furthermore, the insertion point of the transgene(s) could not be determined for this line by inverse PCR. However, the high activity level of the transgene relative to its female/male activity ratio indicates that this line may have more than one copy of the transgene and so this line is not included in the graphs of female/male activity ratio versus male activity/copy.

These results indicate that all the transgenic lines used in this study, with the exception of *arm-lacZ:68D3*, and possibly also *P*-lacZ:75D4* and *I-lacZ-I2:X* represent single copy insertions of the transgene. However, only single insertion sites were detected using inverse PCR for both these lines although Southern hybridization analysis indicates that these may have two insertions of the transgene.

The genes closest to the insertion site of the X-linked transgenes as determined by inverse PCR are identified in Table 23. The *I-arm-lacZ-12:19C3* transgene has inserted between the X-linked *CG1631* and *CG15462* genes, that are uncharacterised with respect to function and expression. The *arm-lacZ:10D8* transgene has inserted in the first intron of the X-linked *inaF* (*CG2457*) gene, which encodes a protein with calcium channel regulator activity involved in rhodopsin mediated signalling that is expressed in the head and eye of adult flies (Li *et al.*, 1999). The *P-lacZ:9B4* transgene has inserted

[1] This site is not present in the sequence, but is recognised during restriction endonuclease digestion, and has been marked on the plasmid maps.

in the first intron of the X-linked *CG15309* gene, which has an unknown function and expression pattern but exhibits similarity to the Yippee family of putative zinc finger binding proteins. The *P-lacZ:12A2* transgene has inserted in the first intron of the X-linked *NFAT* (*CG11172*) gene, which is a transcription factor expressed in many different tissues throughout development (Huang and Rubin, 2000). The *P-lacZ:19E7* transgene has inserted between the X-linked *CG1529* and *Ntf-2* (*CG1740*) genes. *CG1529* encodes a putative zinc finger DNA binding protein with an unknown expression pattern. *The Ntf-2* gene product is involved in the import of proteins to the nucleus, and is expressed in a variety of tissues and developmental stages as the phenotype of various mutant alleles ranges from reduced eyes to larval lethality (Bhattacharya and Steward, 2002). Significantly, some *Ntf-2* mutants have phenotypes that manifest in the fat body tissue of third instar larvae, indicating that *Ntf-2* is expressed in this tissue (Bhattacharya and Steward, 2002). The *P*-lacZ:4D6* transgene has inserted between the X-linked *CG4068* and *Ptp4E* (*CG6899*) genes. The function and expression pattern of *CG4068* have not been characterised, while *Ptp4E* encodes a protein tyrosine phosphatase expressed widely during embryogenesis (Yang *et al.*, 1991). The *P*-lacZ:9B12* transgene has inserted in the first exon of the X-linked *CG17841* gene, which is essential as it exhibits a mutant lethal phenotype (Bourbon *et al.*, 2002).

Lane[1]	Transgene Position[2]	Plasmid used to generate transgene	construct	Transgene copy number
1	70B3	pCaSpeR-arm-βgal	*arm-lacZ*	1
2	65E8	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
3	9B4	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
4	12A2	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
5	83A1	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
6	39A1	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
7	75F6	pCaSpeR-I-arm-βgal-I2	*I-arm-lacZ-I2*	1
8	47A13	pCaSpeR-I-βgal-I2	*I-lacZ-I2*	1
9	68D3	pCaSpeR-arm-βgal	*arm-lacZ*	2
10	62D7	pCaSpeR-arm-βgal	*arm-lacZ*	1
11	19E7	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
12	50C17	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
13	2	pCaSpeR-Lsp1α-βgal	*P-lacZ*	1
14	23A3	pCaSpeR-I-arm-βgal-I2	*I-arm-lacZ-I2*	1
15	57A6	pCaSpeR-I-arm-βgal-I2	*I-arm-lacZ-I2*	1
16	21A2	pCaSpeR-I-arm-βgal-I2	*I-arm-lacZ-I2*	1
17	19C3	pCaSpeR-I-arm-βgal-I2	*I-arm-lacZ-I2*	1
18	10D8	pCaSpeR-arm-βgal	*arm-lacZ*	1
19	79A4	pCaSpeR-arm-βgal	*arm-lacZ*	1
20	3	pCaSpeR-Lsp1αmut-βgal	*P*-lacZ*	1
21	4D6	pCaSpeR-Lsp1αmut-βgal	*P*-lacZ*	1
22	75D4	pCaSpeR-Lsp1αmut-βgal	*P*-lacZ*	1-2?
23	9B12	pCaSpeR-Lsp1αmut-βgal	*P*-lacZ*	1
24	X	pCaSpeR-I-βgal-I2	*I-lacZ-I2*	1?
25	-	-	*y w*	-
26	45D4	pCaSpeR-arm-βgal	*arm-lacZ*	1

Table 22. The copy number and chromosomal position of the transgene in each line as determined by Southern hybridization analysis and inverse PCR

[1] Lane in Southern hybridization analysis (Figure 22).

[2] Determined by inverse PCR or linkage analysis.

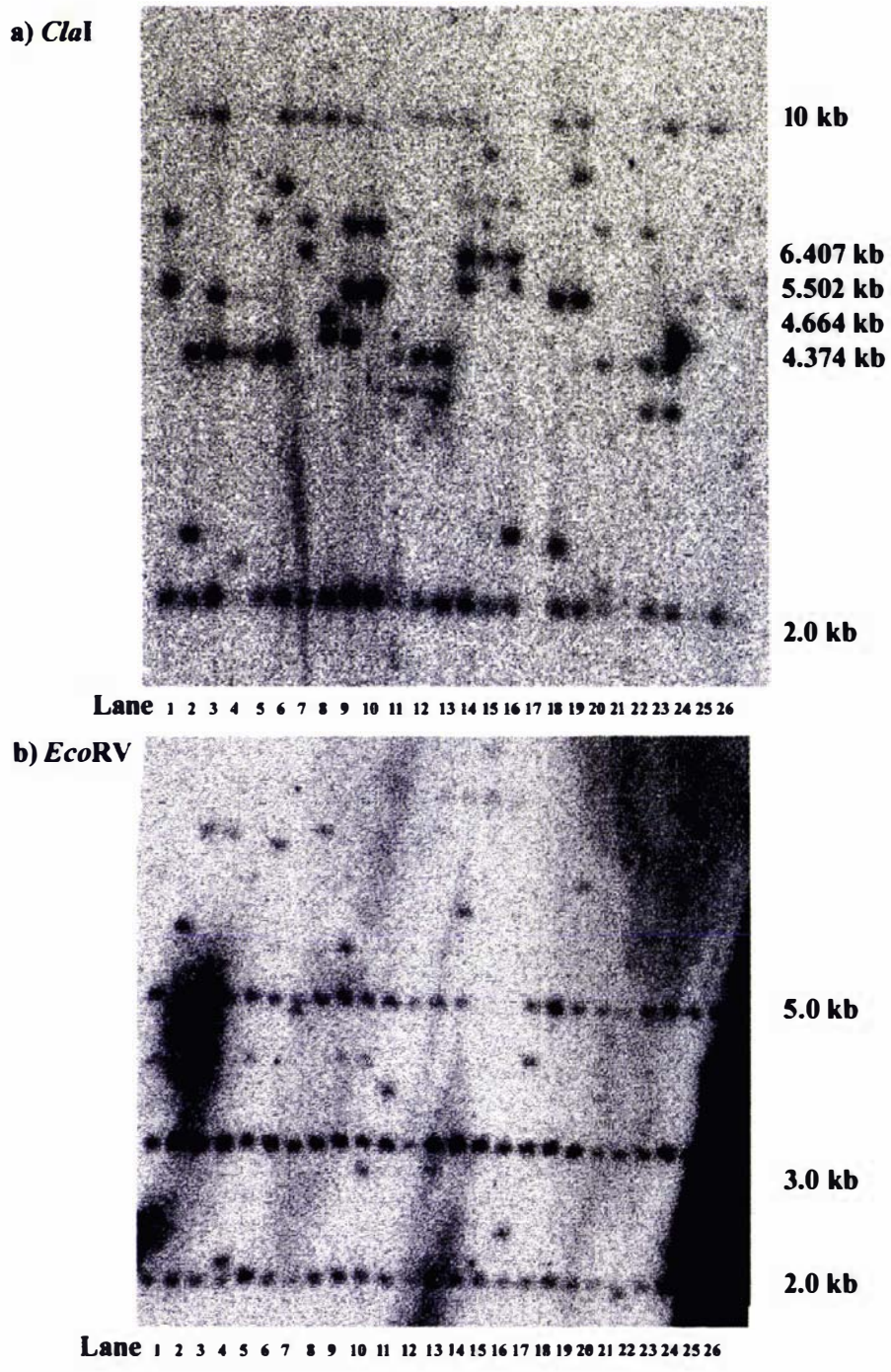


Figure 22. Southern hybridization analysis of transgenic fly lines

Genomic DNA was extracted from each of the fly lines described in Table 22 and subjected to digestion with *Cla*I (A) or *Eco*RV (B). Digested DNA was probed with a 4.7 kb *Hind*III/*Eco*RI fragment from pCaSpeR-arm-βgal consisting of the 5' *P*-element and *mini-white* marker gene.

Transgenic line	Chromosomal position determined by inverse PCR
I-arm-lacZ-I2:19C3	Between *CG1631* and *CG15462*
arm-lacZ:10D8	In the first intron of *inaF* (*CG2457*)
P-lacZ:9B4	In the first intron of *CG15309*
P-lacZ:12A2	In the first intron of *NFAT* (*CG11172*)
P-lacZ:19E7	Between *CG1529* and *Ntf-2* (*CG1740*)
P-lacZ:4D6*	Between *CG4068* and *Ptp4E* (*CG6899*)
P-lacZ:9B12*	In the first exon of *CG17841*

Table 23. The position of insertion of transgenes under the control of the *Lsp1α* promoter on the X chromosome

4.2.2 β-galactosidase activity is linear over a large range

Twenty third instar larvae of the genotype *lsp1α-lacZ:9B4* were staged using the bromophenol blue food method (section 2.5.4.2) or by age/size on standard cornmeal/molasses food (sections 2.5, 2.5.4.1). These larvae were homogenized in 2 ml of phosphate assay buffer, and 0, 10, 20, 30 and 50 µl of the homogenate assayed for β-galactosidase activity in triplicate (section 2.7.1). The mean ΔAbsorbance/Time is plotted against the volume of homogenate assayed in Figure 23. β-galactosidase activity is linear over the entire range of absorbance assayed using both staging methods, with the highest absorbance's measured in both reading >2.6. All β-galactosidase assays described in the following sections were conducted on larvae staged using the bromophenol blue method, or on hemisected adults.

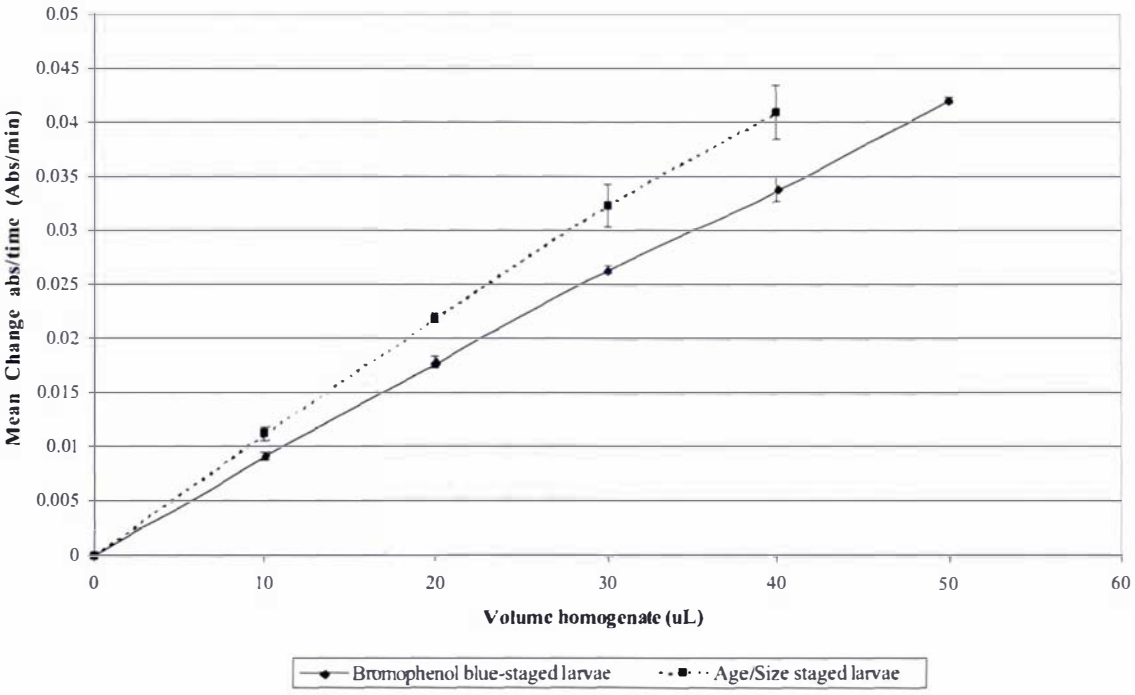


Figure 23. β-galactosidase activity increases in a linear manner with increasing homogenate volume

Mean ΔAbsorbance^{574nm}/time is plotted against increasing volumes of homogenate (0 – 50 µl) consisting of homogenized *lsp1α-lacZ:9B4* third instar larvae staged using the bromophenol blue method or by age/size on standard food. 95% confidence intervals are indicated for 3 replicates.

4.2.3 The difference in the level of protein between female and male third instar larvae does not affect the female/male β-galactosidase activity ratio

Female third instar larvae are slightly bigger than male larvae, and contain around 20% more protein than male larvae. To determine whether this discrepancy affected the female/male β-galactosidase activity ratios obtained, in which β-galactosidase activity is normalized to protein levels (section 2.7), the female/male activity ratio obtained for 6 female and 5 male *arm-lacZ:79A4* third instar larvae was compared to that obtained for 5 female and 5 male *arm-lacZ:79A4* third instar larvae. The mean female/male ratio of OD/min/μg protein using 6 female and 5 male larvae was 0.66 ± 0.05, while that obtained using 5 female and 5 male larvae was 0.70 ± 0.04. Thus the uneven protein levels in female and male third instar larvae do not affect the female/male β-galactosidase activity ratio when activities are normalized to protein levels. Due to this finding, an equal number of male and female third instar larvae were assayed to determine the female/male β-galactosidase activity ratio in all transgenic lines.

4.3 The regions flanking *Lsp1α* do not block dosage compensation of an X-linked transgene

The genes flanking *Lsp1α* are dosage compensated, while *Lsp1α* appears non-dosage compensated (sections 3.5, 3.7). In order to determine whether the genomic regions between *Lsp1α* and the neighbouring genes contained elements able to block the MSL complex from moving into, and dosage compensating *Lsp1α*, these genomic regions were cloned and placed either side of the *arm-lacZ* reporter construct (section 4.1.5, Figure 21). The 883 bp genomic region between the end of the 3' UTR of *CG2560* and the beginning of the *Lsp1α* ORF (I) contains all the elements required for stage and tissue-specific expression of *Lsp1α* (Delaney *et al.*, 1987). Third instar larvae of transgenic fly lines carrying these *I-arm-lacZ-I2* constructs were indistinguishable from transgenic fly lines carrying the *P-lacZ* constructs in so far as these lines produced as much β-galactosidase activity as *P-lacZ* lines, and significantly more than *arm-lacZ* lines (Table 25). It appears that the (I) sequence is either acting as an enhancer or as an upstream promoter. Therefore, it was impossible to determine whether the genomic regions between *Lsp1α* and the neighbouring genes contained elements able to block dosage compensation using third instar larvae, because of the increased reporter gene activity in females caused by the *Lsp1α* promoter sequences. Thus, reporter gene activity of transgenic fly lines carrying the *I-arm-lacZ-I2* constructs was measured in adult flies (section 2.7), as the *armadillo* promoter is active at all stages of *Drosophila* development (Riggleman *et al.*, 1989). In adult flies, the female/male β-galactosidase activity ratio of both X-linked and autosomal *arm-lacZ* lines is *ca.* 1 (Table 24). The autosomal *arm-lacZ:68D3* and *arm-lacZ:79A4* lines show female/male ratios of 0.99 ± 0.09 and 0.92 ± 0.09 respectively. The X-linked *arm-lacZ:10D8* line has a female/male ratio of 1.12 ± 0.26. This is consistent with previously published data (Henry *et al.*, 2001). Transgenic fly lines carrying the *I-arm-lacZ-I2* construct inserted on an autosome, should also exhibit female/male β-galactosidase activity ratios of *ca.* 1, as the MSL complex should only act on X-linked transgenes. As expected, the autosomal *I-arm-lacZ-I2:21A2* and *I-arm-lacZ-I2:57A6* lines show female/male ratios of 1.00 ± 0.05 and 0.99 ± 0.07 respectively. If the genomic regions flanking *Lsp1α* are able to block the MSL complex from moving into, and dosage compensating a reporter gene flanked by these regions, then an X-linked *I-arm-lacZ-I2* line would not be dosage compensated. Hence, females would have twice the level of β-galactosidase activity as

males, and flies homozygous for the construct would have female/male β-galactosidase activity ratios of *ca.* 2. However, the X-linked *1-arm-lacZ-12:19C3* line has a female/male ratio of 1.11 ± 0.06. This indicates that the genomic regions between *Lsp1α* and the neighbouring genes do not contain elements able to block the MSL complex from moving into, and dosage compensating a reporter gene. Hence the MSL complex is not preventing from spreading into *Lsp1α* by boundary elements. However, it is possible that further uncharacterized elements may lie within the large first intron of the gene 3' of *Lsp1α*, *CG2556*. Also, there may be stage or tissue-specific regulatory elements that lie within these regions able to block dosage compensation of *Lsp1α* only in fat body tissue and/or in third instar larvae.

		Dose			Mean Female/ Male ratio	Mean male activity/ copy	Mean female activity/ copy
Construct	location	Male	Female	n[1]	of activity[2]		
arm-lacZ	68D3[3] (A)	1	1	3	0.99±0.09[4]	2.36±0.13	2.33±0.09
arm-lacZ	79A4 (A)	2	2	3	0.92±0.09	2.61±0.16	2.39±0.12
arm-lacZ	10D8 (X)	1	2	3	1.12±0.26	4.27±0.51	2.36±0.27
1-arm-lacZ-12	21A2 (A)	2	2	3	1.00±0.05	2.46±0.43	2.45±0.35
1-arm-lacZ-12	57A6 (A)	2	2	3	0.99±0.07	2.62±0.20	2.58±0.13
1-arm-lacZ-12	19C3 (X)	1	2	3	1.11±0.06	4.02±0.27	2.22±0.12

Table 24. Mean male and female β-galactosidase activities and ratios in adult flies

[1] Number of independent experiments.

[2] 100μOD min^{-1} μg $protein^{-1}$

[3] This line contains 2 insertions of the transgene.

[4] 95% confidence intervals are indicated.

4.4 Autosomal insertions of X-linked promoters show increased activity in males

While female and male β-galactosidase activity levels were equal for the *arm-lacZ* lines in adult flies (Table 24), these levels were not equal in third instar larvae (Table 25). An increase in β-galactosidase activity in males carrying autosomal, but not X-linked, insertions of the *arm-lacZ* and *pgd-lacZ* (Scott and Lucchesi, 1991) reporter constructs was observed. The mean female/male β-galactosidase activity ratio for the 5 autosomal *arm-lacZ* lines is 0.77 ± 0.04, and is 0.73 ± 0.08 for the single autosomal *pgd-lacZ* line (Table 25). In contrast the single X-linked *arm-lacZ:10D8* and *pgd-lacZ* lines have female/male β-galactosidase activity ratios of 1.07 ± 0.04 and 1.09 ± 0.07 respectively. This suggests that the X-linked promoters *armadillo* and *pgd* are partially upregulated in males when inserted on an autosome in larvae, but not in adults. However, only a single X-linked line was obtained for both the *arm-lacZ* and *pgd-lacZ* gene constructs making it difficult to draw firm conclusions. For this reason, the female/male β-galactosidase activity ratios from autosomal lines containing the *Lsp1α* promoter driving *lacZ* expression are shown both with and without correction for this decreased autosomal ratio (Figure 24, section 4.5.1).

4.5 *Lsp1α* promoter activity is higher in female larvae

As discussed at the beginning of the chapter, constructs were made with the *Lsp1α* promoter fused to *lacZ* to determine if the promoter is more active in females than in males. Previously published studies indicated this should not be the case (Ghosh *et al.*, 1989). Surprisingly, many *P-lacZ* lines (Figure 21B, b) showed higher activity in female larvae in comparison to male larvae (Table 25). This indicated that the *Lsp1α* promoter could be sex-specifically regulated. The *I-arm-lacZ-I2* and *I-lacZ-I2* lines also showed higher activity in female compared to male larvae (Table 25). The *I-arm-lacZ-I2* finding is surprising, as *arm-lacZ* alone is not more active in female larvae relative to males. It appears that the *Lsp1α* promoter is hyperactivating *lacZ* as β-galactosidase levels of the *I-arm-lacZ-I2* lines are 5 – 10 fold higher than those of *arm-lacZ* alone as discussed previously (section 4.3). The (I) and (P) genomic fragments contain all the elements necessary for the stage and tissue-specific expression of the reporter gene.

In contrast with the findings of Ghosh *et al.* (1989), many of the transgenic *I-arm-lacZ-I2*, *I-lacZ-I2* and *P-lacZ* lines had female/male β-galactosidase activity ratios greater than one (Table 25). These ranged from 0.90 ± 0.09 to 1.85 ± 0.30 for autosomal insertions, with a median of 1.33. If these ratios are corrected for the difference between autosomal and X-linked insertions (section 4.4), the range alters to between 1.16 ± 0.09 and 2.39 ± 0.30, with a median of 1.72. Homozygous X-linked lines had ratios ranging from 1.61 ± 0.20 to 2.42 ± 0.59, with a median of 1.87. These data strongly indicate that the *Lsp1α* promoter is sex-specifically regulated, as autosomal transgenes show increased activity in females, and that the activity of this promoter is higher in female compared to male larvae. A potential explanation for the high level of variability in the female/male β-galactosidase activity ratios between the different transgenic lines is given below (section 4.5.1).

In order to discount the possibility that the level of *Lsp1α* promoter activity is higher in female larvae because the fat body constitutes a larger proportion of the larva in comparison to male larvae, the level of protein in male and female dissected fat bodies relative to whole third instar larvae was measured (Table 26). The protein levels in dissected fat bodies from three collections of single male and female third instar larvae were measured in triplicate (section 2.7.2). These were compared to the protein levels from three collections of single male and female whole third instar larvae measured in triplicate. Surprisingly, the proportion of fat body as a percentage of the entire larvae is 45% for male larvae and 26% for female larvae. Thus a fat-specific enzyme normalized to total protein would have higher activity in males than in females. While attempts were made to dissect out all of the fat body tissue, it is possible that these dissections may be more efficient with male larvae. However, if the difference reflects the anatomy of male and female larvae, it may explain why *arm-lacZ* and *Pgd-lacZ* activities were higher in male larvae than in female larvae. However, this finding does not explain the higher activity of the *P-lacZ* constructs in female larvae. The higher percentage of protein in the fat body dissected from male larvae may be due to the presence of the large male gonads within this tissue. It is not possible to remove these from the fat body tissue during dissection, and these are significantly larger than those of the female during third instar larval development.

There are several possible explanations for the discrepancies observed between the data from this study and those of Ghosh *et al.* (1989). Their approach involved taking the genomic region containing *Lsp1α*, including most of the *CG2560* and *CG15730* genes, and inserting this elsewhere in the *Drosophila* genome. The transgene was distinguished from the endogenous *Lsp1α* by the insertion of 500 bp of mouse DNA into the ORF, and transcript levels were measured in male and female third instar larvae by northern analysis relative to *sgs-3* and *rp49*. The use of this method and the inclusion of the extra sequence and *Lsp1α* ORF in the transgene may have influenced their results.

However, closer analysis of their data reveals that one of the two X-linked transgenes examined in their study has a male/female transcript ratio of *ca.* 0.80, which is consistent with higher promoter activity in females. Furthermore, 4 of the 5 autosomal transgenes have roughly equivalent transcript levels between males and females, although one autosomal line also exhibits higher activity in females.

4.5.1 The female to male promoter activity ratio decreases with increased activity

The level of upregulation of the *Lsp1α* promoter in females is highly variable between transgenic lines, and is proportional to the activity of the transgene (Figure 24). Lines exhibiting low levels of activity in male larvae have high female to male activity ratios, while lines with high activity levels have female to male activity ratios close to one. When the data is graphed, a clear linear relationship between the level of autosomal transgene activity in males and the female to male activity ratio of the same transgene is observed (Figure 24). This may be caused by increased sensitivity of the *Lsp1α* promoter to position effects in males, but not females. The eye colour of all lines is due to the expression of *mini-white* marker gene, which is immediately adjacent to the *P-lacZ* transgene. The *mini-white* marker is known to be particularly sensitive to position effects (Kelley and Kuroda, 2003). However, lines with particularly low levels of β-galactosidase activity have eye colour comparable to that of lines exhibiting high levels of activity. This indicates that these position effects may be stage or tissue specific, rather than representing general heterochromatin repression.

Autosomal lines have equal copies of the reporter construct in females and males. Homozygous X-linked lines have twice the number of copies in females as in males.

Despite this difference in copy number, the X-linked lines have female to male activity ratios equivalent to those of the autosomal lines that have the same level of transgene expression (Figure 24). This suggests that the single male copy of the X-linked *P-lacZ* transgene is producing the equivalent output of two autosomal copies, indicating hypertranscription by the MSL complex. The presence of active MSL complex on two of these X-linked transgenes was subsequently confirmed using chromatin immunoprecipitation (section 6).

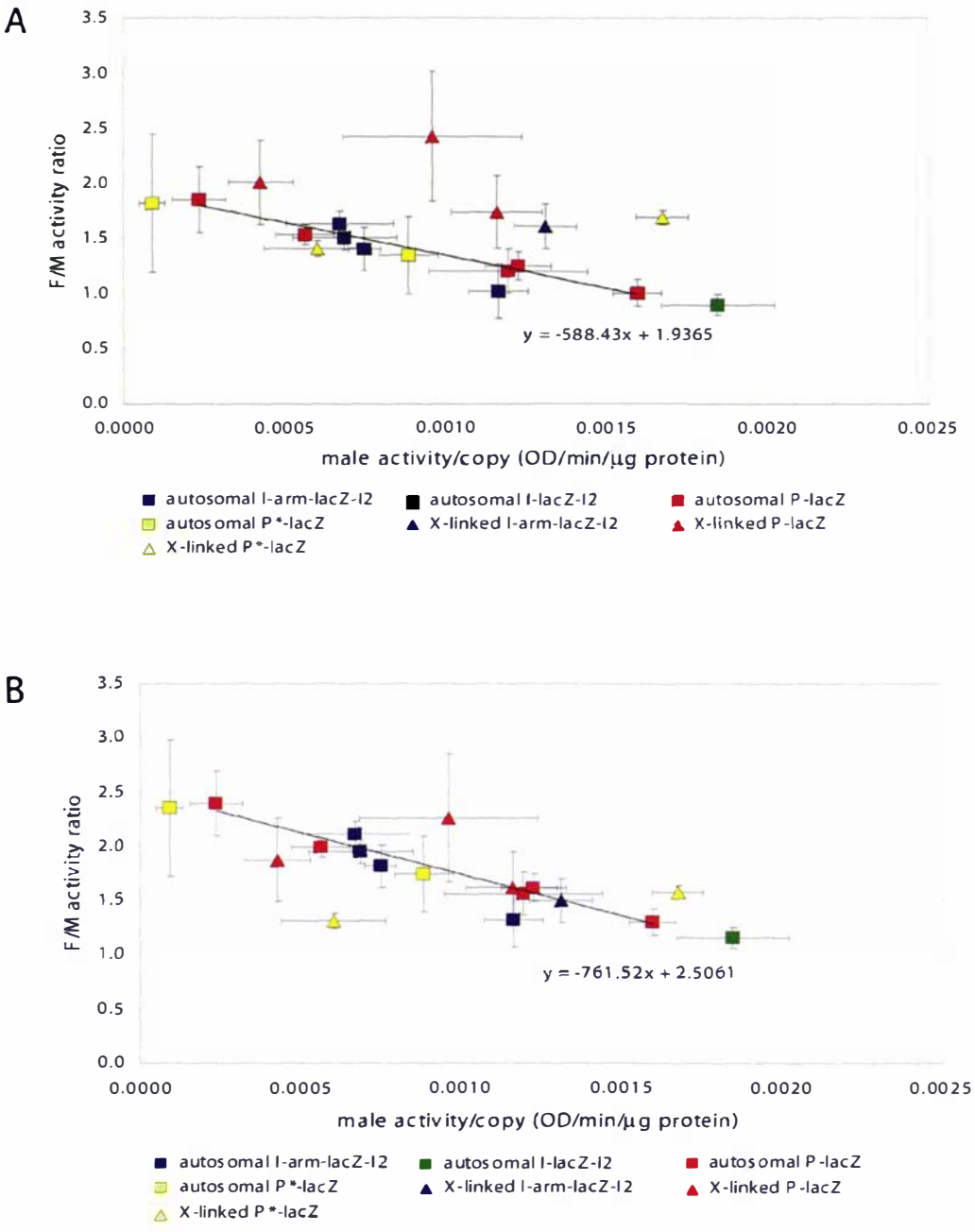


Figure 24. The female to male activity ratio is proportional to the level of male activity

(A) The mean female/male β-galactosidase activity ratio in blue-food staged third instar larvae graphed against mean male β-galactosidase activity/gene copy for X-linked and autosomal *P-lacZ, P*-lacZ, I-lacZ-I2* and *I-arm-lacZ-I2* transgenic fly lines. 95% confidence intervals are indicated for 3 experiments. (B) As in panel (A) but corrected for the difference between the X-linked and autosomal *arm-lacZ* female/male ratios in larvae. The trendlines are based on the autosomal *P-lacZ* lines.

Construct	location	Dose		n[1]	Mean Female/ Male ratio of activity[2]	Mean male activity/copy	Mean female activity/copy
		Male	Female				
arm-lacZ	68D3[3] (A)	1	1	3	0.74±0.08[4]	1.44±0.09	1.06±0.12
arm-lacZ	70B3 (A)	1	1	3	0.80±0.04	1.43±0.03	1.14±0.04
arm-lacZ	62D7 (A)	1	1	3	0.81±0.06	0.99±0.02	0.81±0.07
arm-lacZ	79A4 (A)	2	2	3	0.70±0.04	1.48±0.06	1.04±0.07
arm-lacZ	45D4 (A)	2	2	3	0.82±0.03	1.77±0.02	1.45±0.06
arm-lacZ	10D8 (X)	1	2	3	1.07±0.04	2.09±0.01	1.12±0.03
pgd-lacZ	3 (A)	2	2	5	0.73±0.08	0.04±0.00	0.03±0.00
pgd-lacZ	X	1	2	3	1.09±0.07	0.18±0.02	0.10±0.00
I-arm-lacZ-12	57A6 (A)	2	2	3	1.40±0.20	7.50±0.48	9.00±2.17
I-arm-lacZ-12	21A2 (A)	2	2	3	1.50±0.11	6.89±1.62	10.27±1.61
I-arm-lacZ-12	23A3 (A)	2	2	3	1.63±0.12	6.73±1.66	10.87±2.02
I-arm-lacZ-12	75F6 (A)	2	2	3	1.02±0.25	11.65±0.92	11.83±2.55
I-arm-lacZ-12	19C3 (X)	1	2	6	1.61±0.20	13.12±0.97	10.48±1.04
I-lacZ-12	47A13 (A)	2	2	3	0.90±0.09	18.47±1.77	16.55±2.13
I-lacZ-12	X[5]	1	2	3	2.32±0.13	24.87±3.47	28.86±4.71
P-lacZ	2 (A)	1	1	3	1.01±0.12	15.96±0.73	16.10±2.04
P-lacZ	50C17 (A)	1	1	3	1.20±0.20	11.96±2.46	14.15±0.49
P-lacZ	83A1 (A)	1	1	3	1.25±0.13	12.27±1.02	15.26±0.32
P-lacZ	65E8 (A)	2	2	3	1.53±0.09	5.65±0.91	8.70±1.77
P-lacZ	39A1 (A)	2	2	5	1.85±0.30	2.35±0.83	4.26±1.42
P-lacZ	9B4 (X)	1	2	7	1.74±0.33	11.61±1.42	9.76±0.88
P-lacZ	19E7 (X)	1	2	6	2.01±0.38	4.27±1.01	4.05±0.29
P-lacZ	12A2 (X)	1	2	7	2.42±0.59	9.62±2.78	10.41±1.42
P-lacZ*	75D4[6] (A)	1	1	3	1.35±0.35	8.86±0.93	11.78±1.76
P-lacZ*	3 (A)	1	1	3	1.82±0.55	0.90±0.40	1.50±0.28
P-lacZ*	9B12 (X)	1	2	3	1.69±0.06	16.73±0.80	14.14±0.34
P-lacZ*	4D6 (X)	1	2	3	1.41±0.07	6.02±1.63	4.27±1.08

Table 25. Mean male and female β-galactosidase activities and ratios in bromophenol blue-staged third instar larvae

[1] Number of independent experiments

[2] 100μOD min^{-1} μg $protein^{-1}$

[3] This line contains 2 insertions of the transgene.

[4] 95% confidence intervals are indicated.

[5] This line may contain >1 insertion of the transgene and is not included in Figure 24.

[6] This line may contain >1 insertion of the transgene.

Sample	Sex	Mass protein (μg)	Mean protein (μg)	Proportion protein in fat body compared to whole larvae
Whole larvae	Male	140.58	139.20	
		145.35		
		131.69		
	Female	183.87	180.74	
		185.19		
		173.17		
Fat body	Male	63.21	63.05	45.29
		67.65		
		58.27		
	female	48.72	46.58	25.77
		47.74		
		43.29		

Table 26. Fat body and whole larval protein levels in male and female third instar larvae

Fat bodies from single male and female third instar larvae aged on bromophenol blue food were dissected in triplicate. The mass of protein in these fat bodies was measured and compared to that of single whole male and female third instar larvae in triplicate. The proportion of protein in the fat body relative to the overall level of protein in the whole larvae is shown.

4.5.2 Mutation of a putative DSX binding site in the *Lsp1α* promoter does not reduce female promoter activity

The upregulation in activity of the *Lsp1α* promoter could be due to the sex-specific transcription factor Doublesex (DSX) that is known to regulate female-specific promoters (Coschigano and Wensink, 1993). A putative Doublesex (DSX) binding site exists in the promoter of *Lsp1α* (tgtACAAcgtggc). This site was identified by analysis of the 593 bp *Lsp1α* promoter using the Genomatix MatInspector, which identifies

putative transcription factor binding sites. DSX induces female-specific upregulation of the *yolk protein* genes (Coschigano and Wensink, 1993). In order to determine whether DSX induces upregulation of the *Lsp1α* promoter in females, the putative DSX binding site was mutated by PCR mutagenesis (section 4.1.1.2) and inserted upstream of the reporter gene, *lacZ* (section 4.1.3). Both X-linked and autosomal transgenic *P*-lacZ* fly lines exhibit female-specific upregulation equivalent to that of *P-lacZ, I-arm-lacZ-I2* and *I-lacZ-I2* lines (Table 25, Figure 24), indicating that this putative DSX site is not responsible for the sex-specific difference in the regulation of the *Lsp1α* promoter.

4.5.3 *Lsp1α* promoter activity is reduced by overexpression of MSL2

Sex-specific regulation of the *Lsp1α* promoter could be due to regulation by a member of the MSL complex. For example, MLE regulates *roX2* expression (Lee *et al.*, 2004) and MSL2 regulates expression of both of the *roX* genes (Amrein and Axel, 1997; Meller *et al.*, 1997; Rattner and Meller, 2004). In order to determine whether the *Lsp1α* promoter is capable of responding to the MSL complex, flies in which MSL2 is expressed under the control of the *hsp83* promoter were crossed to X-linked and autosomal *P-lacZ*, *I-arm-lacZ-I2* or *arm-lacZ* lines (section 2.5.2.5). The progeny of these crosses all have a single copy of the transgene and either carry *hsp83-MSL2* or a GFP marker. β-galactosidase activity of male and female GFP and *hsp83-MSL2* third instar larvae was examined in X-linked and autosomal *arm-lacZ*, *P-lacZ* and *I-arm-lacZ-I2* lines. As expected, the presence of MSL2 in male larvae does not affect expression of X-linked or autosomal *arm-lacZ*, *P-lacZ* or *I-arm-lacZ-I2* transgenes (Table 27). However, in females MSL2 increases X-linked *arm-lacZ* expression to *ca.* 1.7-fold (MSL2++ female/ WT-GFP female ratio, Table 27), demonstrating that the presence of MSL2 in females is sufficient to induce dosage compensation of an X-linked transgene. This is consistent with previously published data showing the expression of MSL2 in females is sufficient to induce formation of the MSL complex and its subsequent targeting to the X chromosome (section 1.5.3). Unexpectedly, MSL2 expression in females causes a *ca.* 2-fold decrease in the expression of both X-linked and autosomal *P-lacZ* and *I-arm-lacZ-I2* transgenes (MSL2++ female/ WT-GFP female ratio, Table 27). Based on the 95% confidence intervals, these MSL2++ female/GFP female ratios are significantly less than the expected ratio of one (Table 27).

The simplest explanation for this finding is that the *Lsp1α* promoter is regulated by the MSL complex, or by an individual component of the complex. The former possibility seems unlikely as chromatin immunoprecipitation data indicates that *Lsp1α* is not regulated by the MSL complex (section 6). Hence the most straightforward explanation is that *Lsp1α* is directly repressed by MSL2, independently from the MSL complex. An alternative possibility is that a component of the MSL complex, but not the entire complex itself, partially protects the *Lsp1α* promoter from position effects, and that this component is sequestered by the MSL complex in males. In females ectopically expressing MSL2, the MSL complex forms and also sequesters this component of the MSL complex.

		Mean ratio of activity ($100\mu OD\ min^{-1}\ \mu g\ protein^{-1}$)			
Construct	location	MSL2++[1] male/ WT (GFP) male	MSL2++ female/ WT (GFP) female	MSL2++ male/ MSL2++ female	WT (GFP) male/ WT (GFP) female
P -lacZ	9B4 (X)	0.89±0.09[2]	0.54±0.06	1.58±0.29	0.96±0.07
P -lacZ	12A2 (X)	1.20±0.21	0.44±0.13	1.93±0.79	0.67±0.15
P-lacZ	19E7 (X)	0.94±0.20	0.53±0.28	1.29±0.20	0.74±0.44
P-lacZ	65E8 (A)	1.36±0.09	0.38±0.13	1.45±0.40	0.39±0.09
P -lacZ	39A1 (A)	1.22±0.11	0.30±0.13	1.99±1.02	0.43±0.05
I-arm-lacZ-I2	23A3 (A)	1.12±0.10	0.57±0.12	1.24±0.16	0.62±0.06
arm-lacZ	79A4 (A)	1.00±0.10	1.20±0.15	1.19±0.14	1.43±0.17
arm-lacZ	10D8 (X)	1.02±0.09	1.66±0.11	1.18±0.07	1.93±0.22

Table 27. Mean β-galactosidase activity ratios in staged-third instar larvae overexpressing MSL2

All ratios are from flies heterozygous or hemizygous for the construct *i.e.* carrying one copy of the transgene.

[1] MSL2++ flies carry the *hsp83-MSL2* transgene.

[2] 95% confidence intervals are indicated for three independent experiments.

4.5.4 The *Lsp1α* promoter is upregulated by overexpression of MSL1, but not by other components of the MSL complex

Four members of the MSL complex, MSL1, MSL3, MLE and MOF, were tested as potential regulators of the *Lsp1α* promoter. Flies in which these individual MSL proteins are expressed at high levels were crossed into two autosomal *P-lacZ* lines subject to strong position effects, which exhibited corresponding high female to male activity ratios (sections 2.5.2.6, 2.5.2.7, 2.5.2.8, 2.5.2.9). If an individual MSL protein is responsible for protecting the *Lsp1α* promoter from position effects, the promoter should be more active in males that express the MSL protein, if the levels of this MSL are limiting due to it being sequestered into the MSL complex. Activity in females expressing the MSL protein could also increase, as although females have higher β-galactosidase than males in these lines, the activity is still much less than that of other transgenic lines. All the progeny of these crosses carry one copy of the autosomal transgene and either the MSL transgene under the control of the *hsp70* promoter, or a GFP balancer chromosome. In general, β-galactosidase activity levels in the *P-lacZ* larvae overexpressing some of the MSL proteins were quite variable (Table 28, 95% confidence intervals in MSL1 and MLE *P-lacZ* larvae). This variability was not observed with the control *arm-lacZ* transgene, and is consistent with the relief of position effects on the *P-lacZ* transgene in some, but not all, cells overexpressing the MSL protein, as in variegated or mosaic phenotypes. Despite this variability, some general trends can be observed. Overexpression of MSL1, but not MSL3, MLE or MOF, increases *Lsp1α* promoter activity in both male and female *P-lacZ* larvae (Table 28). This increase is statistically significant ($p<0.025$) in male *P-lacZ:65E8* larvae, but is not sufficient to equalize the activity to levels of WT (GFP) females. A significant β-galactosidase activity increase was not observed in male *P-lacZ:39A1* larvae expressing *hsp70-MSL1* compared to WT (GFP) males. However, female *P-lacZ:39A1* larvae expressing *hsp70-MSL1* did show a significant increase in β-galactosidase activity compared to WT (GFP) females. An increase in the activity of an autosomal *arm-lacZ* transgene in female larvae overexpressing MSL1 or MLE is also observed. While statistically significant, this difference is unlikely to be of biological importance as the difference between 1.44 ± 0.09 and 1.32 ± 0.02, or 1.24 ± 0.02 and 1.19 ± 0.03 is very low. These differences are statistically significant only due to the low errors in these experiments. It is possible that the *P-lacZ:39A1* line, which has a lower level of male

activity than the *P-lacZ:65E8* line, may be subject to stronger position effects. However, MSL1 expression was not sufficient to increase male expression to that of WT females in the *P-lacZ:65E8* line, and it is probable that another component is required for the partial protection of the promoter from position effects. Furthermore, an X-linked *P-lacZ* transgene does not exhibit significantly lower activity in *msl1* mutant females compared to non-mutant females (ratio of $msl1^{L60}/msl1^{L60}$ female to $msl1^{L60}$ */GFP* female activity is 0.93±0.26 in the *P-lacZ:9B4* line). This indicates that MSL1 alone is not sufficient for any protection of the *Lsp1α* promoter from position effects.

construct	location	Mean activity (100μOD min^{-1} μg $protein^{-1}$)			
		MSL1++ male	MSL1++ female	GFP male	GFP female
P-lacZ	65E8 (A)	3.19±0.82[1]	6.11±0.61	1.97±0.07	6.13±0.50
P-lacZ	39A1 (A)	2.28±0.44	5.33±0.88	1.66±0.28	2.56±0.23
arm-lacZ	79A4 (A)	1.92±0.11	1.44±0.09	1.77±0.04	1.32±0.02
		MSL3++ male	MSL3++ female	GFP male	GFP female
P-lacZ	65E8 (A)	3.09±0.33	5.16±0.85	2.95±0.28	5.77±1.18
P-lacZ	39A1 (A)	1.50±0.25	2.17±0.04	1.25±0.19	2.54±0.68
arm-lacZ	79A4 (A)	1.60±0.07	0.95±0.04	1.50±0.15	1.03±0.01
		MLE++ male	MLE++ female	GFP male	GFP female
P-lacZ	65E8 (A)	1.74±0.48	4.85±0.45	1.71±0.20	5.38±0.91
P-lacZ	39A1 (A)	1.13±0.64	2.30±1.01	0.76±0.43	1.82±0.74
arm-lacZ	79A4 (A)	1.71±0.03	1.24±0.02	1.64±0.07	1.19±0.03
		MOF++ male	MOF++ female	GFP male	GFP female
P-lacZ	65E8 (A)	2.32±0.35	6.28±0.17	2.65±0.77	5.96±0.23
P-lacZ	39A1 (A)	0.85±0.31	2.17±0.40	0.76±0.23	2.81±0.98
arm-lacZ	79A4 (A)	1.81±0.06	1.14±0.03	1.75±0.07	1.16±0.04

Table 28. Mean β-galactosidase activities in autosomal *arm-lacZ* and *P-lacZ* staged-third instar larvae overexpressing MSL1, MSL3, MLE or MOF

All activities are from male (M) and female (F) flies heterozygous for the construct, *i.e.* carrying 1 copy of the *lacZ* transgene. The mean activity in larvae expressing an MSL was significantly higher than the mean activity of GFP larvae by T-test analysis [t(16)=X; p<0.025] in the underlined samples.

[1] 95% confidence intervals are indicated.

4.6 Summary of the promoter activity results

The *Lsp1α* promoter exhibits increased activity in female larvae in comparison to male larvae. The level of increased female promoter activity is dependent on the expression level of the transgene. Transgenes with a high level of expression show little difference in promoter activity between male and female larvae. Contrastingly, transgenes that are expressed at lower levels have significantly higher levels of activity in female larvae relative to male larvae. Significantly, the activity of the *Lsp1α* promoter is decreased *ca.* two-fold in the presence of MSL2. The activity of the promoter is slightly increased in the presence of MSL1. The significance of these findings is discussed further in chapter 8.

5 ATTEMPT TO INTRODUCE REPORTER GENES INTO THE ORF OF *LSP1α IN VIVO*

Attempts were made during the course of this study to introduce various reporter constructs into the *Lsp1α* gene in its natural position on the X chromosome using homologous recombination. This would have enabled the level of *Lsp1α* expression relative to the *Lsp1α-lacZ* transgenes to be determined. It would also have enabled direct targeting of the MSL complex via chromatin entry sites into *Lsp1α* in order to determine whether this induced dosage compensation and resulted in equal levels of male and female expression. The constructs required to target a reporter construct into *Lsp1α* consist of a 9.4 kb *Lsp1α* genomic region inserted within a homologous recombination targeting vector (Figure 25). This vector consists of site-specific endonuclease and recombinase recognition sites that flank the genomic sequence and reporter construct. Expression of the endonuclease and recombinase in flies carrying the targeting construct should result in its excision and recombination within the *Lsp1α* region on the X chromosome. Marker genes were inserted in the open reading frame (ORF) of *Lsp1α* within the targeting construct to identify potential recombinants. Two marker genes were tested with an eye-specific or constitutive promoter controlling a gene encoding a fluorescent protein. In addition the construct containing the eye-specific marker contained a *lacZ* reporter gene controlled by a constitutive promoter to quantify gene expression. Unfortunately attempts with both of these constructs were unsuccessful. The potential reasons for this lack of success are discussed in the following sections.

5.1 Homologous recombination construct for insertion of *arm-lacZ* into the *Lsp1α* ORF

5.1.1 Genomic region between *CG2560* and *CG2556* into pGEM®-T Easy

The genomic region between the middle of the ORF of *CG2560* and the middle of the 5' UTR of *CG2556* was amplified by long range PCR with the CG2560ORFF and CG25565'UTRR primers (section 10.1) from *Drosophila y w* genomic DNA (method 3: annealing temperature of 55°C: 35X). The 9413 bp product obtained was gel purified, ligated directly into the pGEM®-T Easy vector (Promega), and transformed into *DH5α*

E. coli cells. The identity of this clone was confirmed by restriction endonuclease digestion and sequencing. This fragment is referred to as the 9.4 kb *Lsp1α* genomic region.

5.1.2 Insertion of *GMR-3xP3-DsRed-hsp70polyA* into pCaSpeR-arm-βgal *StuI/NotI*

The 1.61 kb *GMR-3xP3-DsRed-hsp70polyA* reporter construct was excised from pB/*SCS*-GMR-3xP3-DsRed-hsp70polyA-*SCS*' (provided by Asela Attapatu, Massey University) with *Not*I and *Sal*I, and blunt-ended using Klenow polymerase followed by gel purification. This fragment was ligated into pCaSpeR-arm-βgal *Stu*I/*Not*I (section 4.1.5) that had been digested with *Pst*I and blunt ended using T4 DNA polymerase, and transformed into *DH5α E. coli* cells. Clones in which *GMR-3xP3-DsRed-hsp70polyA* was in the same orientation as *arm-lacZ* were selected by restriction endonuclease digestion and confirmed using sequencing with the armlacZNotIseq sequencing primer (section 10.3). The *GMR-3xP3-DsRed* marker gene is strongly expressed in the eye and is detected using fluorescence microscopy with the SZX-FRFP1 red fluorescence filter[1] set attached to the SZX-RFA3 fluorescence illuminator on the SZX12 stereomicroscope (Olympus).

5.1.3 Replacement of GH02424 cDNA with *Pst*I – *Not*I linker in pOT2/GH02424

The chloroamphenicol-resistant plasmid pOT2 containing the GH02424 cDNA was digested with *Pst*I to remove the cDNA insert, CIP treated, and gel purified. The polynucleotide kinase-treated *Pst*I – *Not*I linker (section 10.4) was ligated into the 1.607 kb pOT2 vector, and transformed into *DH5α E. coli* cells. Clones in which the linker had inserted correctly were identified by restriction endonuclease digestion. The resulting plasmid was named pOT2/*Not*I.

5.1.4 Insertion of the 9.4 kb *Lsp1α* genomic region into pOT2

In order to utilize the *Fsp*I site in the *Lsp1α* open reading frame, it was necessary to transfer the *Lsp1α* genomic region into a plasmid that lacked this restriction endonuclease site. pOT2/*Not*I does not contain a *Fsp*I restriction endonuclease site. The 9.4 kb *Lsp1α* genomic region was excised from the pGEM®-T Easy vector (Promega)

[1] Excitation filter: 520-550nm. Barrier filter: 580nm.

using *Not*I and gel purified. This fragment was ligated into pOT2/*Not*I that had been digested with *Not*I, CIP treated and gel purified. This was transformed into *DH5α E. coli* cells. The identity of this clone was confirmed by restriction endonuclease digestion.

5.1.5 Insertion of the *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* reporter genes into the *Lsp1α* open reading frame

The *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* dual reporter cassette was inserted into the single *Fsp*I restriction endonuclease site within the open reading frame of *Lsp1α* within the 9.4 kb genomic fragment in pOT2. The *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* construct was excised from pCaSpeR-arm-βgal *Stu*I/*Not*I using *Not*I and *Stu*I, blunt ended using Klenow polymerase and gel purified. This was ligated into pOT2/*Not*I carrying the 9.4 kb *Lsp1α* genomic region that had been digested with the blunt-end cutter *Fsp*I, CIP treated and gel purified. The ligation reaction was transformed into *DH5α E. coli* cells. Clones carrying the reporter genes within the *Lsp1α* open reading frame in the same orientation as *Lsp1α* were identified using restriction endonuclease digestion. Clones in which the pCaSpeR fragment had inserted rather than the *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* construct were identified by growth on LB media containing ampicillin, and discarded.

5.1.6 Insertion of the 9.4 kb *Lsp1α* genomic region carrying the *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* reporter genes into the pW30 homologous recombination vector

The pW30 plasmid (Gong and Golic, 2003) consists of FRT site-specific recombinase sites in the same orientation flanking sites for the I-Sce site-specific endonuclease. These FRT and I-SceI sites surround a multiple cloning site that includes a *Not*I restriction endonuclease site (section 10.5.13). A *Not*I restriction endonuclease site was regenerated at the 3' end of the *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* reporter gene cassette after insertion into the *Lsp1α* open reading frame, hence partial digestion was used to separate the genomic fragment/reporter cassette from pOT2. Partial digestion with *Not*I followed by gel purification was used to separate the 17.1 kb cassette containing the 9.4 kb *Lsp1α* genomic region with the *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* reporter genes from the smaller restriction endonuclease digestion products

including the 1.6 kb pOT2 vector. This 17.1 kb *Not*I fragment was ligated into pW30 that had been digested with *Not*I, CIP treated and gel purified. The ligation reaction was transformed into *DH5α E. coli* cells. Clones carrying the 17.1 kb fragment in pW30 were identified by restriction endonuclease digestion. Clones in which the entire 18.6 kb plasmid had inserted into pW30 were identified by growth on LB-media containing chloroamphenicol, and discarded. The position of the *arm-lacZ/GMR-3xP3-DsRed-hsp70polyA* reporter gene cassette within *Lsp1α* was confirmed by sequencing with the armlacZStuIRevseq, GMR-DsRedFseq and GMR-DsRedFseq2 sequencing primers (section 10.3). As expected, the reporter gene cassette had inserted exactly within the *Fsp*I site of *Lsp1α*. This plasmid is referred to as pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed (Figure 25).

5.2 Homologous recombination construct for insertion of *arm-EGFP* into the ORF of *Lsp1α*

5.2.1 Replacement of *lacZ-SV40* with the *Asp718 – Pst*I linker

The *lacZ-SV40* reporter gene of pCaSpeR-arm-βgal (Vincent *et al.*, 1994) was excised with *Asp*718 and *Pst*I, and the resulting fragment gel purified. The *Asp*718-*Pst*I linker (section 10.4) containing the *Bam*HI, *Bgl*II, *Nhe*I and *Not*I restriction endonuclease sites between the *Asp*718 and *Pst*I cohesive ends was ligated into the cut vector. This ligation reaction was transformed into *DH5α E. coli* cells, and clones carrying pCaSpeR-arm-βgal with the *Asp*718-*Pst*I linker in place of the missing *lacZ-SV40* reporter gene were identified by restriction endonuclease digestion. This plasmid was named pCaSpeR-arm.

5.2.2 Replacement of *Pub* promoter in pB[PUbnlsEGFP] with *armadillo* promoter

The *polyubiquitin* promoter of pB[PUbnlsEGFP] (Handler and Harrell, 2001) (section 10.5.11) was excised with *Eco*RI and *Bam*HI, and the resulting 6117 bp fragment retaining the EGFP ORF gel purified. The 1791 bp *armadillo* promoter from pCasper-arm was excised using *Eco*RI and *Bam*HI and gel purified. The 1791 bp *armadillo* *Eco*RI/*Bam*HI fragment was directionally ligated into pB[PUbnlsEGFP] missing the *Pub* promoter, and transformed into *DH5α E. coli* cells. Clones in which *armadillo* was in the same orientation as *EGFP* were selected by restriction endonuclease digestion

and confirmed using sequencing with the armlacZNotIseq sequencing primer (section 10.3).

5.2.3 Insertion of the *arm-EGFP* reporter gene into the *Lsp1α* open reading frame

The 3.8 kb *arm-EGFP* reporter cassette was released from pB[nlsEGFP] by digestion with *Cla*I and *Eco*RI, blunt ended with T4 DNA polymerase and gel purified. This *arm-EGFP* cassette was ligated into the unique *Fsp*I site of the 9.4 kb genomic fragment in pOT2 (section 5.1.4), and transformed into *DH5α E. coli* cells. Clones carrying *arm-EGFP* within the *Lsp1α* open reading frame were identified using restriction endonuclease digestion. Clones in which the pCaSpeR fragment had inserted rather than the *arm-EGFP* construct were identified by growth on LB media containing ampicillin, and discarded.

5.2.4 Insertion of the 9.4 kb *Lsp1α* genomic region carrying the *arm-EGFP* reporter gene into the pW30 homologous recombination vector

The pW30 plasmid (Gong and Golic, 2003) consists of FRT site-specific recombinase sites, in the same orientation, flanking sites for the I-Sce site-specific endonuclease. These FRT and I-SceI sites surround a multiple cloning site that includes a *Not*I restriction endonuclease site (section 10.5.13). A *Not*I restriction endonuclease site was present within the *arm-EGFP* reporter cassette, hence partial digestion was used to separate the genomic fragment/reporter gene from pOT2. The 13.2 kb *Not*I cassette containing the 9.4 kb *Lsp1α* genomic region with the *arm-EGFP* reporter gene was excised by partial digestion with *Not*I, and separated from the smaller restriction endonuclease digestion products including the 1.6 kb pOT2 vector by gel purification. This 13.2 kb *Not*I fragment was ligated into pW30 that had been digested with *Not*I, CIP treated and gel purified. The ligation reaction was transformed into STB2™ *E. coli* cells (10268-019; GibcoBRL/Life Technologies). This cloning step was carried out using STB2™ *E. coli* cells (10268-019; GibcoBRL/Life Technologies) to prevent plasmid rearrangement due to the presence of repetitive DNA elements such as the FRT sites within the plasmid. Clones carrying the 13.2 kb fragment in pW30 were identified by restriction endonuclease digestion. Clones in which the entire 14.8 kb plasmid had inserted into pW30 were identified by growth on LB-media containing chloroamphenicol, and discarded. The position of the *arm-EGFP* reporter gene cassette

within *Lsp1α* was confirmed by sequencing with the armlacZStuIRevseq, Lsp1a1020F and Lsp1a1468R sequencing primers (section 10.3). As expected, the reporter gene cassette had inserted exactly within the *Fsp*I site of *Lsp1α*. Sequencing with the pw30-545F and pw30-1276R sequencing primers (section 10.3) was used to confirm that these FRT and I-SceI sites were intact, and the pW30 sequence was identical to that received from Kent Golic. This plasmid is referred to as pW30-Lsp1α-armEGFP (Figure 25). The *arm-EGFP* marker gene is expressed in all developmental stages but is most easily detected during the third instar larval stage using fluorescence microscopy with the SZX-FGFP green fluorescence filter[1] set attached to the SZX-RFA3 fluorescence illuminator on the SZX12 stereomicroscope (Olympus).

[1] Excitation filter: 460-490nm. Barrier filter: 510nm.

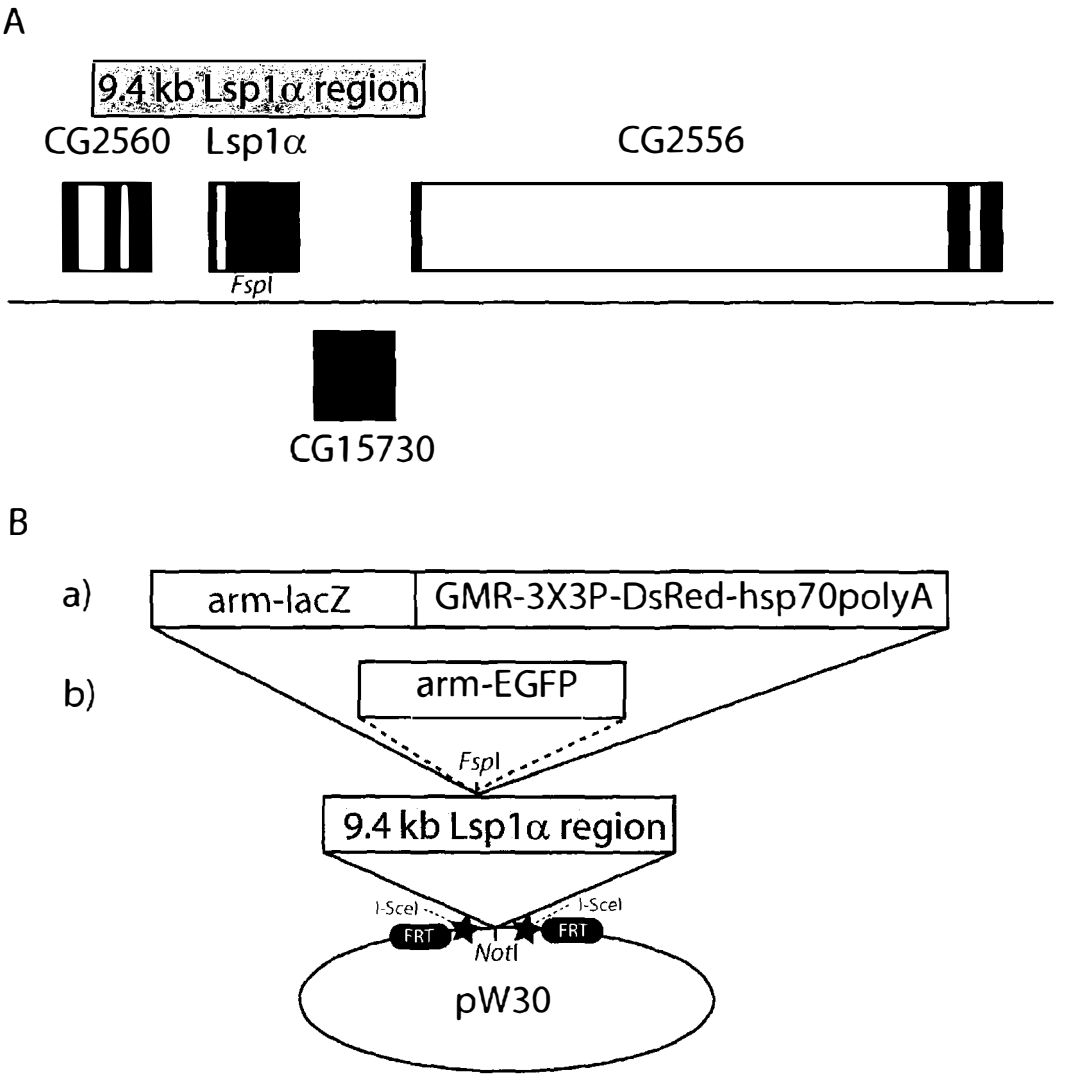


Figure 25. Schematic of homologous recombination constructs

(A) The 9.4 kb *Lsp1α* region begins within the ORF of *CG2560* and extends to the 5' UTR of *CG2556*. Exons (black) and introns (white) are indicated for each gene in the region. (B) Schematic of the *pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed* (a) and *pW30-Lsp1α-armEGFP* (b) constructs. These plasmids contain the *arm-lacZ/GMR-3XP3-DsRed* (a) or *arm-EGFP* (b) reporter genes inserted in the ORF of *Lsp1α* within the 9.4 kb *Lsp1α* region inside the pW30 plasmid, flanked by FRT and I-SceI sites.

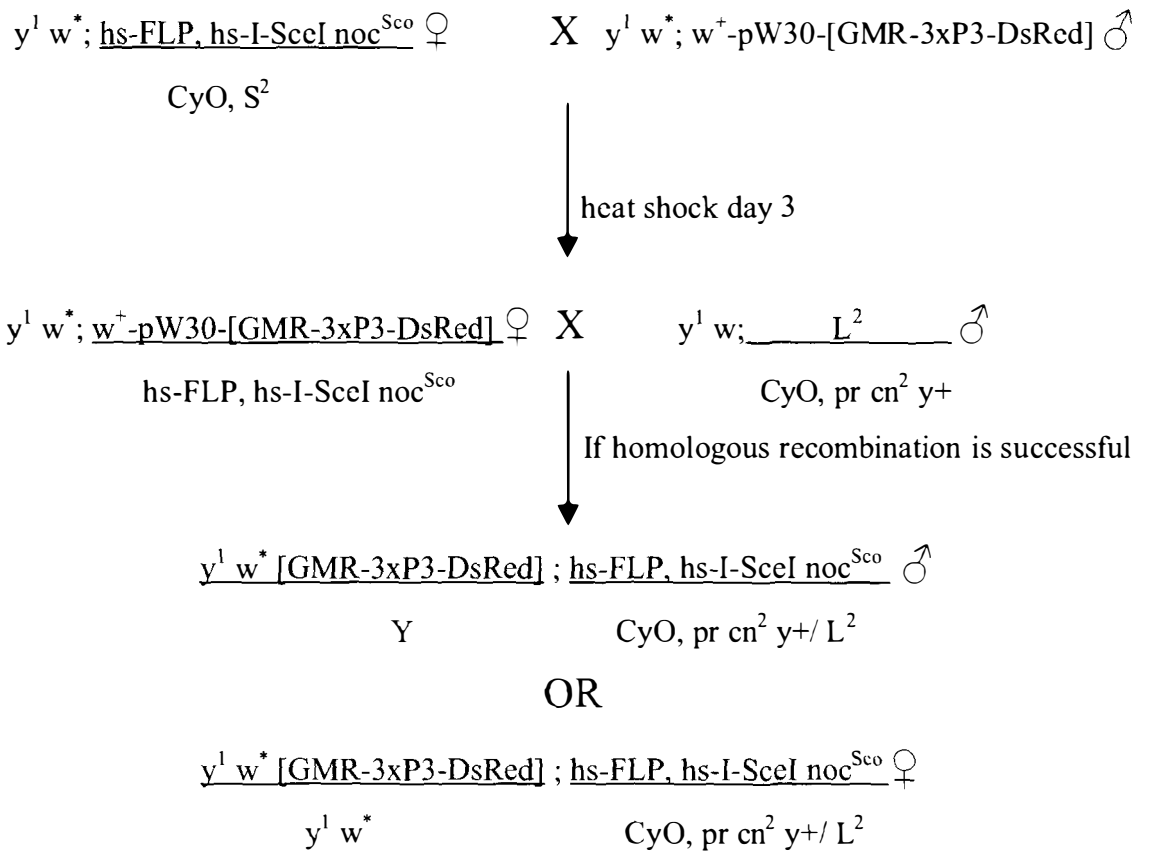


Figure 26. Homologous recombination crossing scheme

All female flies shown in the crossing scheme are virgins. The crossing scheme is identical to that used for the *pW30-Lsp1α-armEGFP* transformant lines. Virgin females expressing the FLP recombinase and I-SceI endonuclease under the control of heat shock (hs) promoters are crossed to males homozygous for the targeting construct. The progeny of this cross are heat shocked on day 3 to induce expression of the FLP recombinase and I-SceI endonuclease. Homologous recombination is much more efficient in the female germline compared to the male germline, thus females are used for the second generation of the cross. If homologous recombination occurs, the *GMR-3xP3-DsRed* cassette within the FRT and I-SceI sites will be excised from the second chromosome and integrated within the *Lsp1α* ORF on the X chromosome. Thus the *DsRed* and *mini-white* eye colour markers will segregate, and any *w- DsRed+* offspring will be potential candidates for successful homologous recombination.

5.3 The *DsRed* marker is not visible in *pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed* transgenic fly lines

Two separate crosses between flies with wild-type eye colour (red) and flies carrying the *GMR-3XP3-DsRed* transgene prior to the initiation of this phase of the project indicated that while the wild-type red eye colour masks *DsRed* expression in the eye itself, *DsRed* expression is observed in the oscilli of flies carrying the transgenic marker. Hence *DsRed* was considered a suitable marker in conjunction with *mini-white* for homologous recombination.

A single line was obtained carrying the *pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed* transgene on the second chromosome. This line was identified by expression of the *mini-white* marker gene. However, *DsRed* expression in the oscilli of the transgenic fly line generated carrying the *pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed* construct was not observed. Larval *DsRed* expression was not observed in this line, or in control lines expressing *GMR-3XP3-DsRed* transgenes. It is possible that the *DsRed* expression in this line was weaker than in the control lines, and was masked by the *mini-white* marker.

5.4 Homologous recombination was not successful with *pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed* transgenic fly lines

Homologous recombination of the *pW30-Lsp1α-armlacZ/GMR-3xP3-DsRed* construct was attempted to insert the *arm-lacZ/GMR-3XP3-DsRed* reporter genes into the ORF of *Lsp1α*. The crossing scheme is described in section 2.5.2.3 and shown in Figure 26. In brief, induction of the FLP-recombinase should lead to the excision of the *arm-lacZ/GMR-3XP3-DsRed* cassette, which is cleaved at the flanking I-SceI sites by the expression of the I-SceI endonuclease. Successful homologous recombination would be indicated by segregation of the *DsRed* and *mini-white* colour markers. However, no progeny expressing *DsRed* were observed out of the 419 examined. It is unclear whether this was due to the lack of homologous recombination or whether the *DsRed* reporter gene was unable to be expressed.

5.5 Homologous recombination was not successful with *pW30-Lsp1α-armEGFP* transgenic fly lines

Due to the failure of the homologous recombination attempt using *DsRed* as a marker, an attempt was made using the *EGFP* marker to determine if it was possible to use this method for the purposes of this study. Two transgenic lines carrying the *pW30-Lsp1α-armEGFP* construct were identified by the presence of the *mini-white* marker, on the second and third chromosomes. Homologous recombination of the *pW30-Lsp1α-armEGFP* construct was attempted to insert the *arm-EGFP* reporter gene into the ORF of *Lsp1α* (section 2.5.2.3). Successful homologous recombination would be indicated by segregation of the *GFP* and eye colour markers. However, all of the *ca.* 200 progeny examined expressing the *arm-EGFP* marker (*ca.* 50% of the total progeny) also had red eyes. This is inconsistent with the results described in (Gong and Golic, 2003), which state that the FLP site-specific recombinase excises the reporter gene flanked by FRT sites in 99.6% of the progeny.

5.6 The *pW30-Lsp1α-armEGFP* construct is not excised by the FLP site-specific recombinase

To determine if the *pW30-Lsp1α-armEGFP* construct could be excised by the FLP site-specific recombinase, as predicted by (Gong and Golic, 2003), a transgenic fly line constitutively expressing FLP was crossed to two separate transgenic fly lines carrying the *pW30-Lsp1α-armEGFP* construct (section 2.5.2.4). If the FLP recombinase excised the intervening DNA between the FRT sites efficiently, then all of the red-eyed progeny should no longer express GFP. However, examination of the larval and adult progeny of these crosses showed that all red-eyed flies still expressed *arm-EGFP*. To prove that the FLP site-specific recombinase could excise DNA between FRT sites efficiently, the constitutive FLP line was crossed to flies carrying the *mini-white* eye colour marker flanked by FRT sites (section 2.5.2.4). As expected, all of the *ca.* 400 progeny of this cross except one had white eyes. Therefore FLP is able to excise the DNA from between FRT sites efficiently. Sequencing of the *pW30-Lsp1α-armEGFP* plasmid with primers designed within pW30 showed that the FRT and I-SceI sites of this plasmid are intact and match the pW30 sequence obtained from Kent Golic (section 5.2.4). Therefore, no explanation is available for the failure of the FLP recombinase to excise the reporter gene from this construct. However, it is obvious that this excision problem

is responsible for the failure of the homologous recombination scheme. It was decided not to continue further with this homologous recombination approach because of the limited time available. However, with the benefit of hindsight, a more useful approach to this scheme would involve using a *piggyBac* vector containing the *polyubiquitin-EGFP* marker, with FRT and I-SceI sites flanking *Lsp1α* genomic DNA in which *the GMR-3XP3-DsRed* marker was inserted. Thus the main source of *Drosophila* DNA would be the *Lsp1α* genomic fragment, whereas pW30 contains the *Drosophila* X-linked *mini-white* and *P*-elements, which could result in spurious targeting, although this is unlikely to prove a problem. "Ends-in" rather than "Ends-out" targeting might also be a more successful approach to homologous recombination, as the "Ends-in" approach has been used successfully by other labs (Robert Saint, Australian National University, Canberra, Australia, personal communications, 2005). Furthermore, it would be more interesting to introduce the 18D10 CES 3' of *Lsp1α* to determine if this could recruit MSL complex. Insertion of the *lacZ* reporter into *Lsp1α* in frame would also be of interest to determine the promoter activity ratio and level in its natural position and compare this with the promoter-reporter constructs inserted elsewhere in the genome.

6 CHROMATIN IMMUNOPRECIPITATION ANALYSIS

There are two potential explanations for the increased expression of the *Lsp1α* gene in female larvae relative to male larvae. Either *Lsp1α* is sex-specifically regulated, or it escapes regulation by the MSL complex. Due to the increased *Lsp1α* promoter activity observed in many of the *Lsp1α-lacZ* transgenes (section 4.5), it is possible that the *Lsp1α* promoter is sex-specifically regulated. In order to determine whether the *Lsp1α* gene is regulated by the MSL complex at its natural position on the X chromosome, *Lsp1α* was examined for the presence of the histone modification, H4K16ac, that is associated with the MSL complex and is required for dosage compensation (section 1.5.5).

Chromatin immunoprecipitation analysis (ChIP) using the anti-H4K16ac antibody (Serotec) was conducted as described in (section 2.6.6) on hand-dissected fat bodies from male third instar larvae of genotypes *P-lacZ:9B4*, *P-lacZ:19E7* and *y w*. Male larvae were selected to increase the efficiency of the chromatin immunoprecipitation, as only the male X chromosome is enriched for the H4K16ac modification. Fat bodies from third instar larvae were also selected to increase the immunoprecipitation efficiency for *Lsp1α*, as the MSL complex is targeted to actively expressed genes and *Lsp1α* is exclusively expressed in fat body tissue (Davies *et al.*, 1986).

Several modifications were made to the chromatin immunoprecipitation protocol before efficient immunoprecipitation was obtained. Key among these was the use of male fat body tissue rather than whole third instar larvae, the addition of the deacetylase-inhibitor sodium butyrate to all buffers and shorter immunoprecipitation times to reduce the effect of proteinases. The improvement to the method of washing as described in (section 2.6.6) and addition of pre-binding/pre-clearing steps may also have improved the fold efficiencies obtained. Most of these modifications were made after advice received from Edwin Smith (Rockefeller University, 2005, personal communications). Nuclei were present following the fat body nuclei extraction procedure as determined by cytometer counting and DAPI staining of aliquots (Figure 27).

Efficient chromatin immunoprecipitation requires that the DNA fragments be sheared to *ca.* 500 bp. It was determined that 6 x 30 s of power level 2.1 on the VirSonic sonicator were required to shear DNA to this size (Figure 28).

Although *Pgd* is expressed in the fat body of third instar larvae, it is also expressed in a subset of other cells within the larva (Williamson *et al.*, 1980). Hence a fat-body specific, X-linked gene was required as a positive control for the level of H4K16ac enrichment expected on such a gene. *Lsd-2* is an X-linked gene expressed only in the fat body of third instar larva[1] and was used as a positive control alongside *Pgd* in chromatin immunoprecipitation analysis of *Lsp1α* (Miura *et al.*, 2002; Teixeira *et al.*, 2003). H4K16ac enrichment is only observed for X-linked genes, so the autosomal gene *glycerol-3-phosphate dehydrogenase* (*Gpdh*) was used as a negative control as described by Smith, Allis *et al.* (2001). All primer sets were designed to the 3' end of the gene to be analysed, as H4K16ac enrichment is greatest at the 3' end of the X-linked *Pgd* gene (Smith *et al.*, 2001).

Real-time PCR was conducted on the undiluted immunoprecipitated DNA and 100-fold diluted input DNA as described in (section 2.6.6). Primers were designed to amplify 150 - 200 bp segments (*i.e.* the average length of DNA wrapped around a single nucleosome, which is *ca.* 147 bp). Most primers were designed to the 3' regions of each of the genes to be analysed using the LightCycler Probe Design Software 2.0 (Roche). Two primer sets were analysed for the X-linked, fat body-expressed gene *Pgd*: pgdf10 and pgdr10 (pgd10; section 10.1); pgdf5 and pgdr5 (pgd5; section 10.1). The pgd5 set amplifies a fragment within the cds of *Pgd*, and the pgd10 set amplifies a sequence near the 3' end of the gene. One primer set was analysed for the autosomal, constitutively expressed gene *Gpdh*: GpdhF and GpdhR (section 10.1). The *Pgd* and *Gpdh* primer sequences were identical to those used previously by Edwin Smith and colleagues (Smith *et al.*, 2001). One primer set was analysed for the X-linked, fat body-specific gene *Lsd-2*: Lsd2-1813F and Lsd2-1921R (section 10.1). These primers amplify a product within the 3' UTR of *Lsd-2*. Two primer sets were analysed for *Lsp1α*: Lsp1a1884F and Lsp1a2085R (Lsp1a-1; section 10.1); Lsp1a-2439F and Lsp1a-2559R (Lsp1a-2; section 10.1). The Lsp1a-1 primer set amplifies a product towards the 3' end

[1] It is also expressed in the germ line of females.

of the *Lsp1α* ORF, while the Lsp1a-2 set amplifies a product that spans the end of the *Lsp1α* ORF and the beginning of the 3' UTR. One primer set was analysed for *lacZ* in the *P-lacZ:9B4* and *P-lacZ:19E7* genotypes only: lacZ-2823F and lacZ-2950R (section 10.1). This primer set amplifies a product within the 3' end of the *lacZ* ORF. Two primer sets were analysed for the gene 3' of *Lsp1α*, *CG2556*: CG2556-5'UTRF and CG2556-5'UTRR (CG2556-5'UTR; section 10.1); CG2556-3'UTRF and CG2556-3'UTRR (CG2556-3'UTR; section 10.1). The CG2556-5'UTR primer set amplify a product within the 5' UTR of CG2556, while the CG2556-3'UTR primer set amplify a product within the 3' UTR of CG2556. The correct identity of the product amplified by each primer set was confirmed by sequencing.

The crossing point was automatically determined by the LightCycler software for each PCR reaction in triplicate. The mean crossing point was used in all further calculations, as the standard deviation was <0.2 cycles. Fold enrichment was determined by:
2^(crossing point input test gene – crossing point ChIP test gene)/
2^(crossing point input *Gpdh* – crossing point ChIP *Gpdh*). The crossing points for each PCR reaction are given in appendix 10.9.2 and a sample trace for one primer set is shown in appendix 10.9.1. The PCR efficiencies were calculated using LinRegPCR 7.0 software (Ramakers *et al.*, 2003). These were generally greater than 1.80, indicating that the reaction efficiency for the different chromatin preparations was the same (Table 30). Importantly, there was little difference in reaction efficiencies between the ChIP template and its corresponding input template. Thus it is valid to calculate the fold enrichments for these experiments using the calculation that is described above.

Chromatin immunoprecipitation using the anti-H4K16ac antibody was successful as the positive X-linked control genes *Pgd* and *Lsd-2* showed mean enrichments of between 3 and 10-fold in the 3 different genotypes analysed (Figure 29, Table 29). The mean fold enrichment for each gene was highly consistent between the 3 different genotypes analysed. The pgd10 primer set showed higher enrichments of *ca.* 10-fold compared to the other primer sets including pgd5 and Lsd-2. This is consistent with the results obtained in (Smith *et al.*, 2001). The fold enrichments obtained using fat bodies were *ca.* 5 fold lower than that obtained using cell culture by (Smith *et al.*, 2001). However, tissue samples are not a homogeneous population of cells, which may decrease the level of H4K16ac within a particular gene. Furthermore fat body tissue is very rich in

enzymes such as proteases and deacetylases, and this may also decrease the efficiency of the immunoprecipitation. Despite this, the reproducible nature of the chromatin immunoprecipitation data using fat body tissue indicates that the fold enrichments obtained are valid for this tissue.

In the *P-lacZ:9B4* and *P-lacZ:19E7* genotypes, *lacZ* has a mean fold enrichment similar to pgd5 and *Lsd-2*. Thus *lacZ* under the control of the *Lsp1α* promoter in X-linked transgenic fly lines is enriched for H4K16ac to the same extent as other X-linked genes. However, in all three genotypes, and using both sets of primers, *Lsp1α* has mean fold enrichments close to that of *Gpdh*, indicating that the H4K16ac modification is not enriched on this gene. Thus although the 95% confidence intervals for each fold enrichment represent two experiments, in essence the control genes and *Lsp1α* were analysed six times. Each were analysed twice in each of the three genotypes tested.

The MSL complex is targeted to actively expressed genes (Sass *et al.*, 2003), and the only gene within the immediate vicinity of *Lsp1α* that is expressed in the same tissue and stage as *Lsp1α*, is *CG2556*. In order to determine whether the H4K16ac modification, and therefore the MSL complex, was present on the end of this gene closest to *Lsp1α*, the mean fold enrichment of H4K16ac on the 5' and 3' UTR of *CG2556* was determined using the CG2556-5'UTR and CG2556-3'UTR primer sets. It was necessary to analyze each of these primer sets in immunoprecipitated chromatin from the two different *P-lacZ* lines due to a lack of remaining immunoprecipitated DNA. However, due to the similar fold enrichments obtained for all the other primer sets analysed in these two lines, the fold enrichments from the 5' and 3' UTR primer sets can be compared. The CG2556-3'UTR primer set shows a fold enrichment of 1.93 ± 0.77, which is lower than that observed for the *Pgd*, *Lsd-2* or *lacZ* primer sets. However, this does indicate that the 3' UTR of *CG2556* is moderately enriched for H4K16ac. The CG2556-5'UTR primer set show a fold enrichment of 1.10 ± 0.09, which is more similar to that observed for the autosomal gene, *Gpdh*. This indicates that the 5' UTR of *CG2556* is not enriched for H4K16ac. Despite the low levels of acetylation present of the 3' UTR of *CG2556*, this gene is clearly expressed at equivalent levels in male and female larvae (section 3.7) and is likely to be regulated by the MSL complex. It is possible that high levels of H4K16ac are not required for regulation by the MSL

complex, and that this moderate enrichment may be sufficient. It is also possible that the enrichment of H4K16ac on the 3' end of *CG2556* may be due to regulation of the neighbouring *CG11146* gene by the MSL complex. It has been demonstrated that the enrichment in acetylation on X-linked genes may extend beyond the transcription unit (Smith *et al.*, 2001), and the 3' end of the *CG11146* and *CG2556* genes are adjacent. This result differs from the previously published finding that while the H4K16ac modification is enriched at the 3' end of the X-linked *Pgd* gene, it is also enriched at the 5' and promoter regions of X-linked genes such as *Pgd* and *Zw* (Smith *et al.*, 2001). This supports the hypothesis that the MSL complex is absent from the chromosomal region immediately surrounding *Lsp1α* because there is no noticeable enrichment for the H4K16ac modification until about 11 kb downstream of *Lsp1α*, at the 3' end of the only gene expressed in the same tissue and at the same developmental stage as *Lsp1α, CG2556.*

Figure 27. DAPI stain of fat body nuclei preparation prior to sonication and immunoprecipitation.

The size difference of the nuclei shown (blue) reflect differences in the sizes of larval (large) and imaginal cell (small) nuclei and cells.

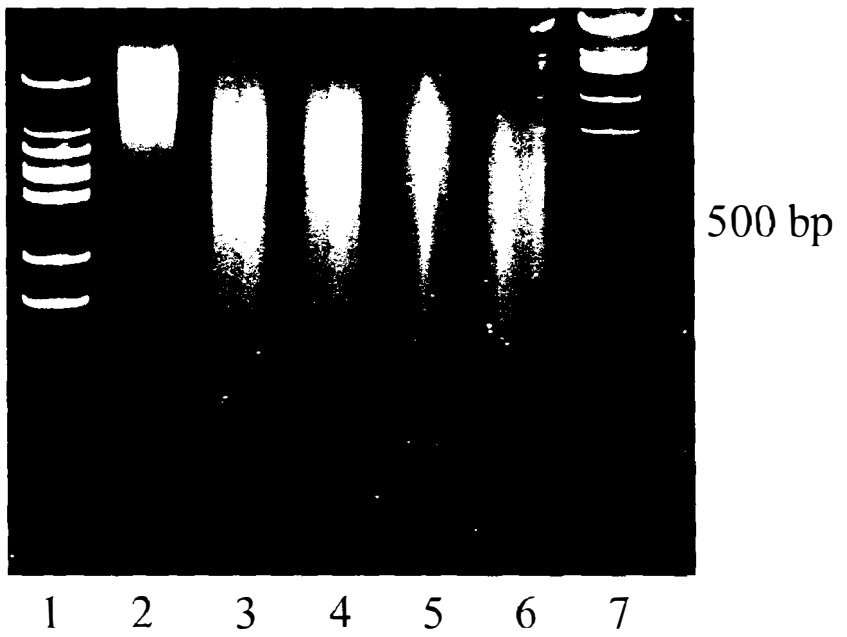


Figure 28. Sonication trial of DNA from nuclei prior to chromatin immunoprecipitation

Nuclei were sonicated at power level 2.1 on VirSonic for 3 x 15 s (lane 2), 6 x 15 s (lane 3), 9 x 15 s (lane 4), 12 x 15 s (lane 5) or 6 x 30 s (lane 6). One μg of DNA purified from these nuclei was electrophoresed on a 5% (w/v) non-denaturing polyacrylamide gel at 40V next to 100 bp molecular size markers (lane 1; section 2.6.2.1) and Lambda *Hind*III/*Sac*II molecular size markers (lane 7; section 2.6.2.1). Sonication at 6 x 30 s resulted in DNA fragments of the optimal size average of 500 bp.

	Pgd-5		*Pgd-10*		*Lsp1α-1*	
Fly Line	F.E.	Mean	F.E.	Mean	F.E.	Mean
P-lacZ:9B4	5.62		10.83		1.14	
	4.25	4.93±1.34	10.01	10.42±0.80	0.99	1.06±0.14
P-lacZ:19E7	3.68		7.36		0.76	
	3.09	3.38±0.58	6.81	7.08±0.55	0.56	0.66±0.20
y w	3.11		7.67		0.32	
y w	4.02	3.56±0.89	8.46	8.07±0.77	0.65	0.49±0.32

	Lsp1α-2		*Lsd-2*		*lacZ*	
Fly Line	F.E.	Mean	F.E.	Mean	F.E.	Mean
P-lacZ:9B4	1.71		4.21		3.66	
	1.18	1.45±0.52	2.96	3.59±1.22	3.15	3.41±0.50
P-lacZ:19E7	1.13		3.00		4.15	
	1.19	1.16±0.06	3.40	3.20±0.39	3.71	3.93±0.43
y w	1.24		2.96			
y w	1.54	1.39±0.29	3.56	3.26±0.59		

	CG2556-5' UTR		*CG2556-3' UTR*	
Fly Line	F.E.	Mean	F.E.	Mean
P-lacZ:9B4			2.32	
			1.54	1.93±0.77
P-lacZ:19E7	1.04			
	1.15	1.10±0.09		

Table 29. Fold enrichment from chromatin immunoprecipitation with anti-H4K16ac on *P-lacZ:9B4*, *P-lacZ:19E7* and *y w* male third instar larval fat bodies

95% confidence intervals are indicated for 2 experiments. Fold enrichment (F.E.) is normalized to *Gpdh*, which is set to 1.

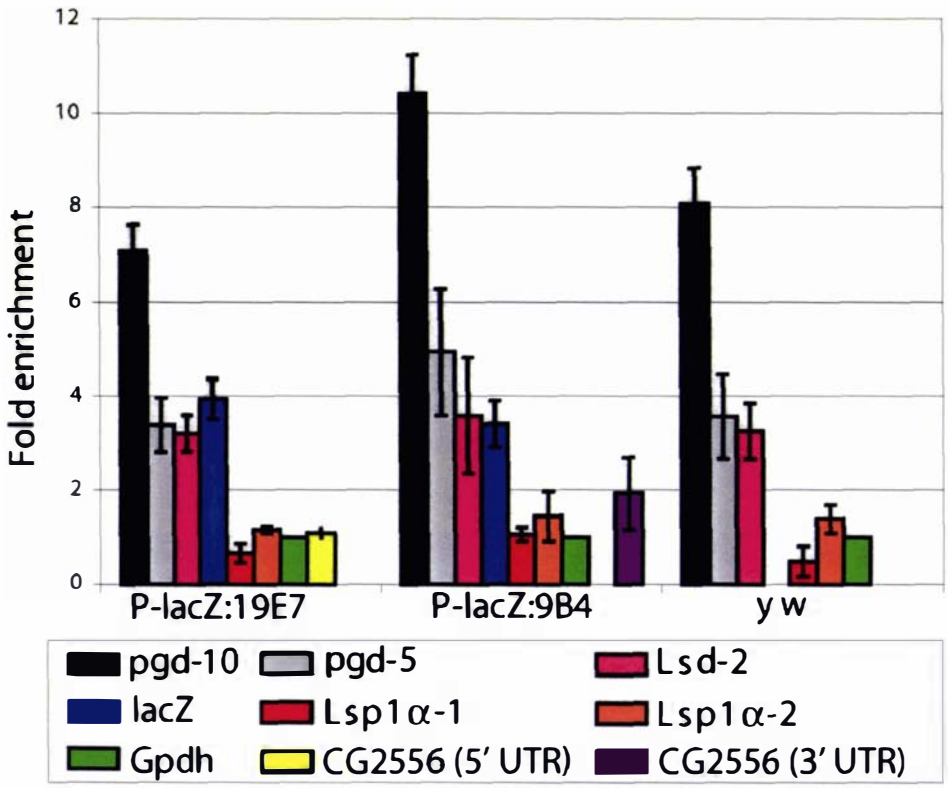


Figure 29. Mean fold enrichment from chromatin immunoprecipitation with anti-H4K16ac on *P-lacZ:9B4*, *P-lacZ:19E7* and *y w* male third instar larval fat body nuclei

Chromatin from male third instar larval fat body nuclei of the appropriate genotype was immunoprecipitated with antibody specific for H4K16ac. The fold enrichment of immunoprecipitated DNA relative to input DNA is shown with 95% confidence intervals indicated for 2 experiments. Fold enrichment is normalized to *Gpdh*, which is set to 1. A 3 - 10 fold enrichment is observed for the control genes *Pgd* and *Lsd-2* and both X-linked *P-lacZ* transgenes. A 2-fold enrichment is observed for the 3' UTR of *CG2556*. However, no enrichment is observed for *Lsp1α* or the 5' UTR of *CG2556*.

Genotype	Sample	Primer set						
		pgd10	*pgd5*	*Lsd-2*	*lacZ*	*Lsp1α-1*	*Lsp1α-2*	*GPDH*
		1.844	1.964	1.929	1.878	1.943	1.810	1.919
	ChIP	1.843	2.031	1.930	1.890	1.956	1.832	1.881
	#1	1.859	1.942	1.930	1.886	1.938	1.826	1.888
		1.860	1.932	1.950	1.892	1.952	1.834	1.886
	Input	1.843	1.966	1.936	1.900	1.953	1.848	1.902
	#1	1.834	1.914	1.922	1.887	1.958	1.836	1.861
		1.840	1.935	1.939	1.873	1.926	1.811	1.876
	ChIP	1.832	1.969	1.928	1.896	1.934	1.824	1.876
	#2	1.829	1.948	1.913	1.900	1.938	1.832	1.875
		1.875	1.925	1.936	1.905	1.963	1.824	1.876
	Input	1.831	1.942	1.930	1.888	1.944	1.839	1.880
P-lacZ:9B4	#2	1.810	1.933	1.930	1.891	1.964	1.834	1.860
		1.862	1.953	1.903	1.909	1.934	1.831	1.888
	ChIP	1.833	1.959	1.927	1.889	1.947	1.844	1.900
	#1	1.844	1.945	1.915	1.890	1.966	1.838	1.907
		1.826	1.933	1.915	1.890	1.962	1.818	1.886
	Input	1.836	1.943	1.904	1.895	1.954	1.835	1.873
	#1	1.826	1.979	1.902	1.888	1.995	1.822	1.885
		1.826	1.941	1.918	1.876	1.930	1.807	1.885
	ChIP	1.811	1.934	1.892	1.877	1.947	1.809	1.872
	#2	1.825	1.958	1.930	1.901	1.963	1.827	1.888
		1.832	1.953	1.878	1.900	1.924	1.808	1.790
	Input	1.830	1.957	1.899	1.885	1.964	1.819	1.903
P-lacZ:19E7	#2	1.836	1.968	1.934	1.892	1.983	1.835	1.908
negative		1.822	1.856	1.929	1.892	1.975	1.794	1.920
		1.830	1.866	1.878		1.810	1.841	1.880
	ChIP	1.814	1.876	1.882		1.806	1.817	1.885
	#1	1.849	1.875	1.906		1.783	1.820	1.868
		1.801	1.848	1.861		1.861	1.832	1.888
	Input	1.822	1.868	1.869		1.831	1.837	1.895
	#1	1.825	1.871	1.878		1.817	1.828	1.830
		1.867	1.855	1.916		1.800	1.828	1.931
	ChIP	1.850	1.859	1.906		1.781	1.842	1.924
	#2	1.649	1.899	1.895		1.764	1.830	1.938
		1.802	1.869	1.830		1.839	1.780	1.880
	Input	1.705	1.852	1.909		1.856	1.769	1.882
y w	#2	1.817	1.869	1.920		1.855	1.834	1.884
negative		1.687		1.883		1.938	1.813	1.892

Table 30. PCR efficiencies of real-time PCR data

Real-time PCR efficiencies were calculated using the LinRegPCR 7.0 program (Ramakers *et al.*, 2003). Triplicate reactions are shown for each sample.

7 PHYLOGENY

Some LSP1 protein sequences were available at the commencement of this study (Table 31). Recently completed and partial genome sequences are available for *D. erecta*, *D. simulans*, *D. yakuba*, *D. ananassae*, *D. mojavensis* and *D. virilis*. This provided an opportunity to determine when LSP1α arose during the evolution of *Drosophila* species. Putative LSP1α, LSP1β and LSP1γ proteins were identified by BLAST searching the recently published genome sequences of *D. erecta*, *D. simulans*, *D. yakuba*, *D. pseudoobscura*, *D. ananassae*, *D. mojavensis* and *D. virilis* with each of the LSP1 protein sequences from *D. melanogaster*. The nucleotide sequences encoding the putative LSP1 proteins were aligned using ClustalW. This alignment indicated the presence of a single conserved intron near the 5' end of each *Lsp1* gene. The putative *Lsp1* genes were translated and the LSP1 proteins aligned using ClustalW (section 10.7). LSP1β and LSP1γ, but not LSP1α, had previously been identified independently in *D. pseudoobscura* and *D. buzzatii* (Gonzalez *et al.*, 2004), as had LSP1γ in *D. simulans*. In this study, LSP1α homologues were identified in *D. erecta, D. simulans* and *D. yakuba*, but not in any of the other *Drosophila* species examined. In these three species, as in *D. melanogaster*, the *Lsp1α* gene is flanked by *CG2560* and *CG15730* (Table 32). In the other species these genes are immediately adjacent. Due to the unfinished nature of the sequences released for some of the *Drosophila* species, the LSP1α sequences of *D. erecta* and *D. simulans* are incomplete. The first 30 residues of *D. erecta* LSP1α are missing, as is some of the central sequence of *D. simulans* LSP1α. For this reason, a Neighbour-Joining phylogenetic tree was constructed from the ungapped regions of the sequence alignment in which all proteins are represented (Figure 31). The *D. miranda* LSP1γ sequence was not included in the alignment or the phylogenetic tree because no LSP1β sequence was available for this species as the genome sequence has not yet been published.

These results show that *Lsp1α* arose within the *melanogaster* subgroup of species after the split of the ancestor of *D. erecta* from the rest of the *Drosophila* species, which is estimated to have occurred around 8 million years ago[1] (Figure 30). Although previous

[1] Divergence times are based on Russo *et al.* (1995).

studies have shown that *Lsp1α* is absent from the *obscura* group of *Drosophila* (Gonzalez *et al.*, 2004), this split occurred between 25 – 30 million years ago, indicating a much more ancient origin for *Lsp1α*. This study shows that *Lsp1α* arose significantly more recently, as it is present in *D. erecta* but not in *D. ananassae*. These species diverged between 8 – 12 million years ago, thus indicating an evolutionary age for *Lsp1α* of 8 – 12 million years (Russo *et al.*, 1995).

The other *Drosophila* genes that appear to escape regulation by the MSL complex do not appear to have a recent autosomal origin, although these have not been well characterized. It is unclear whether there are any other recent evolutionary additions to the X chromosome, although the full release of genome sequence information for additional *Drosophila* species should facilitate this study in the future. In mammalian X chromosome inactivation, recent evolutionary additions to the X chromosome appear to escape the chromatin remodeling and silencing mediated by the *Xist* and *Tsix* noncoding RNAs (Brown and Greally, 2003). It is significant to note that some genes that escape silencing in humans, are inactivated in mice (Brown and Greally, 2003). It is possible that the lack of regulation of *Lsp1α* by the MSL complex exhibits some parallels with this escape from X inactivation in mammalian cells. If this is so, then it might be useful to determine whether the *Lsp1α* homologues in the most closely related *Drosophila* species also escape regulation by the MSL complex.

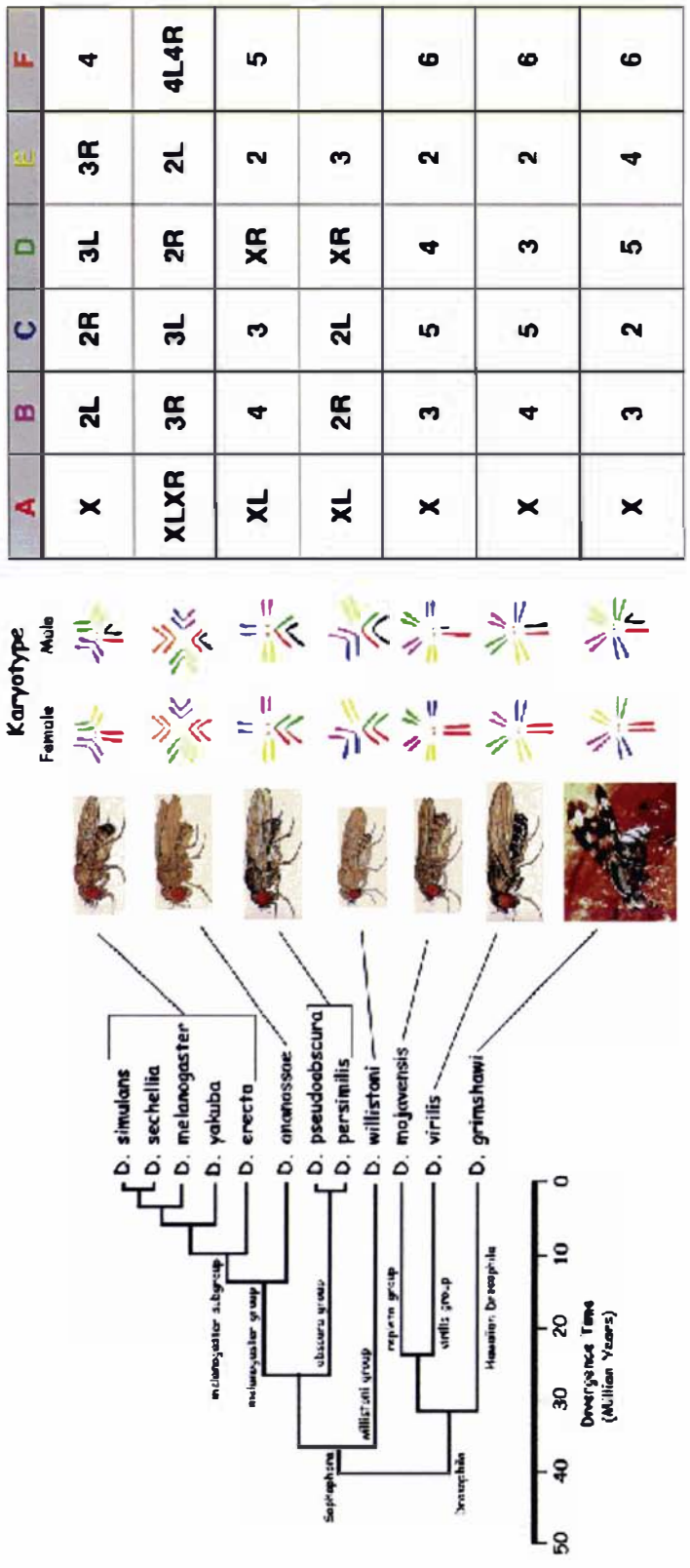


Figure 30. Phylogeny and karyotype of sequenced *Drosophila* species

Taken from http://species.flybase.net/. Divergence times are based on a linearized *Adh* clock (Russo *et al.*, 1995).

Protein	species	Accession Number
LSP1α	*D. melanogaster*	NP_511138
LSP1β	*D. melanogaster*	NP_476624
LSP1γ	*D. melanogaster*	NP_523868
LSP1β	*D. buzzatii*	AAT00494
LSP1γ	*D. buzzatii*	AAT00496
LSP1γ	*D. simulans*	AAB71667
LSP1γ	*D. miranda*	AAS87314

Table 31. Accession Numbers for LSP1 protein sequences

Species	Gene location				
	CG2560	*Lsp1α*	*CG15730*	*Lsp1β*	*Lsp1γ*
D. melanogaster	11A7-9	11A12	11A12	21E2	61A6
D. erecta	c6952[1]	c6951	c6951	c3971	c4915
D. simulans	c135.14 c135.15	c135.15 c135.16	c135.17	c8.95	c57.14
D. yakuba	c56.13	c56.13	c56.13	c14.16	c30.9
D. ananassae	c2236	-	c2235	c2620 c2621	c1157
D. mojavensis	c6725	-	c6725	c1135	c1948
D. virilis	c3192	-	c3192	c3716	c1051
D.pseudoobscura	X	-	X	2	3

Table 32. *Lsp1α* is flanked by *CG2560* and *CG15730* in four *Drosophila* species, but these genes lie immediately adjacent to one another in the other species that lack *Lsp1α*

[1] c: contig

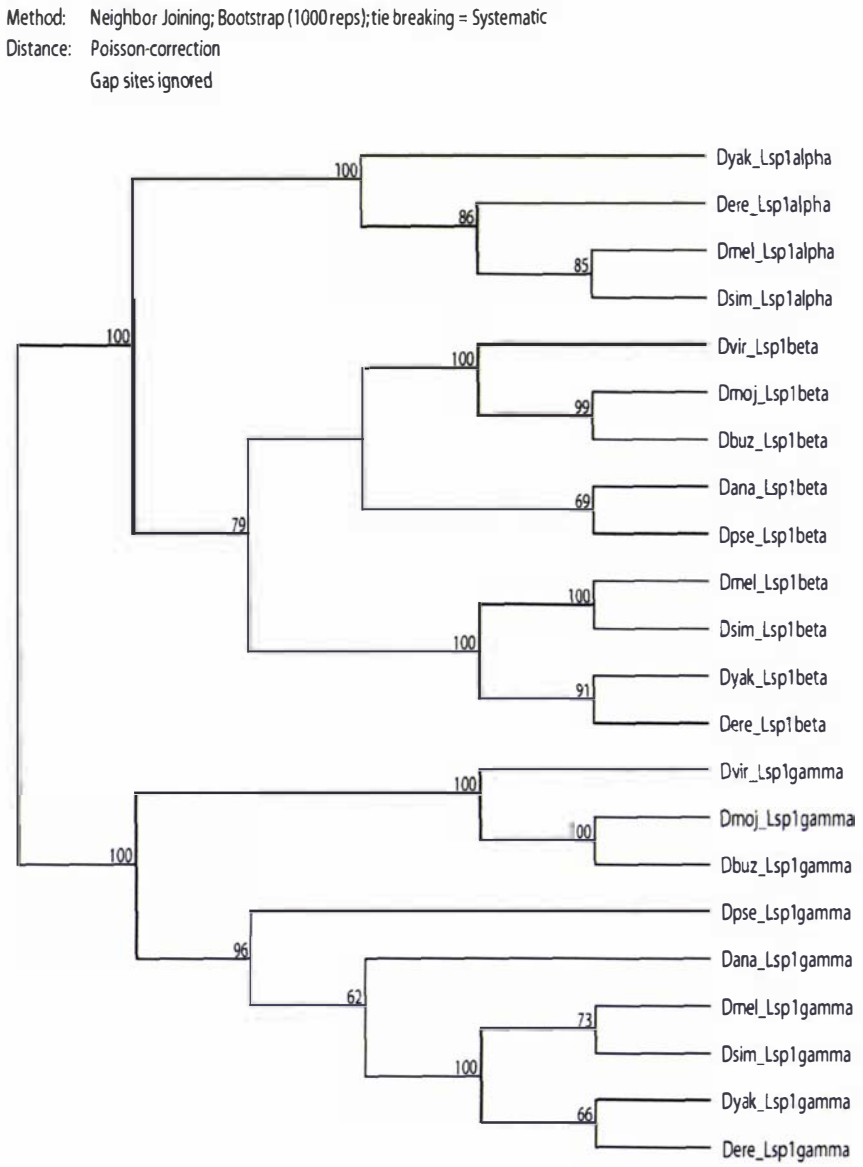


Figure 31. LSP1α is present in four closely related *Drosophila* species but is absent in more distantly related species, although LSP1β and LSP1γ are present in all species

Neighbour-Joining tree based on the ungapped regions of a ClustalW alignment of the LSP1α, LSP1β and LSP1γ protein sequences from *D. melanogaster* (Dmel), *D. yakuba* (Dyak), *D. erecta* (Dere), *D. simulans* (Dsim), *D. mojavensis (*Dmoj), *D. buzzatii* (Dbuz), *D. ananassae* (Dana), *D. pseudoobscura* (Dpse) and *D. virilis* (Dvir). Due to the unfinished nature of the sequence information available, the LSP1α proteins of *D. erecta* and *D. simulans* are incomplete.

8 DISCUSSION

Lsp1α has long been considered to be a well characterized example of a non-dosage compensated gene. However, the evidence for its lack of dosage compensation rested on the two-fold difference in expression levels between male and female larvae (Ghosh *et al.*, 1992; Roberts and Evans-Roberts, 1979a). The higher expression of *Lsp1α* observed in females could also potentially be due to sex-specific regulation of the promoter, or could in fact be caused by a lack of regulation by the MSL complex. In this study it is concluded that in spite of the promoter exhibiting some sex-specific regulation, *Lsp1α* is not regulated by the MSL complex, most likely because of its relatively recent evolutionary origin.

Surprisingly, initial findings in this study indicated that the *Lsp1α* promoter is more active in females than males of the same transgenic fly line. This was in contrast to the results of Ghosh *et al.* (1989), in which most lines showed equal transgene expression in male and female larvae. A clear linear correlation was found between the male activity level of autosomal transgenes and the female to male ratio of transgene activity. When the male activity level is low, the female to male activity level of the transgene is high. Conversely, when the male activity level is high, the female to male activity level of the transgene is low. This suggests that at repressive chromatin locations, in which the promoter is less active, the *Lsp1α* promoter is more active in females than in males.

Constitutive expression of MSL2 lowers *Lsp1α* promoter activity significantly in females, whether the transgene is autosomal or X-linked. This suggests that the *Lsp1α* promoter is regulated by the MSL complex itself, or alternatively by an individual component of the complex that acts independently of the whole complex. It appears unlikely that the *Lsp1α* promoter is regulated by the whole MSL complex, as *Lsp1α* is not enriched for the H4K16ac modification that is associated with regulation by the MSL complex. The simplest explanation is therefore that MSL2 represses the *Lsp1α* promoter. An alternative hypothesis is that an individual MSL component partially protects the *Lsp1α* promoter from position effects. In males, or in females expressing MSL2, this individual MSL protein is sequestered by the complex and is unavailable to protect the *Lsp1α* promoter. Hence transgenes subject to strong position effects, *i.e.*

exhibiting low levels of activity, have higher levels of *Lsp1α* promoter activity in females compared to males because of the partial relief of repressive chromatin effects by this individual MSL component in females, but not in males. High activity transgenes not subject to these strong position effects have equivalent *Lsp1α* promoter activity levels in both sexes, because there are no repressive effects to be overcome in females. Some support for this hypothesis comes from the finding that overexpression of MSL1 causes a slight increase in male *Lsp1α* promoter activity in one autosomal line. This increase was significant, but not large, and an increase in male promoter activity was not observed in another autosomal line when MSL1 was overexpressed. Thus it is difficult to make firm conclusions that MSL1 is responsible for this partial relief of position effects on the *Lsp1α* promoter. Furthermore, is possible that more than one component of the MSL complex, but not the entire complex itself, may be involved in this effect. The MSL complex's propensity to relieve negative position effects has been well documented (Kelley and Kuroda, 2003). There is evidence that individual MSL proteins can regulate gene expression independently of other members of the complex. For example MSL2 regulates *roX* gene expression (Rattner and Meller, 2004) and MLE regulates *roX2* expression (Lee *et al.*, 2004). It is also possible that the effect on the *Lsp1α* promoter may be indirectly mediated by a transcription factor that itself is regulated by MSL1 or MSL2.

There are two potential reasons for the discrepancies between the results observed in this study of the *Lsp1α* promoter, and those observed in the previous study of Ghosh *et al.* (1989). Firstly, the variability in the level of promoter upregulation in females might be sufficient to account for the finding that *Lsp1α* is expressed equally in male and female larvae when moved elsewhere on autosomes or on the X chromosome (Ghosh *et al.*, 1989). Several of the *P-lacZ* lines in this study exhibit female to male ratios close to one, and it is possible that by chance, four of the five autosomal lines in the Ghosh *et al.* (1989) study also inserted in active regions of chromatin and thus had relatively equal expression levels in both sexes. One of the two X-linked transgenes, and one of the five autosomal transgenes in the Ghosh *et al.* (1989) study showed some increased activity in females relative to males, which is more consistent with the results observed in this study. Secondly, it is possible that the larger genomic fragment used in the Ghosh *et al.*

(1989) study may have provided protection of the *Lsp1α* promoter from position effects, such that the transgenes all exhibited high levels of activity.

Despite the higher promoter activity of X-linked *P-lacZ*[1] transgenes in female compared to male larvae, these transgenes are clearly regulated by the MSL complex as evidenced by the enrichment of H4K16ac. *Lsp1α*, itself however, is not enriched for this histone modification. Thus it appears that the sex-specific regulation of the *Lsp1α* promoter is not sufficient to account for the two-fold difference in *Lsp1α* transcript levels between male and female larvae. If this sex-specific regulation was the cause of the non-dosage compensated appearance of *Lsp1α*, it would be expected that this gene would have shown H4K16ac enrichment levels equivalent to those of the X-linked transgenes. Instead this lack of enrichment of H4K16ac strongly indicates that *Lsp1α* is not hypertranscribed by the MSL complex. It is plausible that *Lsp1α* resides within a region in which its promoter is not subject to negative position effects, and instead exhibits high levels of activity with the corresponding equal level of female and male activity. Consistent with this suggestion, *Lsp1α* mRNA is clearly very abundant in third instar larvae (section 3.2) (Davies *et al.*, 1986). These chromatin immunoprecipitation results indicate that *Lsp1α* is not regulated by the MSL complex, and that this is the cause of the two-fold difference in transcript levels between males and females. This lack of regulation by the entire MSL complex does not preclude the possibility discussed previously that the *Lsp1α* promoter is subject to regulation by an individual component of the MSL complex that acts independently of the entire complex. Any individual component of the MSL complex, with the exception of MOF, would not acetylate H4K16, the histone modification essential for regulation by the MSL complex.

Another surprising finding in this study was that the autosomal *arm-lacZ* and *pgd-lacZ* transgenes were *ca.* 25% more active in male compared to female larvae. These reporter constructs are equally expressed in adult male and female flies (Henry *et al.*, 2001; Scott and Lucchesi, 1991). The basis for this difference is not clear. Both *armadillo* and *Pgd* are X-linked genes. Hence it is possible that in some tissues and stages these promoters can recruit the MSL complex. It has recently been shown that regions of the X chromosome that lack a CES can attract the MSL complex (Oh *et al.*, 2004). Consistent

[1] In these transgenes, the *Lsp1α* promoter controls *lacZ* expression.

with this explanation, one X-linked *arm-lacZ* transgene was equally expressed in male and female larvae as was one *pgd-lacZ* transgene. However, additional X-linked lines would be required to confirm these observations.

The genes flanking *Lsp1α* are dosage compensated in the developmental stage at which they are expressed as determined by northern hybridization analysis and RNase protection analysis. The two major models proposed for the recognition and targeting of the MSL complex to the X chromosome in males are consistent with two distinct possibilities for how *Lsp1α* escapes regulation by the MSL complex. If the MSL complex binds to the chromatin entry sites and spreads from these into X chromosome genes, then elements must exist between *Lsp1α* and the flanking dosage compensated genes that block the spread of the MSL complex. Previous work indicated that typical insulator elements, such as the *gypsy* element bound by the Su(Hw) protein are unable to prevent dosage compensation of an X-linked transgene (Roseman *et al.*, 1993). In this study no such boundary elements were found in the intergenic regions between *Lsp1α* and its neighbours that could block dosage compensation of an X-linked reporter gene. However, it is possible that uncharacterized elements may exist outside of the regions tested, for example in the large first intron of *CG2556*, or that these elements are specific to third instar larvae. There is also the potential that the X-linked *armadillo* promoter, used in the reporter construct, is able to attract the MSL complex directly, although the MSL complex does not bind to the reporter construct at autosomal locations.

The second model proposes that X-linked genes have associated sequence motifs that target the MSL complex. Bioinformatic analysis in this study of recently released *Drosophila* species genomes enabled the discovery of *Lsp1α* in four species closely related to *D. melanogaster*. *Lsp1α* is absent however, in the other five species examined. Gonzalez *et al.* independently confirmed that *Lsp1α* is absent from *D. pseudoobscura,* and also showed that it is lacking in *D. buzzatii* (Gonzalez *et al.*, 2004). Phylogenetic analysis in this study of the LSP1 proteins from these species shows that LSP1β and LSP1γ form separate clades, indicating that these proteins were both present in an ancestor of all the *Drosophila* species. LSP1α however, is part of a clade sharing a common ancestor with LSP1β, indicating that LSP1α arose during evolution from a

fairly recent duplication and translocation of LSP1β in an ancestor common to *D. yakuba*, *D. erecta*, *D. melanogaster* and *D. simulans*. LSP1α is not present in *D. ananassae* indicating that it arose between the split of the *melanogaster* subgroup from the *melanogaster* group between 8 – 12 million years ago (Russo *et al.*, 1995). This is *ca.* 13 million years later than the origin proposed by Gonzalez *et al.* (2004), and indicates a very recent evolutionary origin for *Lsp1α*. Since *Lsp1α* appears to have arisen by this fairly recent gene duplication and transposition event, it is possible that it has not yet evolved MSL binding sites.

The *CG2560* gene is dosage compensated and lies less than 1 kb upstream of the 5' end of *Lsp1α*. Although *CG2560* is expressed in the same developmental stage as *Lsp1α* as determined by northern hybridization analysis, it is not expressed in the same tissue (fat body) as determined by RT-PCR. The MSL complex binds to actively expressed genes (Sass *et al.*, 2003). Since *CG2560* is not transcribed in third instar larvae fat body nuclei, the MSL binding sites normally used to regulate *CG2560* transcription may not be accessible to the MSL complex (Figure 32). In contrast, the downstream gene, *CG2556*, is transcribed in third instar larval fat cells. The 5' end of *CG2556* is *ca.* 4.6 kb from the 3' end of *Lsp1α*, but appears to not be modified by the MSL complex, as no enrichment of H4K16ac was detected in the 5' UTR region. A modest enrichment for H4K16ac was found at the 3' end of *CG2556*, which is *ca.* 11 kb from the 3' end of *Lsp1α*. The lack of enrichment of H4K16ac at the 5' end of a gene has not previously been observed, as the X-linked genes *Pgd* and *Zw* both had significant levels of H4K16ac enrichment on their promoter and 5' regions in cell culture (Smith *et al.*, 2001). The overall fold enrichments observed in the chromatin immunoprecipitation experiments in this study were significantly lower than those observed in previous studies, even using identical primers for *Pgd* and GPDH (Smith *et al.*, 2001). It is likely that these lower enrichments were due to the use of fat body tissue rather than cell culture, as fat body tissue is not a homogeneous population as is cell culture and is also rich in enzymes that may adversely affect the level of enrichment of H4K16ac. These fold enrichments were also lower than those obtained from embryos (Smith *et al.*, 2001) and it may be possible that the H4K16ac modification is enriched at higher levels in dividing cells, such as those in embryos or cell culture, compared to terminally differentiated cells, such as those in fat body tissue of third instar larvae.

In the model shown in Figure 32, it is proposed that the MSL complex is targeted to *CG2556* by sequence motifs near its 3' end, but cannot spread over the *ca.* 11 kb region from *CG2556* into *Lsp1α* (Figure 32B). It is also possible that the intron-less gene, *CG15730*, that lies between *CG2556* and *Lsp1α* and appears not to be transcribed in third instar larvae, is in an inactive chromatin state that blocks the spread of the MSL complex (Figure 32C). This possibility is discussed further below (section 8.1). It is suggested that the X-linked *P-lacZ* transgenes showing enrichment for H4K16ac have inserted in actively transcribed regions (*e.g.* Figure 32A), in which there is abundant MSL complex. This is supported by the *P-lacZ:19E7* transgene, which has inserted near the *Ntf-2* gene that exhibits a mutant phenotype in the fat body tissue of third instar larvae (Bhattacharya and Steward, 2002). Since it is likely that *Ntf-2* is expressed in the same tissue and stage as the *Lsp1α* promoter, it follows that the MSL complex should be able to spread from this flanking gene into the *P-lacZ* transgene. In a similar manner, autosomal genes such as *Adh* are fully dosage compensated at sites on the X chromosome (Sass and Meselson, 1991). This does not preclude the possibility that *Lsp1α* exists within a unique chromatin structure that is prohibitive to regulation by the MSL complex. In this model it is proposed that the targeting sites that attract the MSL complex are not necessarily associated directly with each X-linked gene, but may be scattered in nearby intergenic regions. This accounts for the finding that many X-linked genes do not retain full dosage compensation when inserted on an autosome, including the *armadillo* promoter used in this study (section 4.4) (Arkhipova *et al.*, 1997; Qian and Pirrotta, 1995; Sass and Meselson, 1991).

If this model is correct, then a CES or X-linked fat body-expressed gene inserted immediately 3' of *Lsp1α* should induce dosage compensation of *Lsp1α*. In order to test this, it was attempted to insert various reporter constructs within the open reading frame of *Lsp1α* using homologous recombination (Gong and Golic, 2003). However, attempts were unsuccessful (section 5). In brief, the failure of this method was due to the inability of the FLP-recombinase to excise the reporter construct flanked by the FRT sites despite the construct matching the expected pW30 vector sequence. It is unclear whether the FRT sites within the pW30 plasmid are suitable for excision by the FLP-recombinase and an alternative approach would be necessary in future experiments to

insert constructs within the *Lsp1α* gene domain. This alternative approach is discussed below (section 8.1).

In conclusion, the results of this study support a model for dosage compensation similar to that proposed originally by Lucchesi & Manning (1987) and supported by Fagegaltier and Baker (2004), in which affinity sequences for the MSL complex are associated with most X-linked genes. The MSL complex requires clusters of these sequences for this association, and can spread very short distances from these sequences into adjacent actively expressed genes. In this model, inactive genes are inaccessible to the complex, possibly due to the presence of a repressive histone modification.

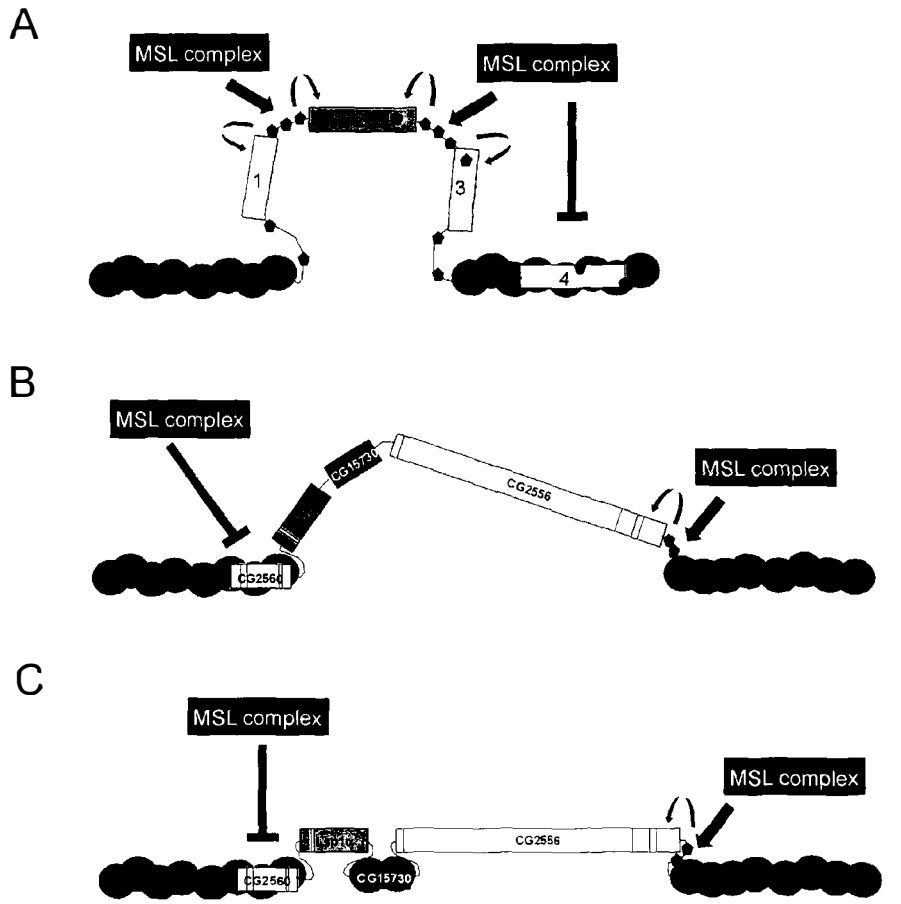


Figure 32. Model: *Lsp1α* is not regulated by the MSL complex because spreading of the complex is limited

(A) The MSL complex is targeted to actively expressed genes (*e.g.* 1, 2 & 3) by clusters of binding sites. Inactive genes (*e.g.* 4) are inaccessible to the complex regardless of the presence of binding sites. The MSL complex can spread over very short distances from these binding sites into nearby chromatin, regulating close-by genes. (B) In fat body nuclei of third instar larvae, the MSL complex is targeted to the actively transcribed *CG2556* by a cluster of binding sites near the 3' end of this gene. The chromatin region downstream of *CG2556,* containing *CG11146*, has not been studied, and may also be active in this tissue. *CG2560* is in an inactive chromatin formation in fat body tissue. *Lsp1α* is active in fat body nuclei, but lacks MSL binding sites. The MSL complex cannot spread from the 3' end of *CG2556* into *Lsp1α.* (C) *CG15730* may also be in an inactive chromatin formation in fat body nuclei. This inactive form may block the MSL complex from spreading from *CG2556* into *Lsp1α.* However, this region (I2) (defined in Figure 21) does not block dosage compensation of an X-linked *arm-lacZ* transgene but is possible that any spreading of the MSL complex may occur through the other, 5' (I) region into the reporter gene.

8.1 Future Directions

This study was complicated somewhat by the unexpected result that the *Lsp1α* promoter exhibits variable sex-specific regulation induced by an individual MSL protein(s), yet still escapes regulation by the MSL complex. Sex-specific regulation of the promoter had not previously been described, and this made initial analysis of the dosage compensation status of *Lsp1α* difficult. Further studies would involve examining the regulation of the promoter in further detail, and confirming how *Lsp1α* escapes dosage compensation.

This study has provided two candidates for the sex-specific regulation of the *Lsp1α* promoter: MSL1 and MSL2. MSL2 downregulates promoter activity two-fold when ectopically expressed in females, and MSL1 induces upregulation of the promoter in one or both sexes when over-expressed ectopically. It is hypothesized that MSL1, potentially in conjunction with other MSL proteins such as MLE, partially protects the *Lsp1α* promoter from negative position effects in females. MSL1 is sequestered by the MSL complex in males due to the presence of MSL2, and is unavailable to protect the promoter. The DNaseI hypersensitive sites (DHS) in the *Lsp1α* promoter provide good candidates for potential binding sites for MSL1, and other factors (Jowett, 1985). Examination of the *Lsp1α* promoter for sex-specific DHS may uncover potential binding sites for MSL1 and any additional factors. The model proposed in this study would suggest that there might exist female-specific DHS. Once any sex-specific potential binding sites were discovered, gel-mobility assays with extract from male and female third instar larvae, and with *in vitro* translated MSL1 could be conducted to determine if factors specifically bound these sites.

In the model proposed to explain how *Lsp1α* escapes regulation by the MSL complex, it is suggested that *Lsp1α* lacks the sequences required to attract the MSL complex because of its recent autosomal evolutionary origins and because the targeting sites surrounding it are inaccessible or too far away from it. It follows therefore that *Lsp1α*, similar to the *P-lacZ* transgenes, is capable of being regulated by the MSL complex. A perfect test of this would be to insert a known MSL targeting site, such as the 18D10 multimer, or an X-linked gene expressed in the same tissue such as *Lsd-2*, into *Lsp1α*, and examine if these are sufficient to equalize male and female expression and

induce H4K16ac of *Lsp1α*. However, homologous recombination in *Drosophila* is technically difficult and attempts in this study were not successful. Generally this approach did not work because of problems with excision by the FRT site-specific recombinase. If this approach was to be used in the future, it would be desirable to generate targeting plasmids that do not contain X-linked DNA such as *mini-white* or the *P*-elements (*e.g.* piggyBac), and that use fluorescent markers as described in (section 5.6). The generation of a successful homologous recombination method to insert transgenes into the ORF of *Lsp1α* would provide a powerful system in which to examine sequences required to recruit the MSL complex to X-linked genes.

An alternative and complementary approach would involve increasing the concentration of MSL complexes on the X chromosome. Overexpression of MSL1 and MSL2 leads to a puffed X chromosome (Oh *et al.*, 2003). This suggests that as the concentration of complexes increases, the MSL complex may bind to usually non-dosage compensated regions as its affinity for weak binding sites is increased (Fagegaltier and Baker, 2004). It is possible therefore that *Lsp1α* may become dosage compensated in an *hsp83-MSL1*, *hsp83-MSL2* fly line.

Further chromatin immunoprecipitation experiments could be used to help determine how *Lsp1α* escapes regulation by the MSL complex. For example, finer mapping of H4K16ac enrichment throughout the *Lsp1α* domain, including 3' of *Lsp1α* within *CG15730*, and 5' of *Lsp1α* in *CG2560,* might provide further support for the argument that the MSL complex is not active in the region immediately surrounding *Lsp1α* in the tissue and stage at which it is expressed. However, these experiments are particularly time consuming as they require the manual dissection of 200 larval fat bodies which are individually snap frozen in liquid nitrogen, prior to nuclei extraction and immunoprecipitation. Chromatin immunoprecipitation to determine whether NURF is enriched on *Lsp1α* may indicate whether its presence precludes that of the MSL complex. Intriguingly, a recent presentation at the Keystone Chromatin Modification Symposia (2005) indicates a role for NURF in regulating ecdysone-stimulated genes, of which *Lsp1α* is one (Carl Wu, National Cancer Institute, Bethesda, USA, 2005). It is thus possible that NURF may be involved in the regulation of the *Lsp1α* promoter. NURF functions antagonistically to the MSL complex (Corona *et al.*, 2002; Nishioka *et*

al., 2002a; Nishioka *et al.*, 2002b), and it is possible that NURF may prevent the MSL complex from regulating and dosage compensating *Lsp1α*.

It may also be possible using chromatin immunoprecipitation to determine whether there is an enrichment on the *Lsp1α* promoter of any of the individual MSL proteins that might be involved in the sex-specific regulation of *Lsp1α*, *e.g.* MSL1 or MSL2. This could be facilitated by performing chromatin immunoprecipitation on fat body nuclei extracted from third instar larvae expressing MSL2 containing a HA or Biotin tag.

Another histone modification of interest that could be examined by chromatin immunoprecipitation is H4K20 methylation. The presence of this methylation inversely correlates with the presence of H4K16ac, and H4K20 methylation is under-represented on the male X chromosome (Nishioka *et al.*, 2002a; Nishioka *et al.*, 2002b). It is possible that H4K20 methylation is enriched on *Lsp1α* and that this prevents the acetylation of H4K16, but this is unlikely as *Lsp1α* is highly expressed in the fat body tissue of third instar larvae, and H4K20 methylation is more typically associated with silencing (Karachentsev *et al.*, 2005). However, although this study has shown that there are no insulator elements between *Lsp1α* and the flanking dosage compensated genes able to block the MSL complex from dosage compensating an X-linked reporter gene, it is possible that the silent state of the flanking genes, *CG15730* and *CG2560*, prevents the MSL complex from spreading from *CG2556* into *Lsp1α*. When inserted elsewhere on the X chromosome, the (I) region between *CG2560* and *Lsp1α* is not silenced and therefore the MSL complex could spread from an active gene at the 5' end of the construct into the reporter gene, dosage compensating it. As the H4K20 methylation is associated with silencing, and acts antagonistically to H4K16ac, is it possible that this modification is enriched on the silenced genes flanking *Lsp1α* in the fat body tissue of third instar larvae. Determining whether H4K20 methylation is enriched on *CG15730* and *CG2560* in this tissue could provide some explanation as to how the MSL complex is prevented from spreading into *Lsp1α*. However, H4K20 can be subjected to mono-, di- and tri- methylation, and the published data do not indicate which of these modifications is present at reduced levels on the male X chromosome. Thus the specific antibody to be used in this type of chromatin immunoprecipitation experiment is

unclear. An alternative approach to examining this possibility would be to assay male and female β-galactosidase activities of an X-linked *I2-arm-lacZ-I2* transgene, in which the *arm-lacZ* reporter construct is flanked by the (I2) regions containing the *CG15730* gene. If the repressive chromatin state of the *CG15730* gene is able to block the MSL complex from spreading into and dosage compensating *arm-lacZ*, then females would exhibit twice the level of β-galactosidase activity compared to males. However, this proposal is complicated by the fact that while the *armadillo* promoter is constitutively active, the tissue or stage in which *CG15730* is expressed has not been determined.

In conclusion, this study has shown that the two-fold difference in *Lsp1α* transcript levels between male and female larvae is due to a lack of regulation by the MSL complex as evidenced by the absence of H4K16ac. Surprisingly, the *Lsp1α* promoter was found to be more active in females than in males at repressive chromatin locations. Some evidence indicates that this may be due to sex-specific regulation by an individual MSL component that acts independently of the entire MSL complex. Two of the genes flanking *Lsp1α* are dosage compensated, and the regions between *Lsp1α* and these flanking genes do not block the MSL complex from regulating an X-linked transgene. Only the gene 3' of *Lsp1α* is expressed in the same tissue and stage as *Lsp1α*, and this gene is enriched for H4K16ac at its 3', but not at its 5' end which lies closest to *Lsp1α*. Phylogenetic analysis indicates that *Lsp1α* has arisen fairly recently in evolutionary terms, and may not have yet acquired the targeting sites required to attract the MSL complex. A model is proposed in which the MSL complex does not spread into and regulate *Lsp1α* due to its lack of targeting sites, despite its presence on the 3' end of the neighbouring gene, 11 kb downstream. This study has provided some insight into the mechanism by which *Lsp1α* escapes regulation by the MSL complex, but further work is required to elucidate the exact reason for the failure of the MSL complex to regulate *Lsp1α*.

9 REFERENCES

Adams, M. D. Celniker, S. E. Holt, R. A. Evans, C. A. Gocayne, J. D. Amanatides, P. G. Scherer, S. E. Li, P. W. Hoskins, R. A. Galle, R. F. *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.

Akhtar, A. (2003). Dosage compensation: an intertwined world of RNA and chromatin remodeling. *Curr Opin Genet Dev* **13**, 161-9.

Akhtar, A. and Becker, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell* **5**, 367-75.

Akhtar, A. and Becker, P. B. (2001). The histone H4 acetyltransferase MOF uses a C2HC zinc finger for substrate recognition. *EMBO Rep* **2**, 113-8.

Akhtar, A., Zink, D. and Becker, P. B. (2000). Chromodomains are protein-RNA interaction modules. *Nature* **407**, 405-409.

Amrein, H. and Axel, R. (1997). Genes expressed in neurons of adult male *Drosophila*. *Cell* **88**, 459-469.

Arkhipova, I., Li, J. and Meselson, M. (1997). On the mode of gene-dosage compensation in *Drosophila*. *Genetics* **145**, 729-736.

Ashburner, M. (1989a). *Drosophila*: A laboratory handbook. New York: Cold Spring Harbor Laboratory Press.

Ashburner, M. (1989b). *Drosophila*: A laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1997). Current protocols in molecular biology, vol. 1 (ed. V. Benson Chando): John Wiley & Sons, Inc.

Badenhorst, P., Voas, M., Rebay, I. and Wu, C. (2002). Biological functions of the ISWI chromatin remodeling complex NURF. *Genes & Development* **16**, 3186-3198.

Bai, X., Alekseyenko, A. A. and Kuroda, M. I. (2004). Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *Embo J* **23**, 2853-61.

Barges, S., Mihaly, J., Galloni, M., Hagstrom, K., Muller, M., Shanower, G., Schedl, P., Gyurkovics, H. and Karch, F. (2000). The *Fab-8* boundary defines the distal limit of the bithorax complex *iab-7* domain and insulates *iab-7* from initiation elements and a PRE in the adjacent *iab-8* domain. *Development* **127**, 779-790.

Bashaw, G. J. and Baker, B. S. (1995). The *msl-2* dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. *Development* **121**, 3245-3258.

Bell, A. C., West, A. G. and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387-396.

Belote, J. M. (1983). Male-specific lethal mutations of *Drosophila melanogaster*. II. Parameteters of gene action during male development. *Genetics* **105**, 881-896.

Belote, J. M. and Lucchesi, J. C. (1980a). Male-specific lethal mutations of *Drosophila melanogaster*. *Genetics* **96**, 165-186.

Belote, J. M. and Lucchesi, J. C. (1980b). Control of X chromosome transcription by the *maleless* gene in *Drosophila*. *Nature* **285**, 573-575.

Berger, S. L. (2002). Histone modifications in transcriptional regulation. *Current Opinion in Genetics and Development* **12**, 142-148.

Berger, S. L. and Felsenfeld, G. (2001). Chromatin goes global. *Molecular Cell* **8**, 263-268.

Bernstein, E. and Allis, C. D. (2005). RNA meets chromatin. *Genes Dev* **19**, 1635-55.

Bernstein, M. and Cline, T. W. (1994). Differential effects of Sex-lethal mutations on dosage compensation early in *Drosophila* development. *Genetics* **136**, 1051-1061.

Bhadra, U., Pal-Bhadra, M. and Birchler, J. A. (1999). Role of the male specific lethal (msl) genes in modifying the effects of sex chromosomal dosage in Drosophila. *Genetics* **152**, 249-68.

Bhattacharya, A. and Steward, R. (2002). The Drosophila homolog of NTF-2, the nuclear transport factor-2, is essential for immune response. *EMBO Rep* **3**, 378-83.

Blanton, J., Gaszner, M. and Schedl, P. (2001). Interaction between two boundary proteins, zeste-white 5 and BEAF. In *Annual Drosophila Research Conference 42nd*, (ed. Washington DC.

Bone, J. R., Lavender, J., Richman, R., Palmer, M. J., Turner, B. M. and Kuroda, M. I. (1994). Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila*. *Genes & Development* **8**, 96-104.

Bourbon, H. M., Gonzy-Treboul, G., Peronnet, F., Alin, M. F., Ardourel, C., Benassayag, C., Cribbs, D., Deutsch, J., Ferrer, P., Haenlin, M. *et al.* (2002). A P-insertion screen identifying novel X-linked essential genes in Drosophila. *Mech Dev* **110**, 71-83.

Breen, T. R. and Lucchesi, J. C. (1986). Analysis of the dosage compensation of a specific transcript in *Drosophila melanogaster*. *Genetics* **112**, 483-491.

Brock, H. W. and Roberts, D. B. (1982). The *Lsp1-alpha* gene is not dosage compensated in the *Drosophila melanogaster* species subgroup. *Biochemical Genetics* **20**, 287-296.

Brock, H. W. and Roberts, D. B. (1983). Location of the *Lsp-1* genes in *Drosophila* species by *in situ* hybridization. *Genetics* **103**, 75-92.

Brown, C. J. and Greally, J. M. (2003). A stain upon the silence: genes escaping X inactivation. *Trends Genet* **19**, 432-8.

Buscaino, A., Kocher, T., Kind, J. H., Holz, H., Taipale, M., Wagner, K., Wilm, M. and Akhtar, A. (2003). MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. *Mol Cell* **11**, 1265-77.

Cai, H. N. and Levine, M. (1995). Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature* **376**, 533-536.

Cai, H. N. and Levine, M. (1997). The *gypsy* insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *EMBO Journal* **16**, 1732-1741.

Cai, H. N. and Shen, P. (2001). Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* **291**, 493-495.

Camfield, G. R. (1974). A genetic and biochemical study of a temperature-sensitive vermilion mutation in *Drosophila melanogaster*. In *Department of Zoology*, (ed. Vancouver, Canada: University of British Columbia.

Cavalli, G. (2002). Chromatin as a eukaryotic template of genetic information. *Current Opinion in Cell Biology* **14**, 269-278.

Chen, S. and Corces, V. G. (2001). The *gypsy* insulator of *Drosophila* affects chromatin structure in a directional manner. *Genetics* **159**, 1649-1658.

Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M. and Allis, C. D. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Molecular Cell* **5**, 905-915.

Chung, J. H., Bell, A. C. and Felsenfeld, G. (1997). Characterization of the chicken beta-globin insulator. *Proc Natl Acad Sci U S A* **94**, 575-80.

Chung, J. H., Whiteley, M. and Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila. Cell* **74**, 505-514.

Copps, K., Richman, R., Lyman, L. M., Chang, K. A., Rampersad-Ammons, J. and Kuroda, M. I. (1998). Complex formation by the *Drosophila* MSL proteins: Role of the MSL2 RING finger in protein complex assembly. *EMBO Journal* **17**, 5409-5417.

Corona, D. F., Clapier, C. R., Becker, P. B. and Tamkun, J. W. (2002). Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep* **3**, 242-7.

Corona, D. F., Langst, G., Clapier, C. R., Bonte, E. J., Ferrari, S., Tamkun, J. W. and Becker, P. B. (1999). ISWI is an ATP-dependent nucleosome remodeling factor. *Molecular Cell* **3**, 239-245.

Coschigano, K. T. and Wensink, P. C. (1993). Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila. Genes & Development* **7**, 42-54.

Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J. *et al.* (2004). Histone deimination antagonizes arginine methylation. *Cell* **118**, 545-53.

Davies, J. A., Addison, C. F., Delaney, S. J., Sunkel, C. and Glover, D. M. (1986). Expression of the prokaryotic gene for chloramphenicol acetyl transferase in Drosophila under the control of larval serum protein 1 gene promoters. *J Mol Biol* **189**, 13-24.

Delaney, S. J., Sunkel, C. E., Genova-Seminova, G., Davies, J. E. and Glover, D. M. (1987). *Cis*-acting sequences sufficient for correct tissue and temporal specificity of larval serum protein 1 genes of *Drosophila. EMBO Journal* **6**, 3849-3854.

Demakova, O. V., Kotlikova, I. V., Gordadze, P. R., Alekseyenko, A. A., Kuroda, M. I. and Zhimulev, I. F. (2003). The MSL complex levels are critical for its correct targeting to the chromosomes in *Drosophila melanogaster. Chromosoma* **112**, 103-15.

Deryckere, F. and Gannon, F. (1994). A one-hour minipreparation technique for extraction of DNA-binding proteins from animal tissues. *Biotechniques* **16**, 405.

Deuring, R., Fanti, L., Armstrong, J. A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S. L., Berloco, M. *et al.* (2000). The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure *in vivo. Molecular Cell* **5**, 355-365.

Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L. and Roeder, R. G. (2005). Physical association and

coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873-885.

Fagegaltier, D. and Baker, B. S. (2004). X chromosome sites autonomously recruit the dosage compensation complex in Drosophila males. *PLoS Biol* **2**, e341.

Farkas, G., Gausz, J., Galloni, M., Reuter, G., Gyukovics, H. and Karch, F. (1994). The *trithorax-like* gene encodes the *Drosophila* GAGA factor. *Nature* **371**, 806-808.

Fitzsimons, H. (1998). Development of a reporter gene assay to identify control elements required for dosage compensation in *Drosophila melanogaster*. In *Institute of Molecular BioSciences*, (ed. Palmerston North: Massey University.

Fitzsimons, H. L., Henry, R. A. and Scott, M. J. (1999). Development of an insulated reporter system to search for *cis*-acting DNA sequences required for dosage compensation in *Drosophila*. *Genetica* **105**, 215-226.

Franke, A. and Baker, B. S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. *Mol Cell* **4**, 117-22.

Fukunaga, A., Tanaka, A. and Oishi, K. (1975). *Maleless*, a recessive autosomal mutant of *Drosophila melanogaster* that specifically kills male zygotes. *Genetics* **81**, 135-141.

Gaszner, M., Vazquez, J. and Schedl, P. (1999). The Zw5 protein, a component of the *scs* chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes & Development* **13**, 2098-2107.

Gdula, D. A., Gerasimova, T. I. and Corces, V. G. (1996). Genetic and molecular analysis of the *gypsy* chromatin insulator of *Drosophila*. *PNAS* **93**, 9378-9383.

Gdula, D. A., Sandaltzopoulos, R., Tsukiyama, T., Ossipow, V. and Wu, C. (1998). Inorganic pyrophosphatase is a component of the Drosophila nucleosome remodeling factor complex. *Genes Dev* **12**, 3206-16.

Gebauer, F., Grskovic, M. and Hentze, M. W. (2003). Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. *Mol Cell* **11**, 1397-404.

Georgel, P. T., Tsukiyama, T. and Wu, C. (1997). Role of histone tails in nucleosome remodeling by Drosophila NURF. *EMBO J.* **16**, 4717-26.

Gerasimova, T. I. and Corces, V. G. (2001). Chromatin insulators and boundaries: Effects on transcription and nuclear organization. *Annual Review of Genetics* **35**, 193-208.

Gergen, J. P. (1987). Dosage compensation in *Drosophila*: Evidence that *daughterless* and *sex-lethal* control X chromosome activity at the blastoderm stage of embryogenesis. *Genetics* **117**, 477-485.

Geyer, P. K., Green, M. M. and Corces, V. G. (1988). Reversion of a *gypsy*-induced mutation at the yellow (*y*) locus of *Drosophila melanogaster* is associated with the insertion of a newly defined transposable element. *PNAS* **85**, 3938-3942.

Ghosh, A. K. and Mukherjee, A. S. (1992). Replicative activity of X-chromosome and autosomes of *Drosophila melanogaster* in autosomal hyperploids and the problem of dosage compensation. *Indian J Exp Biol* **30**, 557-66.

Ghosh, S., Chatterjee, R. N., Bunick, D., Manning, J. E. and Lucchesi, J. C. (1989). The *Lsp1-alpha* gene of *Drosophila melanogaster* exhibits dosage compensation when it is relocated to a different site on the X chromosome. *EMBO Journal* **8**, 1191-1196.

Ghosh, S., Lucchesi, J. C. and Manning, J. E. (1992). The non-dosage compensated *Lsp1-alpha* gene of *Drosophila melanogaster* lies immediately downstream of the dosage compensated *L12* gene. *Molecular General Genetics* **233**, 49-52.

Golubovsky, M. D. and Ivanov, Y. N. (1972). Autosomal mutations in *Drosophila melanogaster* killing the males and connected with female sterility. *Drosophila Information Service* **49**, 117.

Gong, W. J. and Golic, K. G. (2003). End-out, or replacement, gene targeting in *Drosophila*. *PNAS* **100**, 2556-2561.

Gonzalez, J., Casals, F. and Ruiz, A. (2004). Duplicative and conservative transpositions of larval serum protein 1 genes in the genus Drosophila. *Genetics* **168**, 253-64.

Gorman, M., Franke, A. and Baker, B. S. (1995). Molecular characterization of the male-specific lethal-3 gene and investigations of the regulation of dosage compensation in Drosophila. *Development* **121**, 463-75.

Gorman, M., Kuroda, M. I. and Baker, B. S. (1993). Regulation of the sex-specific binding of the maleless dosage compensation protein to the male X chromosome in Drosophila. *Cell* **72**, 39-49.

Grant, P. A., Duggan, L., Cote, J., Roberts, S. M., Brownell, J. E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C. D., Winston, F. *et al.* (1997). Yeast Gcn5 functions in two multisubunit
complexes to acetylate nucleosomal histones: Characterization
of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* **11**, 1640–1650.

Gu, W., Szauter, P. and Lucchesi, J. C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Developmental Genetics* **22**, 56-64.

Gu, W., Wei, X., Pannuti, A. and Lucchesi, J. C. (2000). Targeting the chromatin-remodeling MSL complex of *Drosophila* to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. *EMBO Journal* **19**, 5202-5211.

Hagiwara, T., Hidaka, Y. and Yamada, M. (2005). Deimination of Histone H2A and H4 at Arginine 3 in HL-60 Granulocytes. *Biochemistry* **44**, 5827-34.

Hagiwara, T., Nakashima, K., Hirano, H., Senshu, T. and Yamada, M. (2002). Deimination of arginine residues in nucleophosmin/B23 and histones in HL-60 granulocytes. *Biochem Biophys Res Commun* **290**, 979-83.

Hagstrom, K., Muller, M. and Schedl, P. (1996). *Fab-7* functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. *Genes & Development* **1996**, 3202-3215.

Hamiche, A., Sandaltzopoulos, R., Gdula, D. A. and Wu, C. (1999). ATP-dependent histone octamer sliding mediated by the chromatin remodeling complex NURF. *Cell* **97**, 833-842.

Handler, A. M. and Harrell, R. A., 2nd. (2001). Transformation of the Caribbean fruit fly, Anastrepha suspensa, with a piggyBac vector marked with polyubiquitin-regulated GFP. *Insect Biochem Mol Biol* **31**, 199-205.

Hart, C. M., Cuvier, O. and Laemmli, U. K. (1999). Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. *Chromosoma* **108**, 375-383.

Hart, C. M., Zhao, K. and Laemmli, U. K. (1997). The *scs'* boundary element: Characterisation of boundary element-associated factors. *Molecular & Cellular Biology* **17**, 999-1009.

Hebbes, T. R., Clayton, A. L., Thorne, A. W. and Crane-Robinson, C. (1994). Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO Journal* **13**, 1823-1830.

Henry, R. A., Tews, B., Li, X. and Scott, M. J. (2001). Recruitment of the male-specific lethal (MSL) dosage compensation complex to an autosomally integrated *roX* chromatin entry site correlates with an increased expression of an adjacent reporter gene in male *Drosophila*. *Journal of Biological Chemistry* **276**, 31953-31958.

Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A. and Lucchesi, J. C. (1997). *mof*, a putative actetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO Journal* **16**, 2054-2060.

Hogga, I., Mihaly, J., Barges, S. and Karch, F. (2001). Replacement of *Fab-7* by the *gypsy* or *scs* insulator disrupts long-distance regulatory interactions in the *Abd-B* gene of the Bithorax complex. *Molecular Cell* **8**, 1145-1151.

Huang, A. M. and Rubin, G. M. (2000). A misexpression screen identifies genes that can modulate RAS1 pathway signaling in *Drosophila melanogaster*. *Genetics* **156**, 1219-30.

Inoue, H., Nojima, H. and Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-28.

Ito, T., Bulger, M., Pazin, M. J., Kobayashi, R. and Kadonaga, J. T. (1997). ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* **90**, 145-55.

Ito, T., Levenstein, M. E., Fyodorov, D. V., Kutach, A. K., Kobayashi, R. and Kadonaga, J. T. (1999). ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes & Development* **13**, 1529-1539.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science* **293**, 1074-1080.

Jin, Y., Wang, Y., Johansen, J. and Johansen, K. M. (2000). JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* **149**, 1005-10.

Jin, Y., Wang, Y., Walker, D. L., Dong, H., Conley, C., Johansen, J. and Johansen, K. M. (1999). JIL-1: a novel chromosomal tandem kinase implicated in transcriptional regulation in *Drosophila*. *Molecular Cell* **4**, 129-135.

Jowett, T. (1985). The regulatory domain of a larval serum protein gene in *Drosophila melanogaster*. *EMBO J.* **4**, 3789-3795.

Kadam, S. and Emerson, B. M. (2002). Mechanisms of chromatin assembly and transcription. *Current Opinion in Cell Biology* **14**, 262-268.

Kageyama, Y., Mengus, G., Gilfillan, G., Kennedy, H. G., Stuckenholz, C., Kelley, R. L., Becker, P. B. and Kuroda, M. I. (2001). Association and spreading of the

Drosophila dosage compensation complex from a discrete *roX1* chromatin entry site. *EMBO Journal* **20**, 2236-2245.

Karachentsev, D., Sarma, K., Reinberg, D. and Steward, R. (2005). PR-Set7-dependent methylation of histone H4 Lys 20 functions in repression of gene expression and is essential for mitosis. *Genes Dev* **19**, 431-5.

Kelley, R. L. (2004). Path to equality strewn with *roX*. *Developmental Biology* **269**, 18-25.

Kelley, R. L. and Kuroda, M. I. (2003). The Drosophila roX1 RNA gene can overcome silent chromatin by recruiting the male-specific lethal dosage compensation complex. *Genetics* **164**, 565-74.

Kelley, R. L., Meller, V. H., Gordadze, P. R., Roman, G., Davis, R. L. and Kuroda, M. I. (1999). Epigenetic spreading of the *Drosophila* dosage compensation compex from *roX* RNA genes into flanking chromatin. *Cell* **98**, 513-522.

Kelley, R. L., Solovyeva, I., Lyman, L. M., Richman, R., Solovyev, V. and Kuroda, M. I. (1995). Expression of MSL-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell* **81**, 867-877.

Kellum, R. and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941-950.

Kellum, R. and Schedl, P. (1992). A group of *scs* elements function as domain boundary elements in an enhancer-blocking assay. *Molecular & Cellular Biology* **12**, 2424-2431.

Koonin, E. V., Zhou, S. and Lucchesi, J. C. (1995). The chromo superfamily: new members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucleic Acids Res* **23**, 4229-33.

Korge, G. (1981). Genetic analysis of the larval secretion gene Sgs-4 and its regulatory chromosome sites in *Drosophila melanogaster*. *Chromosoma* **84**, 373-390.

Kornberg, R. D. (1974). Chromatin structure: A repeating unit of histones and DNA. *Science* **184**, 868-871.

Kornberg, R. D. and Thomas, J. O. (1974). Chromatin structure: Oligomers of the histones. *Science* **184**, 865-868.

Kouzarides, T. (2002). Histone methylation in transcriptional control. *Current Opinion in Genetics & Development* **12**, 198-209.

Kuhn, E. J., Hart, C. M. and Geyer, P. K. (2004). Studies of the role of the Drosophila scs and scs' insulators in defining boundaries of a chromosome puff. *Mol Cell Biol* **24**, 1470-80.

Kuroda, M. I., Kernan, M. J., Kreber, R., Ganetzky, B. and Baker, B. S. (1991). The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. *Cell* **66**, 935-947.

Lachner, M. and Jenuwein, T. (2002). The many faces of histone lysine methylation. *Current Opinion in Cell Biology* **14**, 286-298.

Lakhotia, S. C. (1970). Chromosomal basis of dosage compensation in Drosophila. II. DNA replication patterns in an autosome-X insertion in D. melanogaster. *Genetic Research* **15**, 301-307.

Langst, G., Bonte, E. J., Corona, D. F. V. and Becker, P. B. (1999). Nucleosome movement by CHRAC and ISWI without disruption or *trans*-displacement of the histone octamer. *Cell* **97**, 843-852.

Lee, C. G., Chang, K. A., Kuroda, M. I. and Hurwitz, J. (1997). The ATPase/helicase activities of *Drosophila maleless*, an essential factor in dosage compensation. *Embo J* **16**, 2671-81.

Lee, C. G., Reichman, T. W., Baik, T. and Mathews, M. B. (2004). MLE functions as a transcriptional regulator of the roX2 gene. *J Biol Chem* **279**, 47740-5.

Lehmann, M., Siegmund, T., Lintermann, K. and Korge, G. (1998). The *pipsqueak* protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain. *J. Biol. Chem.* **273**, 28504-28509.

Li, C., Geng, C., Leung, H. T., Hong, Y. S., Strong, L. L., Schneuwly, S. and Pak, W. L. (1999). INAF, a protein required for transient receptor potential Ca(2+) channel function. *Proc Natl Acad Sci U S A* **96**, 13474-9.

Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R. and Berger, S. L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked *in vitro* and *in vivo* to Gcn5-mediated acetylation at lysine 14. *Molecular Cell* **5**, 917-926.

Lucchesi, J. C. and Manning, J. E. (1987). Gene dosage compensation in *Drosophila melanogaster*. *Advances in Genetics* **24**, 371-429.

Lucchesi, J. C., Skripsky, T. and Tax, F. E. (1982). A new male-specific lethal mutation in *Drosophila melanogaster*. *Genetics* **100**, s42.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Angstrom resolution. *Nature* **389**, 251-260.

Lyman, L. M., Copps, K., Rastelli, L., Kelley, R. L. and Kuroda, M. I. (1997). Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**, 1743-53.

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H. and MacNeil, T. (1992). Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61-68.

Maroni, G. and Stamey, S. C. (1983). Use of blue food to select synchronous, late third instar larvae. *Drosophila Information Service* **59**, 142-143.

Martinez-Balbas, M. A., Tsukiyama, T., Gdula, D. and Wu, C. (1998). Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proc Natl Acad Sci U S A* **95**, 132-7.

Meller, V. H., Gordadze, P. R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R. L. and Kuroda, M. I. (2000). Ordered assembly of *roX* RNAs into MSL compexes on the dosage-compensated X chromosome in *Drosophila*. *Current Biology* **10**, 136-143.

Meller, V. H. and Rattner, B. P. (2002). The *roX* genes encode redundant *male-specific lethal* transcripts required for targeting of the MSL complex. *EMBO Journal* **21**, 1084-1091.

Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. and Davis, R. L. (1997). *roX1* RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**, 445-457.

Mihaly, J., Hogga, I., Barges, S., Galloni, M., Mishra, R. K., Hagstrom, K., Muller, M., Schedl, P., Sipos, L., Gausz, J. *et al.* (1998). Chromatin domain boundaries in the Bithorax complex. *Cellular & Molecular Life Science* **54**, 60-70.

Miura, K., Wang, S. and Raikhel, A. S. (1999). Two distinct subpopulations of ecdysone receptor complex in the female mosquito during vitellogenesis. *Molecular and Cellular Endocrinology* **156**, 111-120.

Miura, S., Gan, J., Brzostowski, J., Parisi, M. J., Schultz, C. J., Londos, C., Oliver, B. and Kimmel, A. R. (2002). Functional conservation for lipid storage droplet association among Perilipin, ADRP, and TIP47 (PAT)-related proteins in mammals, *Drosophila*, and *Dictyostelium*. *Journal of Biological Chemistry* **227**, 32253-32257.

Mizuguchi, G., Tsukiyama, T., Wisniewski, J. and Wu, C. (1997). Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. *Mol Cell* **1**, 141-50.

Morales, V., Straub, T., Neumann, M. F., Mengus, G., Akhtar, A. and Becker, P. B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *Embo J* **23**, 2258-68.

Morillon, A., Karabetsou, N., Nair, A. and Mellor, J. (2005). Dynamic lysine methylation on histone h3 defines the regulatory phase of gene transcription. *Mol Cell* **18**, 723-34.

Mukherjee, A. S. and Beermann, W. (1965). Synthesis of ribonucleic acid by the X-chromosomes of *Drosophila melanogaster* and the problem of dosage compensation. *Nature* **207**, 785-786.

Muller, H. J., League, B. B. and Offermann, C. A. (1931). Effects of dosage changes of sex-linked genes, and the compensatory effects of other gene differences between male and female. *Anat. Rec. Suppl.* **51**.

Muller-Hermelink, H. K. (1932). Further studies on the nature and causes of gene mutations. *Proc. Int. Congr. Genet* **6**, 213-255.

Muravyova, E., Golovnin, A., Gracheva, E., Parshikov, A., Belenkaya, T., Pirrotta, V. and Georgiev, P. (2001). Loss of insulator activity by paired Su(Hw) chromatin insulators. *Science* **291**, 495-498.

Nakashima, K., Hagiwara, T. and Yamada, M. (2002). Nuclear localization of peptidylarginine deiminase V and histone deimination in granulocytes. *J Biol Chem* **277**, 49562-8.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P. and Reinberg, D. (2002a). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev* **16**, 479-89.

Nishioka, K., Rice, J. C., Sarma, K., Erdjument-Bromage, H., Werner, J., Wang, Y., Chuikov, S., Valenzuela, P., Tempst, P., Steward, R. *et al.* (2002b). PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol Cell* **9**, 1201-13.

Nowak, S. J. and Corces, V. G. (2000). Phosphorylation of histone H3 correlates with transcriptionally active loci. *Genes & Development* **14**, 3003-3013.

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

Offermann, C. A. (1936). Branched chromosomes as symmetrical duplications. *Journal of Genetics* **32**, 103-116.

Oh, H., Bone, J. R. and Kuroda, M. I. (2004). Multiple classes of MSL binding sites target dosage compensation to the X chromosome of Drosophila. *Curr Biol* **14**, 481-7.

Oh, H., Park, Y. and Kuroda, M. I. (2003). Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. *Genes Dev* **17**, 1334-9.

Ota, T., Fukunaga, A., Kawabe, M. and Oishi, K. (1981). Interactions between sex-transformation mutants of *Drosophila melanogaster*. I. Hemolymph vitellogenins and gonad morphology. *Genetics* **81**, 429-441.

Palmer, M. J., Mergner, V. A., Richman, R., Manning, J. E., Kuroda, M. I. and Lucchesi, J. C. (1993). The male-specific lethal-one (msl-1) gene of *Drosophila melanogaster* encodes a novel protein that associates with the X chromosome in males. *Genetics* **134**, 545-57.

Palmer, M. J., Richman, R., Richter, L. and Kuroda, M. I. (1994). Sex-specific regulation of the male-specific lethal-1 dosage compensation gene in Drosophila. *Genes Dev* **8**, 698-706.

Pardue, M. L., Lowenhaupt, K., Rich, A. and Nordheim, A. (1987). (dC-dA)n.(dG-dT)n sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. *Embo J* **6**, 1781-9.

Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I. and Meller, V. H. (2002). Extent of chromatin spreading determined by *roX* RNA recruitment of MSL proteins. *Science* **298**, 1620-1623.

Park, Y., Mengus, G., Bai, X., Kageyama, Y., Meller, V. H., Becker, P. B. and Kuroda, M. I. (2003). Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex. *Mol Cell* **11**, 977-86.

Pham, A. D. and Sauer, F. (2000). Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. *Science* **287**, 501-504.

Qian, S. and Pirrotta, V. (1995). Dosage compensation of the *Drosophila white* gene requires both the X chromosome environment and multiple intragenic elements. *Genetics* **139**, 733-744.

Ramakers, C., Ruijter, J. M., Deprez, R. H. and Moorman, A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* **339**, 62-6.

Rattner, B. P. and Meller, V. H. (2004). Drosophila male-specific lethal 2 protein controls sex-specific expression of the roX genes. *Genetics* **166**, 1825-32.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. *et al.* (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599.

Recillas-Targa, F., Bell, A. C. and Felsenfeld, G. (1999). Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *PNAS* **96**, 14354-14359.

Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M. and Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci U S A* **99**, 6883-8.

Rehm, E. J. (2004). Inverse PCR & cycle sequencing of P element insertions for STS generation, vol. 2004 (ed.: Berkeley Drosophila Genome Project.

Reitman, M. and Felsenfeld, G. (1990). Developmental regulation of Topoisomerase II sites and DNase I hypersensitive sites in the chicken beta-globin locus. *Molecular & Cellular Biology* **10**, 2774-2786.

Rice, J. C., Nishioka, K., Sarma, K., Steward, R., Reinberg, D. and Allis, C. D. (2002). Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev* **16**, 2225-30.

Richter, L., Bone, J. R. and Kuroda, M. I. (1996). RNA-dependent association of the *Drosophila* maleless protein with the male X chromosome. *Genes Cells* **1**, 325-336.

Riggleman, B., Wieschaus, E. and Schedl, P. (1989). Molecular analysis of the *armadillo* locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a *Drosophila* segment polarity gene. *Genes Dev* **3**, 96-113.

Rio, D. C. and Rubin, G. M. (1985). Transformation of cultured *Drosophila melanogaster* cells with a dominant selectable marker. *Mol Cell Biol* **5**, 1833-8.

Roberts, D. B. and Evans-Roberts, S. (1979a). The X-linked alpha-chain gene of *Drosophila* LSP-1 does not show dosage compensation. *Nature* **280**, 691-692.

Roberts, D. B. and Evans-Roberts, S. (1979b). The genetic and cytogenetic localization of the three structural genes coding for the major protein of *Drosophila* larval serum. *Genetics* **93**, 663-679.

Roberts, D. B., Jowett, T., Hughes, J., Smith, D. F. and Glover, D. M. (1991b). The major serum protein of *Drosophila* larvae, larval serum protein 1, is dispensible. *European Journal of Biochemistry* **195**, 195-201.

Roberts, D. B., Turing, J. D. and Loughlin, S. A. R. (1991a). The advantages that accrue to *Drosophila melanogaster* possessing laral serum protein 1. *Journal of Insect Physiology* **37**, 391-400.

Roberts, D. B., Wolfe, J. and Akam, M. E. (1977). The developmental profiles of two major haemolymph proteins from *Drosophila melanogaster*. *Journal of Insect Physiology* **23**, 871-878.

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

Robzyk, K. and Recht, J. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**, 501-504.

Roseman, R. R., Pirrotta, V. and Geyer, P. K. (1993). The su(Hw) protein insulates expression of the *Drosophila melanogaster white* gene from chromosomal position-effects. *EMBO Journal* **12**, 435-442.

Roth, S. Y., Denu, J. M. and Allis, C. D. (2001). Histone acetytransferases. *Annual Review of Biochemistry* **70**, 81-120.

Rubin, G. M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M. and Harvey, D. A. (2000). A *Drosophila* complementary DNA resource. *Science* **287**, 2222-2224.

Russo, C. A., Takezaki, N. and Nei, M. (1995). Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* **12**, 391-404.

Sachs, L. M. and Shi, Y. (2000). Targeted chromatin binding and histone acetylation *in vivo* by thyroid hormone receptor during amphibian development. *PNAS* **97**, 13138-13143.

Saitoh, N., Bell, A. C., Recillas-Targa, F., West, A. G., Simpson, M., Pikaart, M. and Felsenfeld, G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO Journal* **19**, 2315-2322.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Sass, G. L., Pannuti, A. and Lucchesi, J. C. (2003). Male-specific lethal complex of Drosophila targets activated regions of the X chromosome for chromatin remodeling. *Proc Natl Acad Sci U S A* **100**, 8287-91.

Sass, H. and Meselson, M. (1991). Dosage compensation of the Drosophila pseudoobscura Hsp82 gene and the *Drosophila melanogaster* Adh gene at ectopic sites in D. melanogaster. *Proc Natl Acad Sci U S A* **88**, 6795-9.

Scott, K. C., Taubmana, A. D. and Geyera, P. K. (1999). Enhancer blocking by the *Drosophila* gypsy insulator depends upon insulator anatomy and enhancer strength. *Genetics* **153**, 787-798.

Scott, M. J. and Lucchesi, J. C. (1991). Structure and expression of the *Drosophila melanogaster* gene encoding 6-phosphogluconate dehydrogenase. *Gene* **109**, 177-183.

Scott, M. J., Pan, L. L., Cleland, S. B., Knox, A. L. and Heinrich, J. (2000). MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in *Drosophila*. *EMBO Journal* **19**, 144-155.

Seecof, R. L., Kaplan, W. D. and Futch, D. G. (1969). Dosage compensation for enzyme activities in *Drosophila melanogaster*. *PNAS* **62**, 528-535.

Shermoen, A. W. and Beckendorf, S. K. (1982). A complex of interacting DNAase I-hypersensitive sites near the Drosophila glue protein gene, Sgs4. *Cell* **29**, 601-7.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A. and Casero, R. A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-53.

Sims, R. J., 3rd, Belotserkovskaya, R. and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev* **18**, 2437-68.

Smith, D. F., McClelland, A., White, B. N., Addison, C. F. and Glover, D. M. (1981). The molecular cloning of a dispersed set of developmentally regulated genes which encode the major larval serum protein of *D. melanogaster*. *Cell* **23**, 441-449.

Smith, E. R., Allis, C. D. and Lucchesi, J. C. (2001). Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *Journal of Biological Chemistry* **276**, 1483-1486.

Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D. and Lucchesi, J. C. (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol Cell Biol* **20**, 312-8.

Smith, P. D. and Lucchesi, J. (1969). The role of sexuality in dosage compensation in Drosophila. *Genetics* **61**, 607-618.

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

Stuckenholz, C., Meller, V. H. and Kuroda, M. I. (2003). Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in *Drosophila melanogaster*. *Genetics* **164**, 1003-14.

Teixeira, L., Rabouille, C., Rorth, P., Ephrussi, A. and Vanzo, N. F. (2003). *Drosophila* Perilipin/ADRP homologue Lsd2 regulates lipid metabolism. *Mechanisms of Development* **120**, 1071-1081.

Thummel, C. S., Boulet, A. M. and Lipshitz, H. D. (1988). Vectors for *Drosophila* P-element-mediated transformation and tissue culture transformation. *Gene* **74**, 445-456.

Tsukiyama, T., Daniel, C., Tamkun, J. and Wu, C. (1995). ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* **83**, 1021-6.

Tsukiyama, T. and Wu, C. (1995). Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**, 1011-20.

Turner, B. M., Birley, A. J. and Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* **69**, 375-384.

Uchida, S., Uenoyama, T. and Oishi, K. (1981). Studies on the sex-specific lethals of *Drosophila melanogaster*. III. A third chromosome male-specific mutant. *Japanese Journal of Genetics* **56**, 523-527.

Udvardy, A., Maine, E. and Schedl, P. (1985). The 87A7 chromomere: Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *Journal of Molecular Biology* **185**, 341-358.

Varga-Weisz, P. D., Wilm, M., Bonte, E., Dumas, K., Mann, M. and Becker, P. B. (1997). Chromatin-remodeling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* **388**, 598-602.

Vazquez, J. and Schedl, P. (1994). Sequences required for enhancer blocking activity of *scs* are located within two nuclease-hypersensitive regions. *EMBO Journal* **13**, 5984-5993.

Vincent, J. P., Girdham, C. H. and O'Farrell, P. H. (1994). A cell-autonomous, ubiquitous marker for the analysis of Drosophila genetic mosaics. *Dev Biol* **164**, 328-31.

Wang, Y., Zhang, W., Jin, Y., Johansen, J. and Johansen, K. M. (2001). The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in *Drosophila*. *Cell* **105**, 433-443.

West, A. G., Huang, S., Gaszner, M., Litt, M. D. and Felsenfeld, G. (2004). Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol Cell* **16**, 453-63.

Williamson, J. H., Krochko, D. and Geer, B. W. (1980). 6-phosphogluconate dehydrogenase from *Drosophila melanogaster*. I. Purification and properties of the A isozyme. *Biochemical Genetics* **18**, 87-101.

Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., Roeder, R. G., Brivanlou, A. H. and Allis, C. D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859-872.

Xiao, H., Sandaltzopoulos, R., Wang, H. M., Hamiche, A., Ranallo, R., Lee, K. M., Fu, D. and Wu, C. (2001). Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* **8**, 531-43.

Xu, F., Zhang, K. and Grunstein, M. (2005). Acetylation in histone h3 globular domain regulates gene expression in yeast. *Cell* **121**, 375-85.

Yang, X. H., Seow, K. T., Bahri, S. M., Oon, S. H. and Chia, W. (1991). Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a subset of developing axons and pioneer neurons in the embryonic CNS. *Cell* **67**, 661-73.

Zhang, W., Wang, Y., Long, J., Girton, J., Johansen, J. and Johansen, K. M. (2003). A developmentally regulated splice variant from the complex lola locus encoding multiple different zinc finger domain proteins interacts with the chromosomal kinase JIL-1. *J Biol Chem* **278**, 11696-704.

Zhang, Y. (2003). Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev* **17**, 2733-2740.

Zhao, K., Hart, C. M. and Laemmli, U. K. (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* **81**, 879-889.

Zhou, J., Barolo, S., Szymanski, P. and Levine, M. (1996). The *Fab-7* element of the bithorax complex attenuates enhancer-promoter interactions in the *Drosophila* embryo. *Genes & Development* **10**, 3195-3201.

Zhou, J. and Levine, M. (1999). A novel *cis*-regulatory element, the PTS, mediates an anti-insulator activity in the *Drsoophila* embryo. *Cell* **99**, 567-575.

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. ***et al.*** (1995). Male-specific lethal 2, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO Journal* **14**, 2884-2895.

10 APPENDICES

10.1 Primer sequences

These primers were used for all PCR-based analyses of these genes or genomic regions.

Gene/ Region	Primer	Primer sequence	Size product amplifi d (bp)
Lsp1α	Forw-Lsp1α	5' ATCGCATTCCTGGCTTGTGT	2344
	Rev-Lsp1α	5' ACTTGGTGTAGTAGTCGTTG	
Lsp1α promoter	Lsp1pForw	5' GCAACAAACGTATGAAATTC	593
	Lsp1pRev	5' GATCCAATGCCTAAAGACTG	
Lsp1α 3' UTR	Lsp1αfor3'reg	5' GCCACAGCACGAGCAGTTCT	575
	Lsp1αrev3'reg	5' TTCGCATTTTTTGTAAGGGG	
CG2560-Lsp1α genomic region	InsulForw	5' CGATCAAAGTTTTAATGTGTTTTAAT	883
	Lsp1pRev	5' GATCCAATGCCTAAAGACTG	
Lsp1α-CG2556 genomic region	Lsp1aCG2556F	5' CTGATAACATTCCGATTTTGACC	4596
	Lsp1aCG2556R#2	5' AAAAATGAAACGATAGTAAATGT	
CG2560-CG2556 genomic region	CG2560ORFF	5' TCCTTCTGGGTTTCGCTGCTCTGG	9413
	CG25565'UTRR	5' CTTTTCTCTGTTTATCGGCGGTTGCG	
CG15730	AT20771F	5' AAGCGGAACTGGAGCAATA	1478
	AT20771R	5' CCACAAAGGCGGCCATTAC	
CG2556 exon 2 – exon 4	CG2556E2F	5' TATGTCCCCGTTGAGCAGAAGATT	636 cDNA 820 ger
	CG2556E4R	5' AAGAGGGAGAAGAAGGAGAAGGAC	

Lsp1α promoter – mutagenesis of putative Dsx binding site	Lsp1pdsxmutRev Lsp1pdsxmutForw	5' AAGTGTGTTTCCCGTGGCGTATGAGC 5' GCCACGGGAAACACACTTTTCTCAA	
ChIP negative control in ORF of *Gpdh*. From (Smith *et al.*, 2001)	GpdhF GpdhR	5' GTGCCCGACCTGGTTGAG 5' CTTGCCTTCAGGTGACGC	200
ChIP test primers in ORF and 3' UTR of *Lsp1α*	Lsp1a1884F Lsp1a2085R	5' CTCGCTGACGGACAAC 5' GGGCTCAGTAAGGTCCA	195
	Lsp1a-2439F Lsp1a-2559R	5' CTTCAAGTTGTAGATCTGATAACATT CCGA 5' GCTTAACAGAGAAATCTGAATACGT TGG	202
ChIP positive control in ORF and 3' UTR of *Pgd*. From (Smith *et al.*, 2001)	pgdf5 pgdr5	5' GGATAAGCAGGTGGAAGTAGGAAG 5' ACACTTGTGGTTACGGTTTTCG	121
	pgdf10 pgdr10	5' GAAGGGCACGGGCAAGTG 5' CAATGCCGCCGTAATTAAGTCTC	200
ChIP positive control in ORF and 3' UTR of *Lsd-2*.	Lsd2-1490F Lsd2-1674R	5' AGTGTACTAGCCGATACG 5' TCTGACTCCCGGATCT	185
	Lsd2-1813F Lsd2-1921R	5' GAAACACACGCACACGG 5' TCCCAGCGAGCGTACAA	109

ChIP test primers in ORF of *lacZ*	lacZ-2823F lacZ-2950R	5' GCGCGAATTGAATTATGGCCC 5' GCCATGTGCCTTCTTCCG	128
ChIP test primers in 5' UTR and 3' UTR of *CG2556*	CG25565'UTRF CG25565'UTRR	5' GATTGCTCTGCGTGCTGGTCGTTC 5' CTTTTCTCTGTTTATCGGCGGTTGCG	117
	CG25563'UTRF CG25563'UTRR	5' AGTGTCCACACTAACACACTTG 5' CAATTCCCATTGATTGGCATTAAGAT	150
Real-time RT-PCR in exon 2 of *rp49*	rp49-596F rp49-720R	5' CGGTTACGGATCGAACA 5' CGATCTCGCCGCAGTAAA	125
Real-time RT-PCR in exon 4 of *CG2560*	CG2560-696F CG2560-875R	5' ATGGCAATGCTTACGGT 5' GAGGTGGCTGATAATCGTAG	180
Real-time RT-PCR in exon 2 of *CG2556*	CG2556-1129F CG2556-1229R	5' TGGTAATGGCGGCCTAAA 5' TGCGAGTGTTCAGCTTG	101

10.2 Inverse PCR primers

These primers were used for inverse PCR conducted on transgenic fly lines to determine the site of transgene integration (Rehm, 2004).

Gene/Region	Primer	Primer sequence
PEP *P*-element 5' end	Plac1	5' CACCCAAGGCTCTGCTCCCACAAT
	Pwht1	5' GTAACGCTAATCACTCCGAACAGGTCACA
PEP *P*-element 3' end	Pry4	5' CAATCATATCGCTGTCTCACTCA
	Pry1	5' CCTTAGCATGTCCGTGGGGTTTGAAT
Sequencing of products from the PEP *P*-element 5' end	Sp1	5' ACACAACCTTTCCTCTCAACAA
Sequencing of products from the PEP *P*-element 3' end	Spep1	5' GACACTCAGAATACTATTC

10.3 Sequencing primers

Purpose	Primer name	Sequence
sequencing of inserts in *Stu*I/*Eco*RI sites 5' of *lacZ-SV40* in pCaSpeR-arm-βgal based plasmids	arm-lacZseq	5' CAGAGAAGGAGGCAAACAGC
sequencing of inserts in *Not*I site 3' of *lacZ-SV40* in pCaSpeR-arm-βgal based plasmids	armlacZNotIseq	5' AATAAAGCAATAGCATCACA
sequencing of region where *arm-lacZ* construct has integrated. Reverse from beginning *arm* sequence.	armlacZStuIRevseq	5' CGTAGAAGTGGGTTCATCGGT
sequencing of *GMR-3xP3-DsRed-hsp70polyA* construct. Reverse from middle *3xP3* sequence.	GMR-DsRedFseq	5' CGCATGGAAGGAACGGTCAAT
sequencing of region where *GMR-3xP3-DsRed-hsp70polyA* construct has integrated. Reverse from *hsp70polyA* sequence.	GMR-DsRedFseq2	5' ACTGATAAGAATGTTTCGATCG
Sequencing of homologous recombination constructs	Lsp1a1020F Lsp1a1468R	5' CAAGACTTTTGGCCTGC 5' AGTGGGTATCGGCGAA
Sequencing of FRT and I-SceI sites within pW30 constructs	pW30-545F pW30-1276R	5' TTTTGTGGAAGCGGTATTCG 5' GCGACAGAGTGAGAGAGCAA

10.4 Oligonucleotide pairs annealed as linkers in cloning

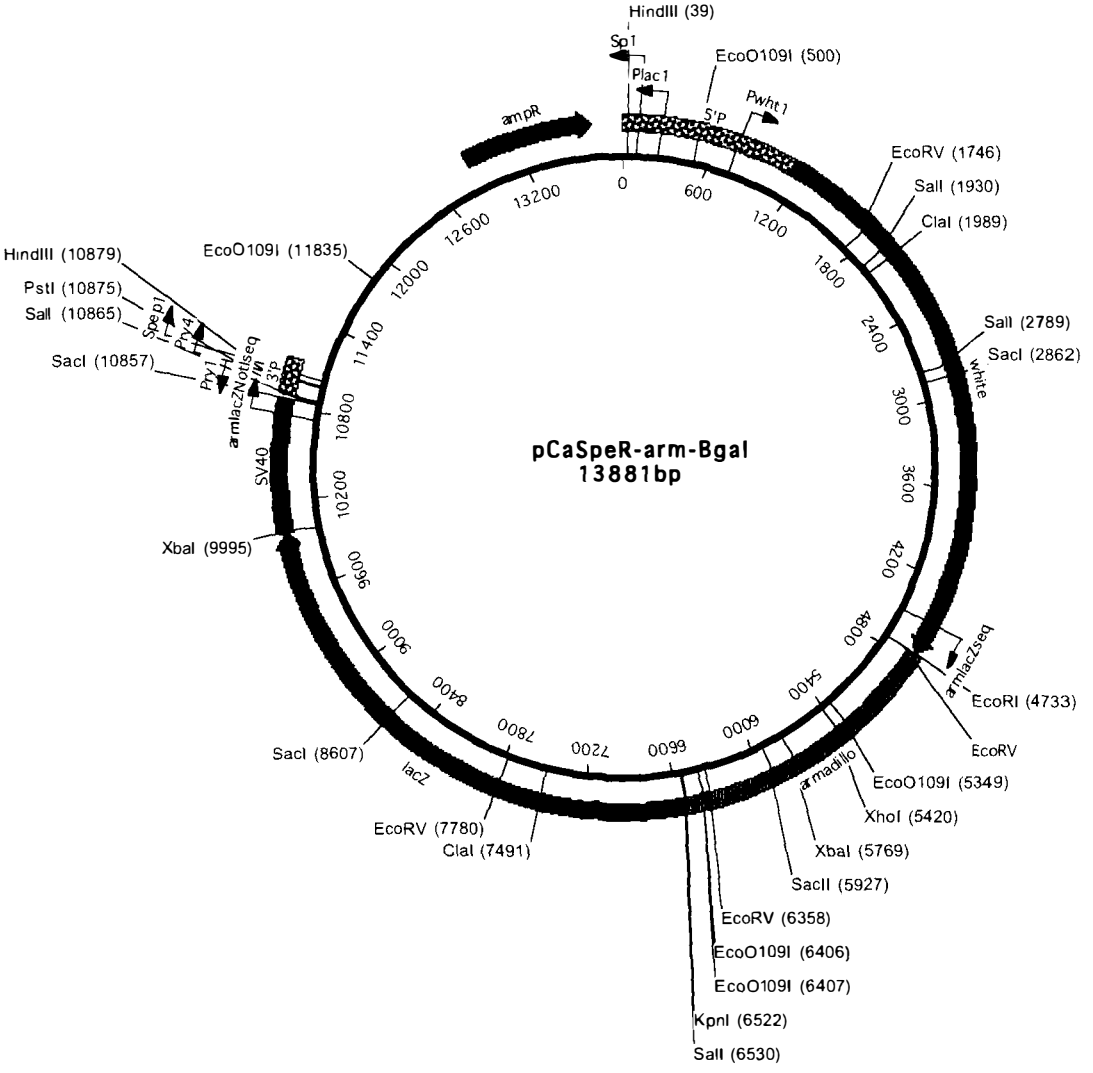
Italics indicate cohesive end only. *Underlined Italics* indicates where the restriction site was not regenerated following ligation into the vector.

Restriction sites	oligo name	oligonucleotide sequence (5' – 3')
*Asp*718 – *Bam*HI – *Bgl*II – *Nhe*I – *Not*I – *Pst*I	Asp718-PstIup	GTACCGGATCCAGATCTGCGGCCGCGCTA GCCTGCA
	Asp718-PstIlow	GGCTAGCGCGGCCGCAGATCTGGATCCG
PstI - *Not*I – *Pst*I	PstI-NotIlinkup	AGCGGCCGCAGCTTACACTGCA
	PstI-NotIlinklow	GTGTAAGCTGCGGCCGCTTGCA
EcoRI - *Stu*I – *Eco*RI	*EcoR*I-StuIlinkup	AATTGAGAGAAGGCCTAGCGTG
	*EcoR*I-StuIlinklow	AATTCACGCTAGGCCTTCTCTC
*Eco*RI - *Bam*HI - *Nhe*I - *Stu*I - *Asp*718	EcoAsplinkup	AATTCGGATCCGCTAGCAGGCCTG
	EcoAsplinklow	GTACCAGGCCTGCTAGCGGATCCG

10.5 Plasmid maps

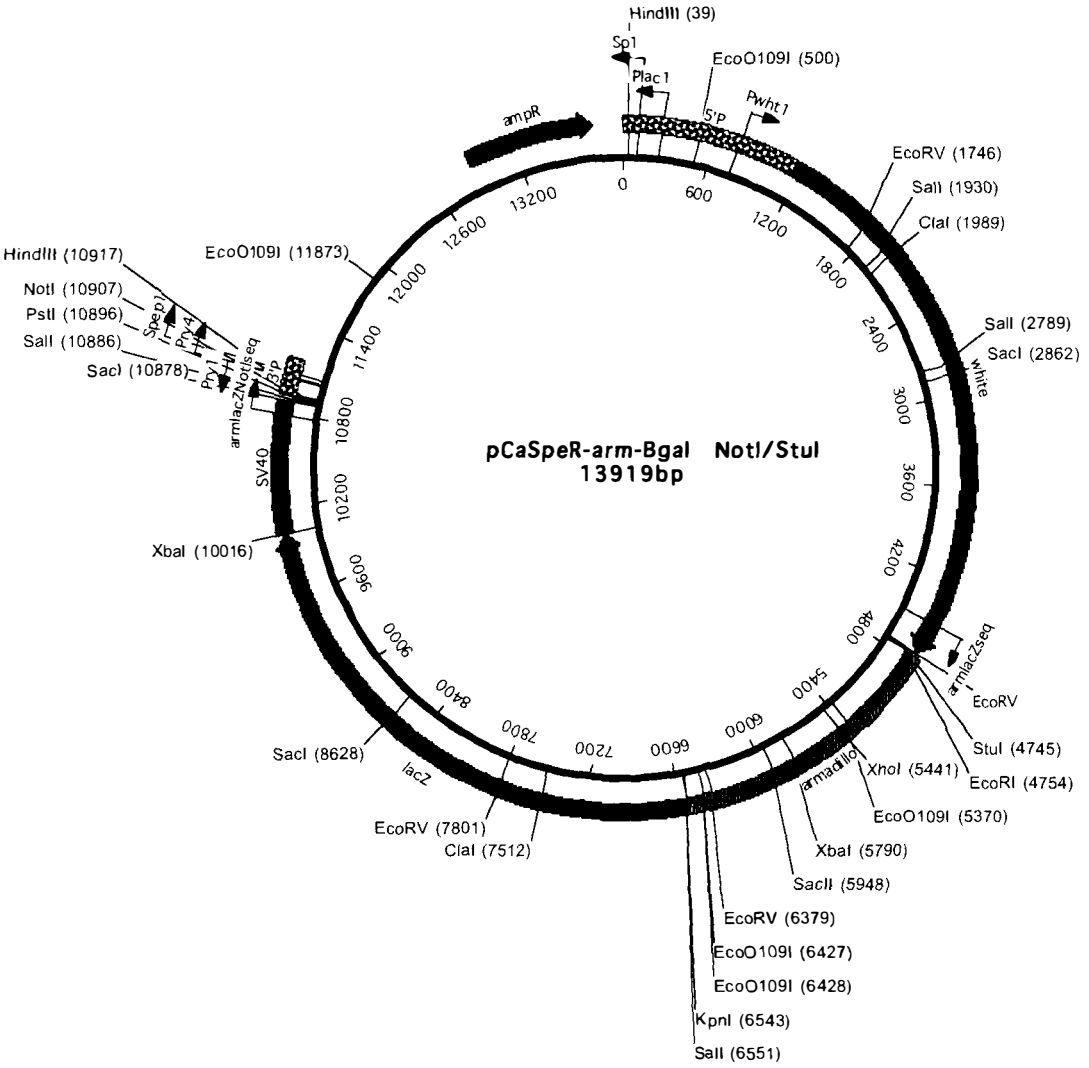
10.5.1 pCaSpeR-arm-βgal (13881 bp)

This vector is based on pCaSpeR-AUG-βgaI (Thummel *et al.*, 1988). It consists of a pCaSpeR (pUC, *mini-white*, *P*-element 5' and 3' ends) backbone containing the 1.7 kb *armadillo* promoter inserted into the *Eco*RI and *Kpn*I sites of the MCS to drive expression of the 4.4 kb *lacZ* reporter gene with the SV40 3' UTR (Fitzsimons *et al.*, 1999; Vincent *et al.*, 1994).



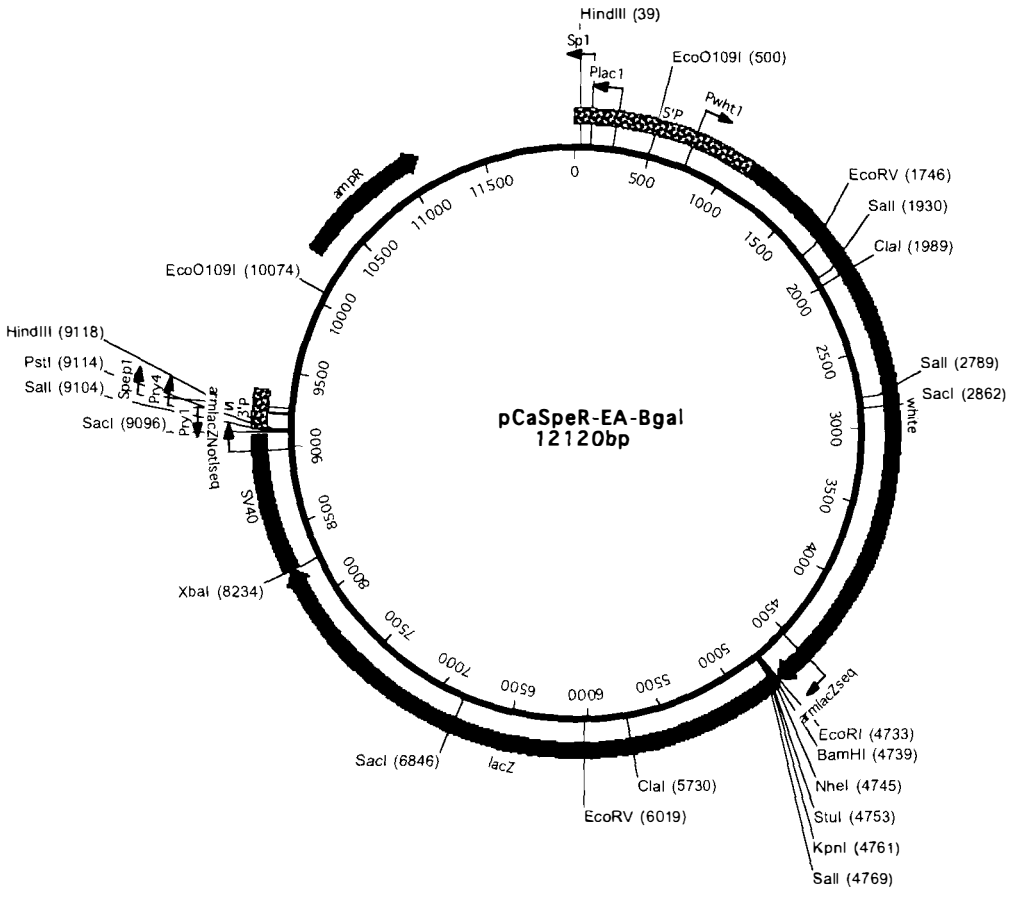
10.5.2 pCaSpeR-arm-βgal *Stu*I/*Not*I (13919 bp)

This vector is based on pCaSpeR-arm-βgal (section 10.5.1). A linker containing an extra *Stu*I site has been inserted in the *Eco*RI site (section 10.4). A linker containing an extra *Not*I site has been inserted in the *Pst*I site (section 10.4). The *Not*I and *Stu*I restriction endonuclease sites provide insertion points for putative insulator elements cloned using PCR directly into pGEM®-T Easy (Promega).



10.5.3 pCaSpeR-EA-βgal (12120 bp)

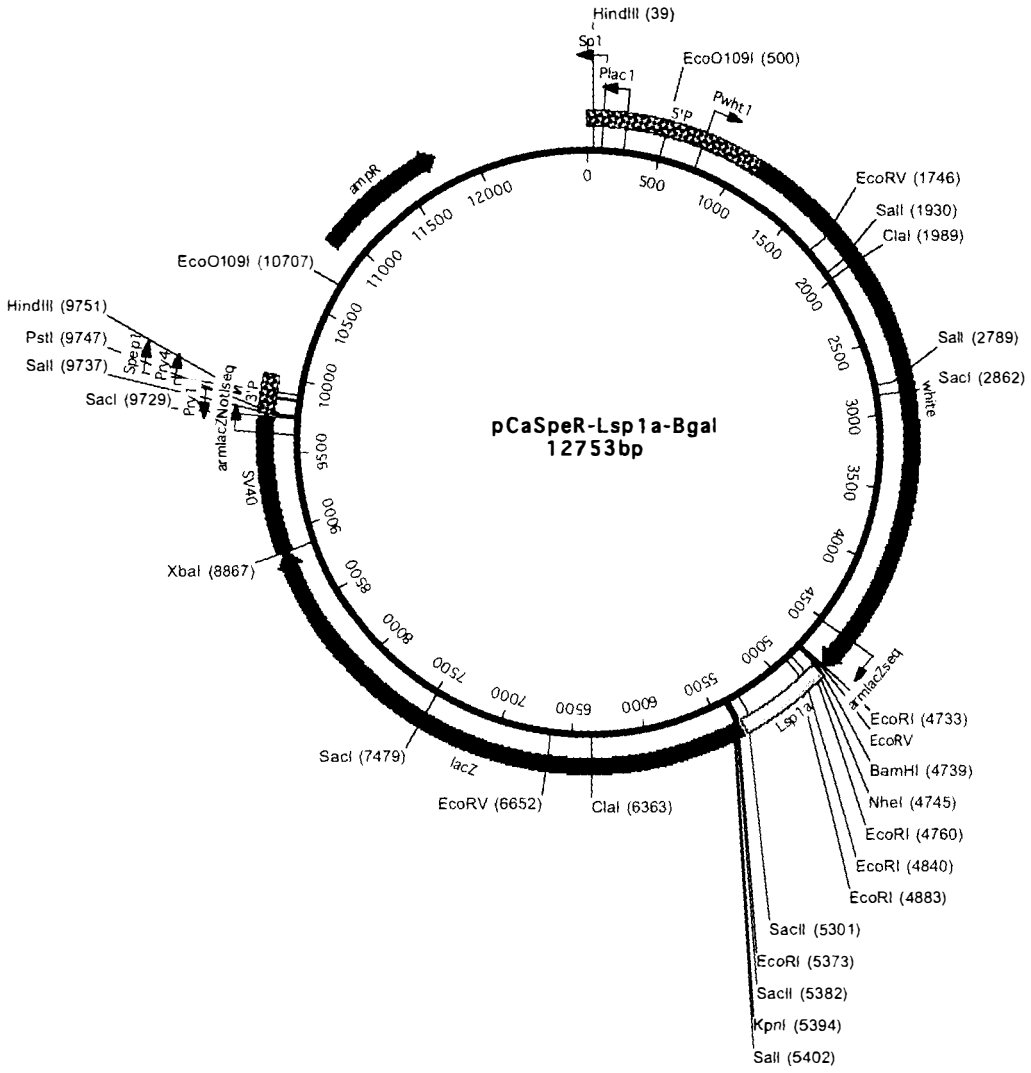
This vector is based on pCaSpeR-arm-βgal. The *armadillo* promoter has been replaced with an *Eco*RI – *Kpn*I linker (section 10.4). This provides suitable restriction sites for the insertion of the full and extended *Lsp1α* promoter fragments



.

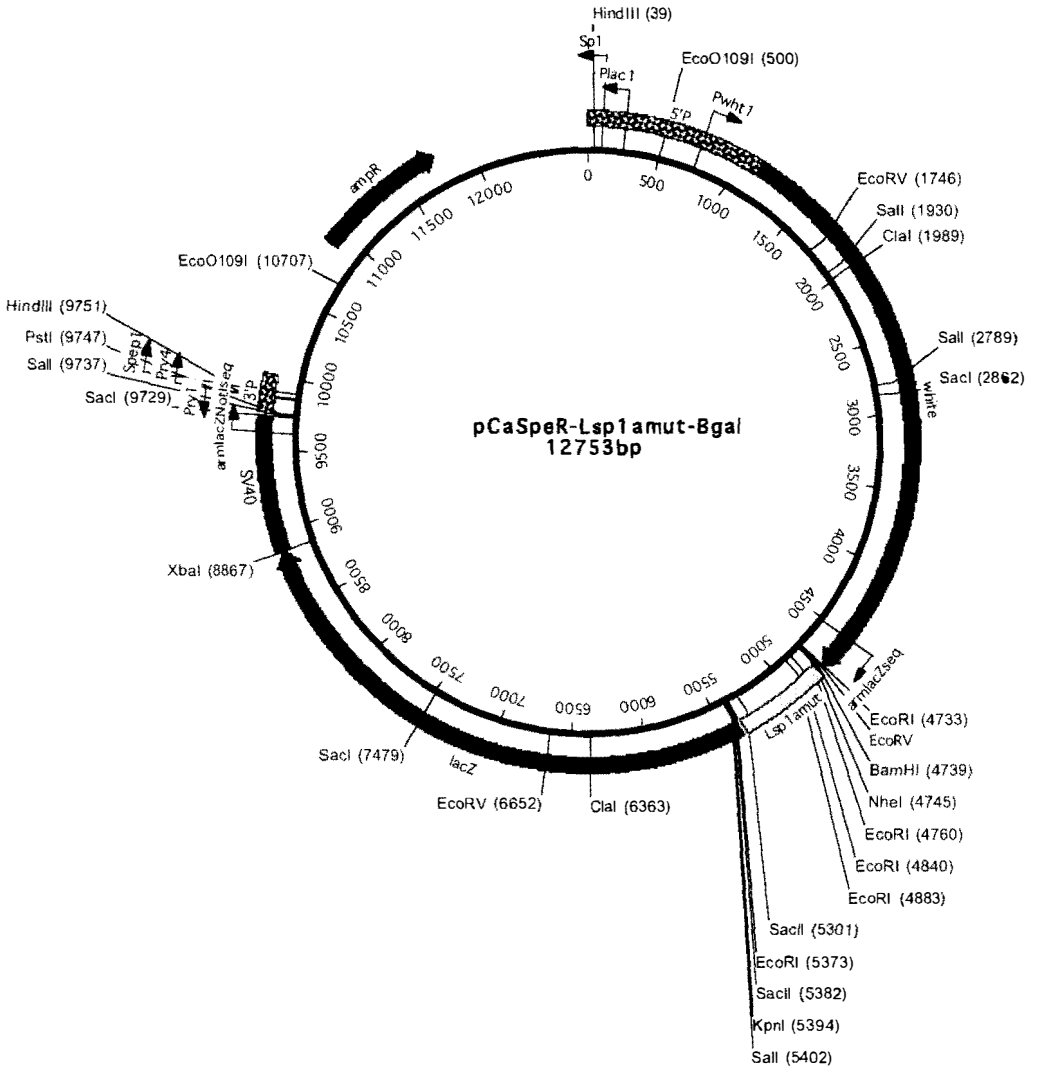
10.5.4 pCaSpeR-Lsp1α-βgal (12753 bp)

This vector is based on pCaSpeR-βgal (section 10.5.3). The 593 bp *Lsp1α* promoter (Lsp1α/P) is inserted in the *Stu*I site of pCaSpeR-βgal and controls expression of the *lacZ* reporter gene.



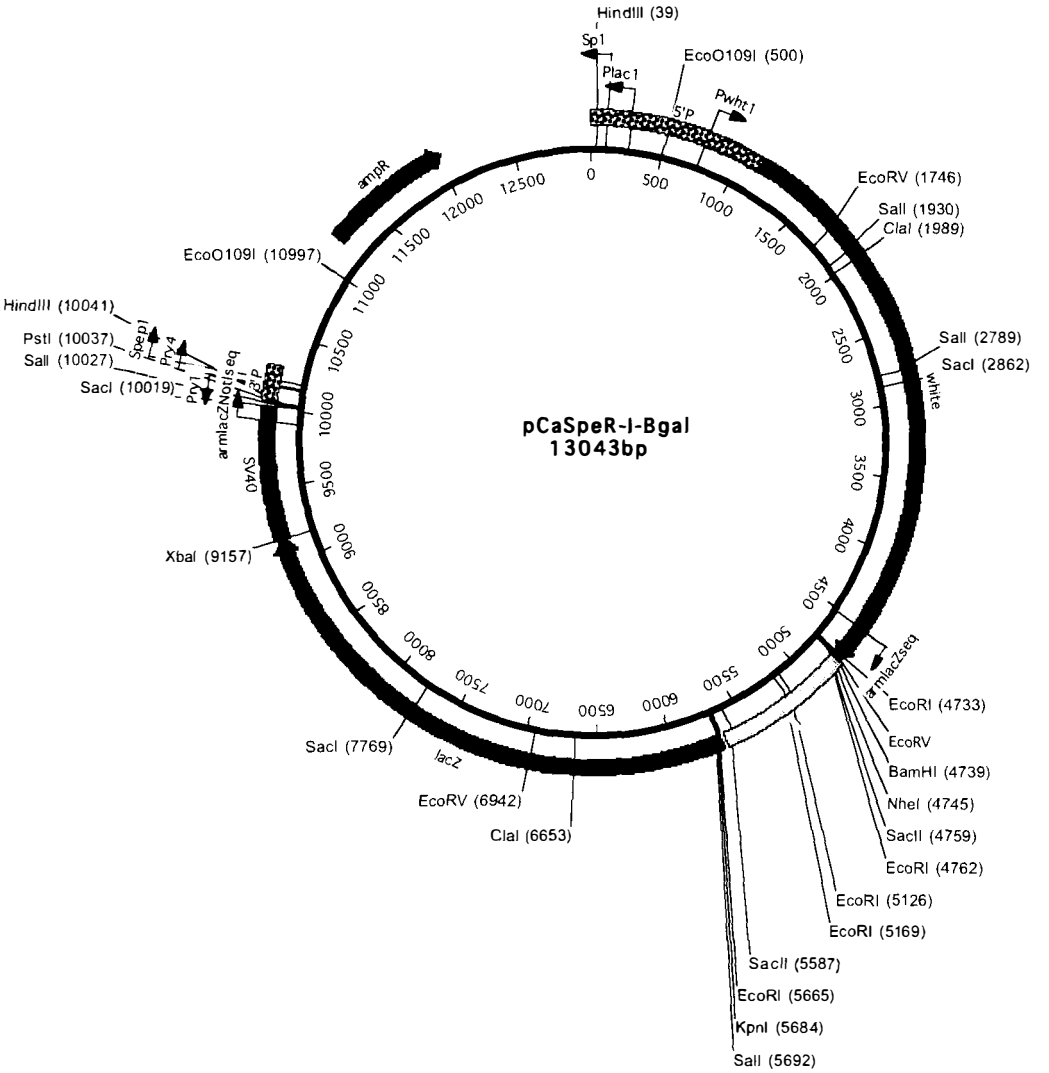
10.5.5 pCaSpeR-Lsp1αmut–βgal (12753 bp)

This vector is based on pCaSpeR-βgal (section 10.5.3). The 593 bp mutant *Lsp1α* promoter (Lsp1αmut/P*) is inserted in the *StuI* site of pCaSpeR-βgal and controls expression of the *lacZ* reporter gene.



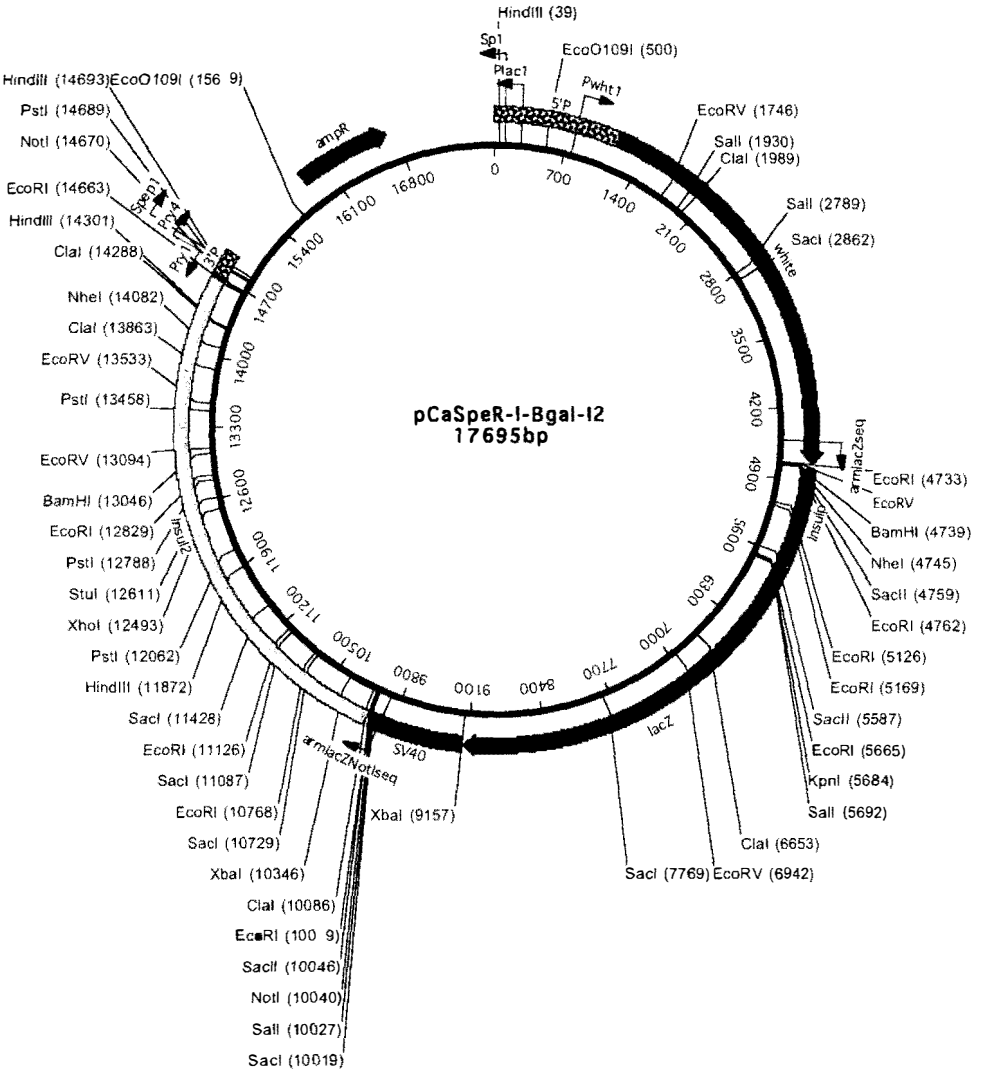
10.5.6 pCaSpeR-I-βgal (13043 bp)

This vector is based on pCaSpeR-βgal (section 10.5.3). The 883 bp genomic region between *CG2560* and *Lsp1α*, including the *Lsp1α* promoter (I) is inserted in the *Stu*I site of pCaSpeR-βgal and controls expression of the *lacZ* reporter gene.



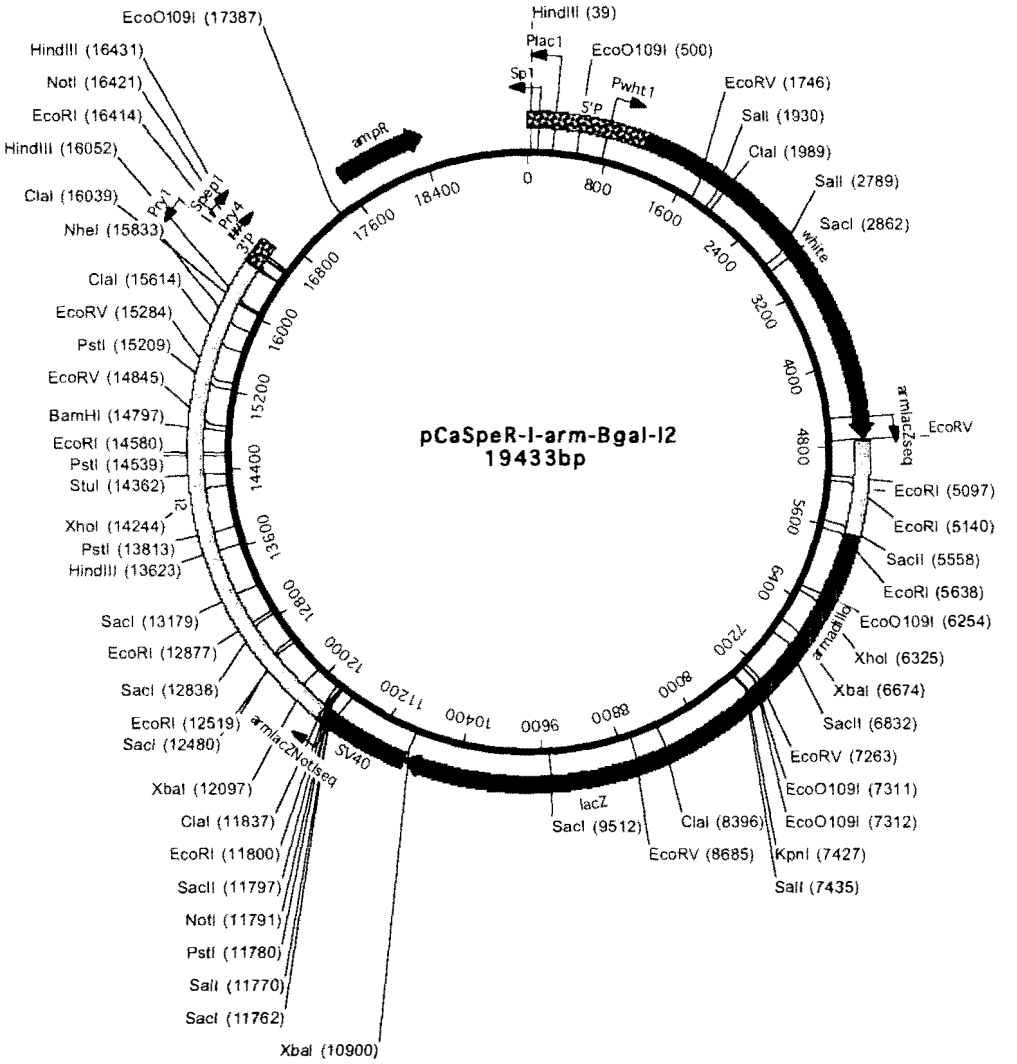
10.5.7 pCaSpeR-I-βgal-I2 (17695 bp)

This vector is based on pCaSpeR-βgal (section 10.5.3). The 883 bp genomic region between *CG2560* and *Lsp1α*, including the *Lsp1α* promoter (I) is inserted in the *Stu*I site of pCaSpeR-βgal and controls expression of the *lacZ* reporter gene. The 4596 bp genomic region between *Lsp1α* and *CG2556* (I2) is inserted in the *Not*I site 3' of *arm-lacZ*.



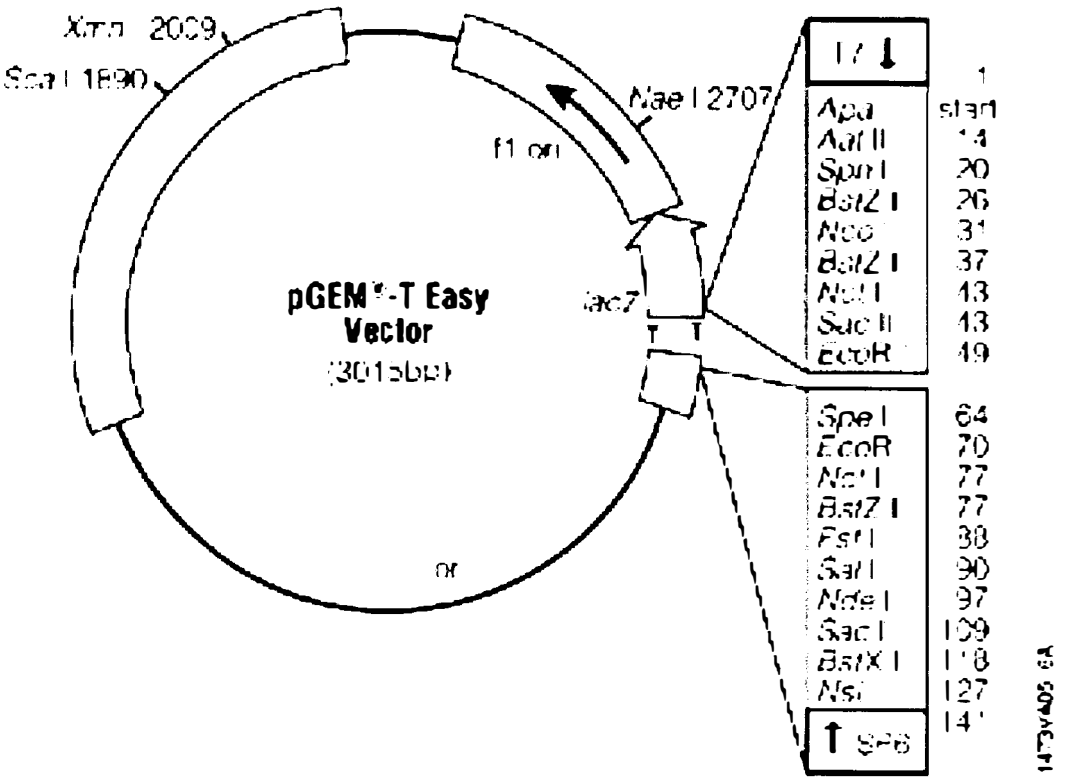
10.5.8 pCaSpeR-I-arm-βgal-I2 (19433 bp)

This vector is based on pCaSpeR-arm-βgal *Stu*I/*Not*I (section 10.5.2). The 883 bp genomic region between *CG2560* and *Lsp1α* (I) is inserted in the *Stu*I site 5' of *arm-lacZ*. The 4596 bp genomic region between *Lsp1α* and *CG2556* (I2) is inserted in the *Not*I site 3' of *arm-lacZ*.

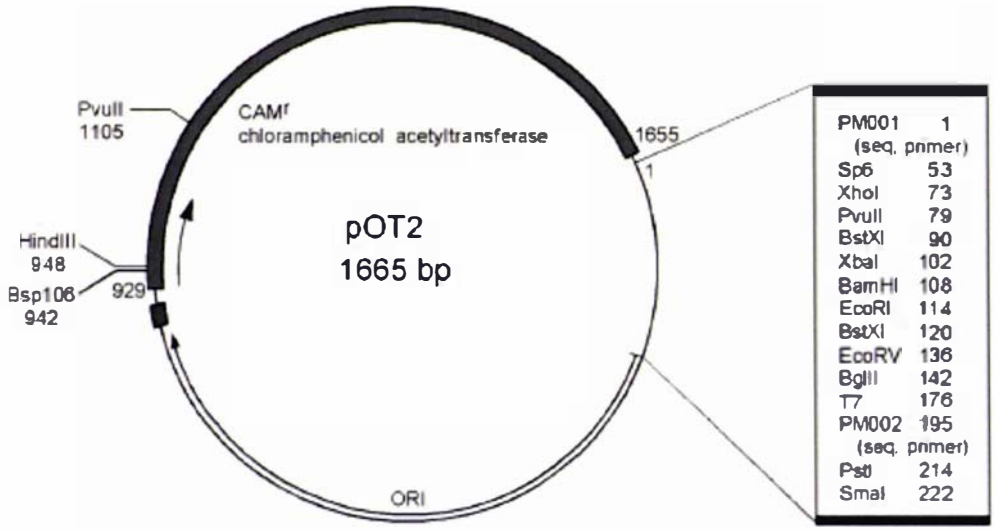


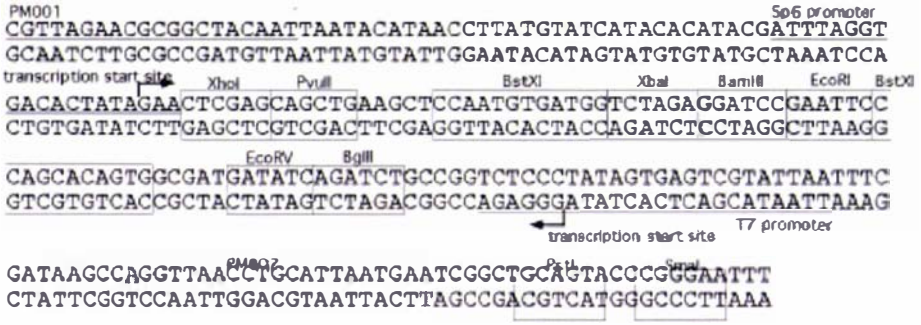
10.5.9 pGEM®-T Easy (3015 bp)

This vector was used for cloning of PCR products (Promega).



10.5.10 pOT2 (1665 bp)





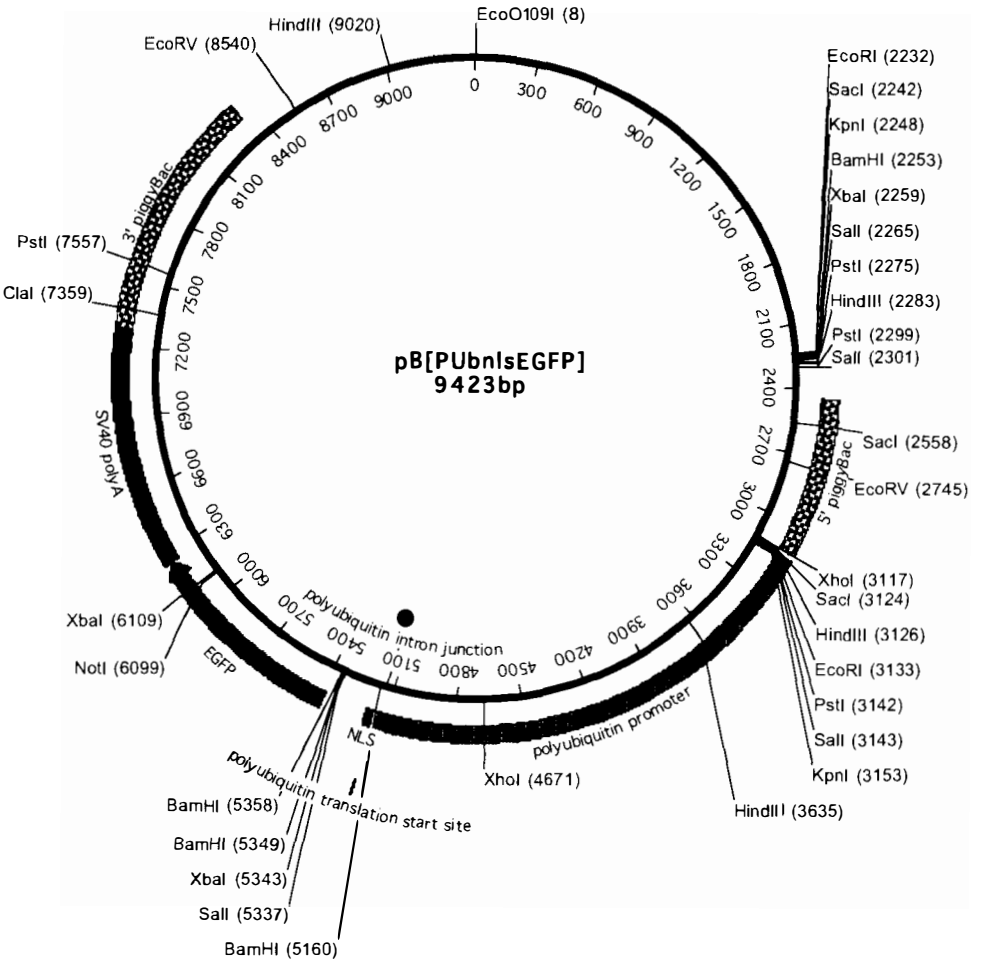
The original plasmid DNA was provided by M. Palazzolo. The plasmid was sequenced by Martha Evans-Holm and Damon Harvey of the Gerald M. Rubin Laboratory at the Howard Hughes Medical Institute, University of California, Berkeley.

The pOT2 vector was used in over 90% of the D. melanogaster cDNA libraries constructed for the BDGP EST sequencing project. The cDNAs were directionally cloned into the 5' EcoRI site and the 3' XhoI site. The 5' ESTs were sequenced using the T7 sequencing primer.

cutters: BamHI, BglII, Bsp106, BstXI, EcoRI, EcoRV, HindIII, PstI, PvuII, SmaI, XbaI, XhoI
non-cutters: ApaI, DraII, KpnI, NotI, SacI, SpeI, XmaIII

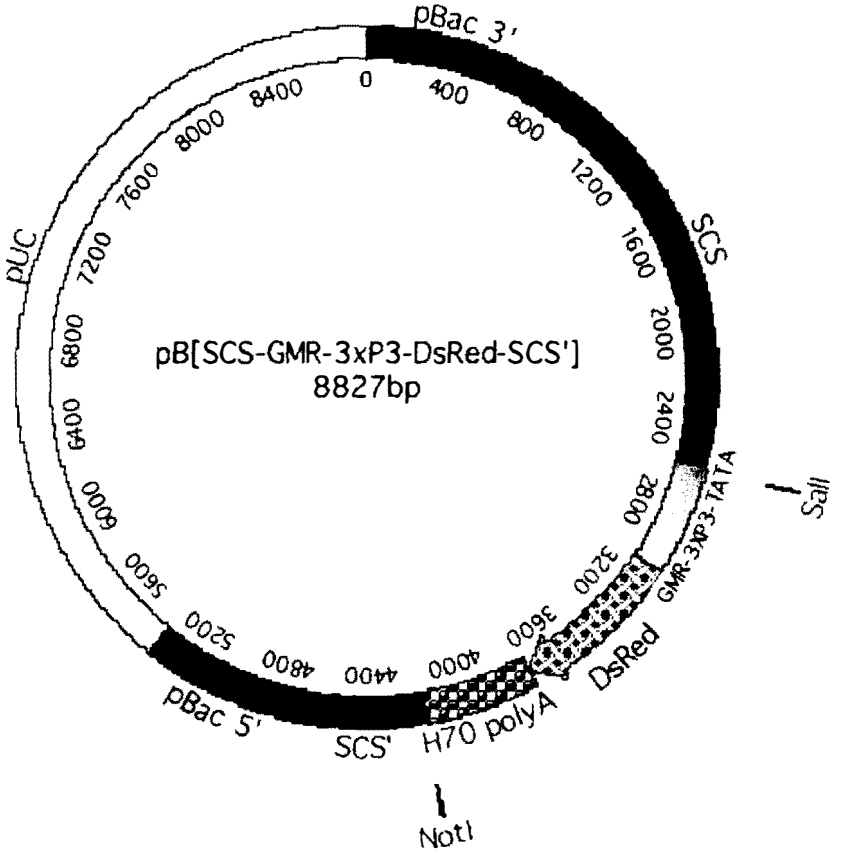
10.5.11 pB[PUbnlsEGFP] (9423 bp)

This plasmid was provided by Al Handler and consists of the 4121 bp *Pub-nls-EGFP-SV40* reporter construct flanked by piggyBack transposon sequences in a pUC vector (Handler and Harrell, 2001).



10.5.12 pB[SCS-GMR-3xP3-DsRed-SCS'] (8827 bp)

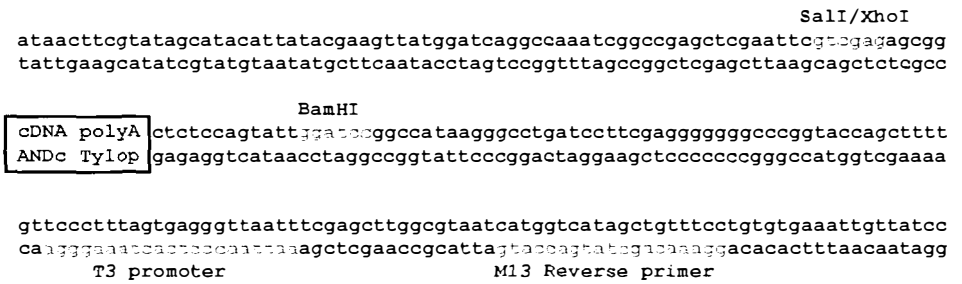
This plasmid was provided by Asela Attapatu (Massey University, 2004) and consists of the *GMR-3xP3-DsRed* reporter gene flanked by *SCS/SCS'* insulators in a PUC-based piggyBac vector.



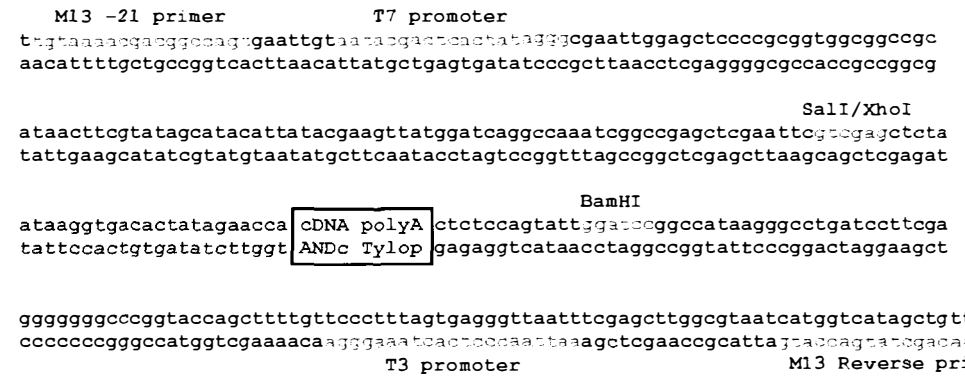
10.5.13 pFLC-1 (3005 bp)

This vector contains the oligo-dT primed cDNA clones integrated in the sense direction from the T7 polymerase binding site (Rubin *et al.*, 2000). The pFLC-I vector is a derivative of the ampicillin-resistant plasmid pBlueScriptII-SK(+). cDNA sequences were inserted into vector with the *Xho*I site at the 5' end of the cDNA and the *Bam*HI site at the 3' end.

pFLC-I Polylinker for RE clones

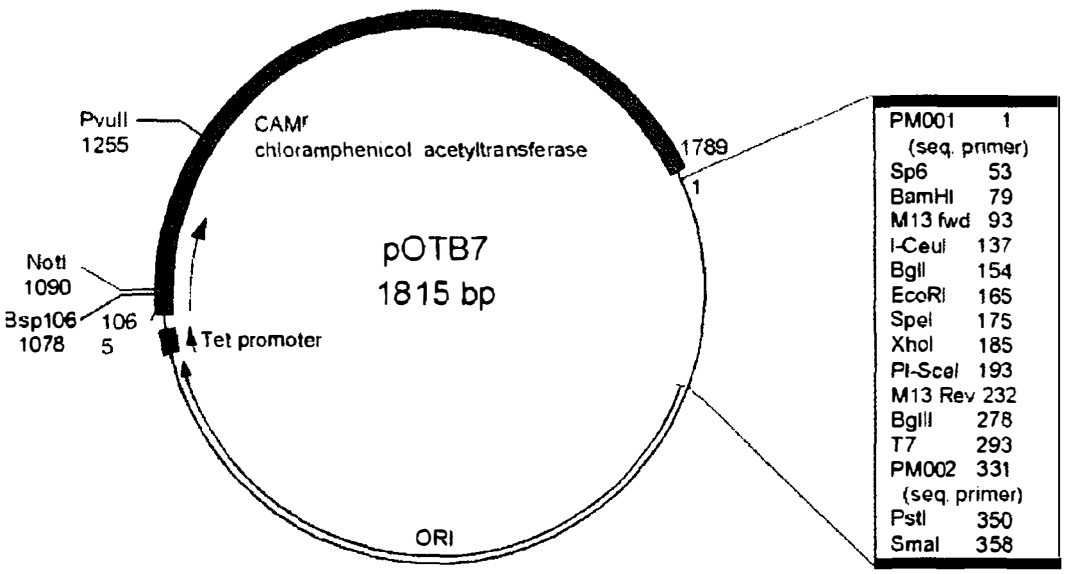


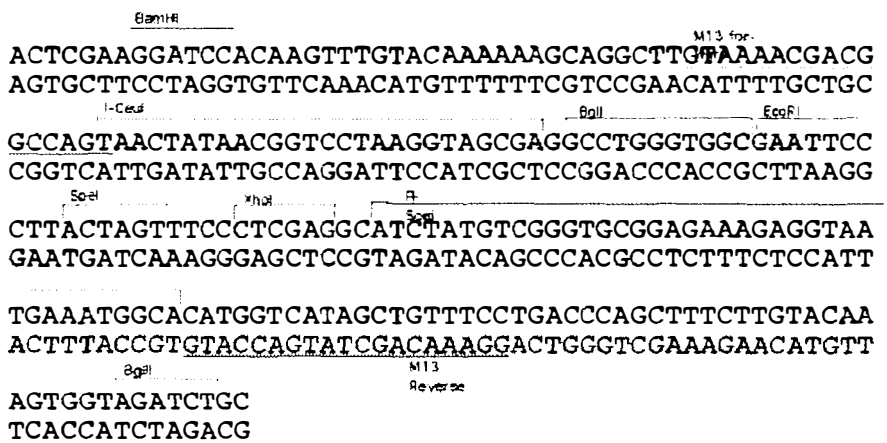
pFLC-I Polylinker for RH clones



10.5.14 pOTB7 (1815 bp)

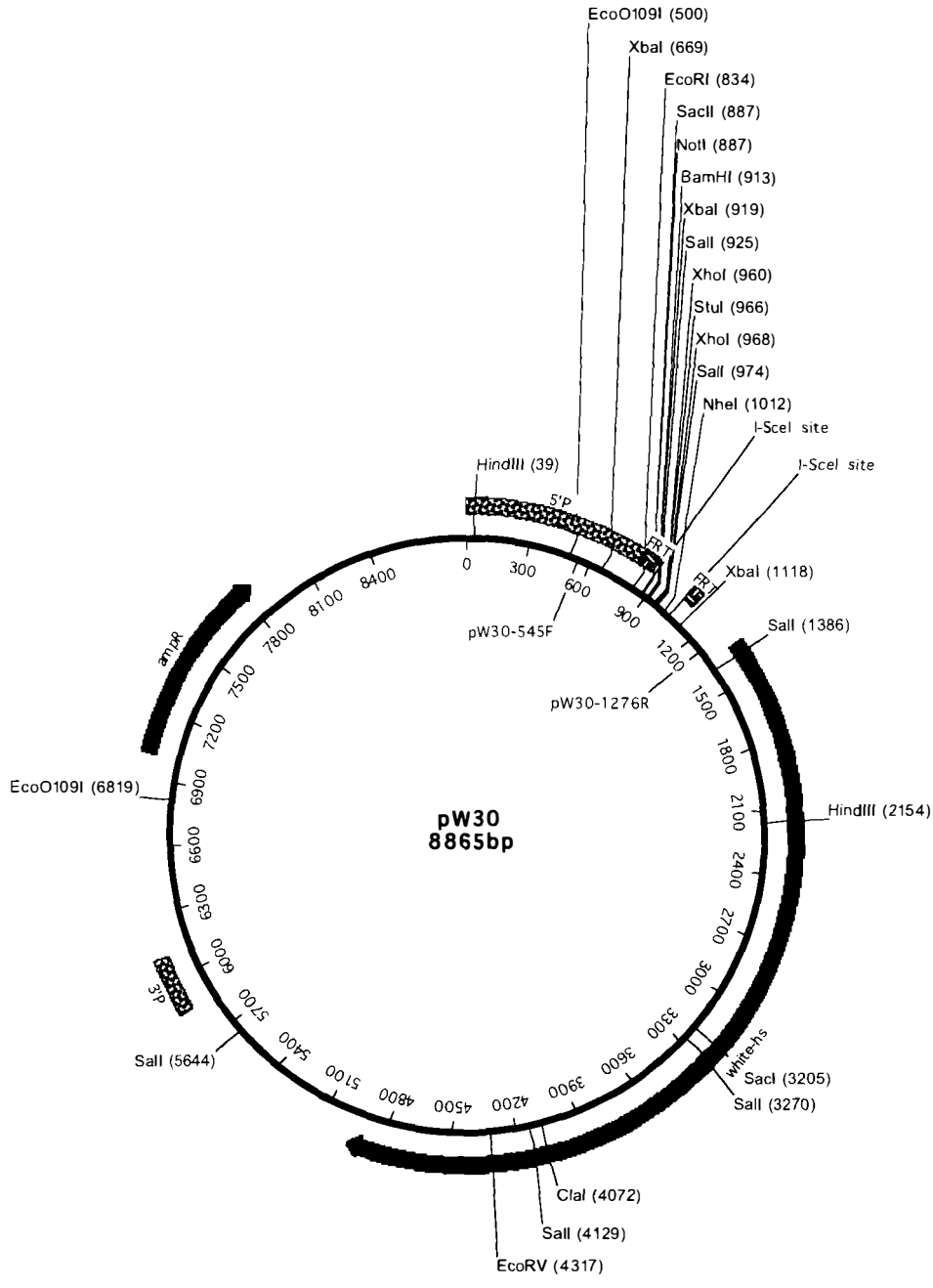
This vector contains the oligo-dT primed cDNA clones integrated in the antisense direction from the T7 polymerase binding site (Rubin *et al.*, 2000). The pOTB7 plasmid is based on the pOT2A backbone with a 204 bp inserted fragment designed to facilitate the use of the *Gateway* expression system from Life Technologies. pOTB7 contains an *attB1* site between the *Bam*HI and M13 forward sites. An *attB2* site is situated between *Bgl*II and M13 reverse sites. The 204 bp insert was ligated into *Xho*I and *Eco*RI sites within the pOT2A vector, destroying the *Xho*I site and altering the *Eco*RI site to a *Bgl*II site. The *Hind*III site in pOT2A was replaced by a *Not*I site. It is important to note that the *Eco*RI and *Xho*I sites are reversed with regard to the Sp6 and T7 sequencing primers compared to the pOT2A vector.





10.5.15 pW30 (8865 bp)

The pW30 plasmid map was obtained from Wei Gong (Gong and Golic, 2003). pW30 contains the FRT sites flanking I-SceI sites within an ampicillin-resistant *P*-element vector containing the *mini-white* marker gene.



10.6 Alignment of genomic sequences from *y w* strain with *y2; cn bw sp* strain used for BDGP genomic sequence

10.6.1 Alignment of *Lsp1α* 3' UTR genomic sequences from *y w* strain *and y2; cn bw sp* BDGP strain

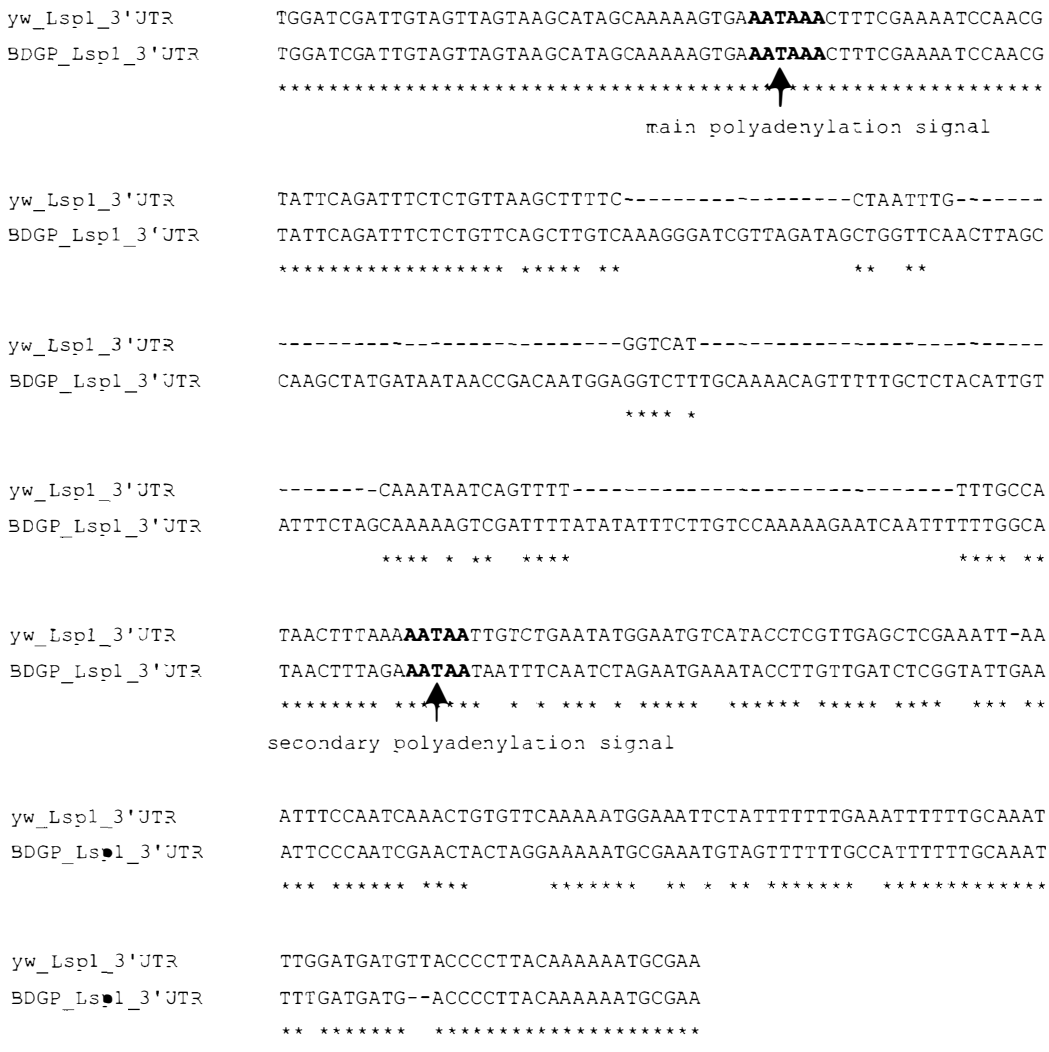
Consensus key

* - single, fully conserved residue

- no consensus

CLUSTAL W (1.81) multiple sequence alignment

```
yw_Lsp1_3'UTR       GCCACAGCACGAGCAGTTCTCCACCTTTGACTACACCTACTCGTGCGGCATCGGATCCGG
BDGP_Lsp1_3'UTR     GCCACAGCACGAGCAGTTCTCCACTTTTGACTACACCTACTCGTGCGGCATCGGATCCGG
                    ************************ ***********************************

yw_Lsp1_3'UTR       AGCTCGACACGTGGACAGCCTGCCTTTTGGATATCCCTTCGATCGTGAAATCAACGAGTA
BDGP_Lsp1_3'UTR     AGCTCGACACGTGGACAGCCTGCCTTTTGGATATCCCTTCGATCGTGAAATCAACGAGTA
                    ********************************^***************************
                                      EcoRV cuts at position 90 (bold)

yw_Lsp1_3'UTR       CGAGTTCCACGTGCCCAACATGTACTTCAAGGATGTGACCATCTATCATGCGGATACAAT
BDGP_Lsp1_3'UTR     CGAGTTCCACGTGCCCAACATGTACTTCAAGGATGTGACCATCTATCATGCGGATACAAT
                    ************************************************************

yw_Lsp1_3'UTR       GGAGAAGTACTATAACTACAAGGAATACACCAACTACGGTCACTTCGATTATTCCTTCTT
BDGP_Lsp1_3'UTR     GGAGAAGTACTATAACTACAAGGAATACACCAACTACGGTCACTTCGATTACTCCTTCTT
                    ************************************************* ********

yw_Lsp1_3'UTR       CAACGACTACTACACCAAGTACTTCAAGTTGTAGATCTGATAACATTCCGATTTTGACCA
BDGP_Lsp1_3'UTR     CAACGACTACTACACCAAGTACTTCAAGTTGTAGATCTGATAACATTCCGATTTTGACCA
                    ********************************^***************************
                                                ORF end (bold)
```


yw_Lspl_3'UTR TGGATCGATTGTAGTTAGTAAGCATAGCAAAAAGTGAAATAAACTTTCGAAAATCCAACG
BDGP_Lsp1_3'UTR TGGATCGATTGTAGTTAGTAAGCATAGCAAAAAGTGAAATAAACTTTCGAAAATCCAACG
main polyadenylation signal
yw_Lspl_3'UTR TATTCAGATTTCTCTGTTAAGCTTTTC----------CTAATTTG-------
BDGP_Lsp1_3'UTR TATTCAGATTTCTCTGTTCAGCTTGTCAAAGGGATCGTTAGATAGCTGGTTCAACTTAGC
yw_Lspl_3'UTR ----------GGTCAT----------
BDGP_Lspl_3'UTR CAAGCTATGATAATAACCGACAATGGAGGTCTTTGCAAAACAGTTTTTGCTCTACATTGT
yw_Lspl_3'UTR --------CAAATAATCAGTTTT----------TTTGCCA
BDGP_Lspl_3'UTR ATTTCTAGCAAAAAGTCGATTTTATATATTTCTTGTCCAAAAAGAATCAATTTTTTGGCA
yw_Lspl_3'UTR TAACTTTAAAAATAATTGTCTGAATATGGAATGTCATACCTCGTTGAGCTCGAAATT-AA
BDGP_Lspl_3'UTR TAACTTTAGAAATAATAATTTCAATCTAGAATGAAATACCTTGTTGATCTCGGTATTGAA
secondary polyadenylation signal
yw_Lspl_3'UTR ATTTCCAATCAAACTGTGTTCAAAAATGGAAATTCTATTTTTTTGAAATTTTTTGCAAAT
BDGP_Lspl_3'UTR ATTCCCAATCGAACTACTAGGAAAAATGCGAAATGTAGTTTTTTGCCATTTTTTGCAAAT
yw_Lspl_3'UTR TTGGATGATGTTACCCCTTACAAAAAATGCGAA
BDGP_Lspl_3'UTR TTTGATGATG--ACCCCTTACAAAAAATGCGAA

10.6.2 Alignment of genomic region from *CG2560 – Lsp1α*, including *Lsp1α* promoter, from *y w* flies and the *y2; cn bw sp* BDGP strain

Consensus key

* - single, fully conserved residue

- no consensus

yw_Lsp1alpha_promoter = (P)

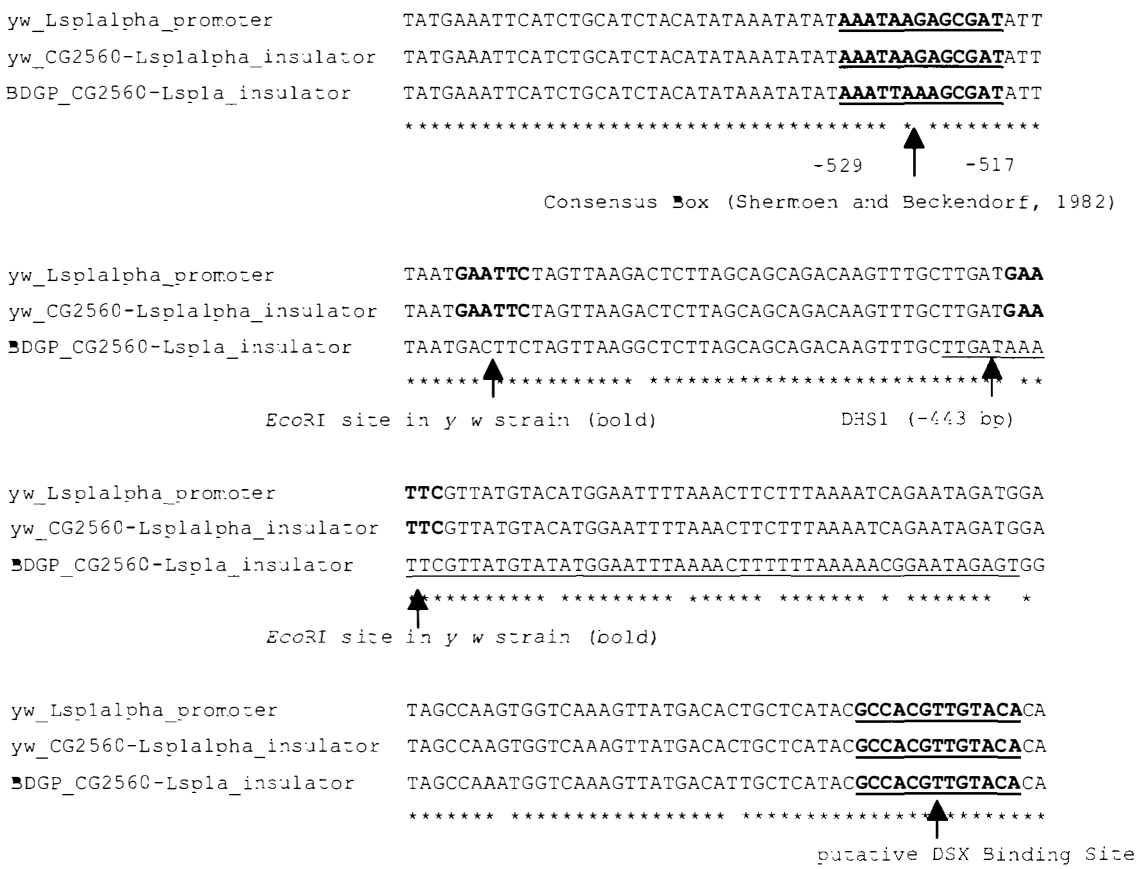
yw_CG2560-Lsp1alpha_insulator = (I)

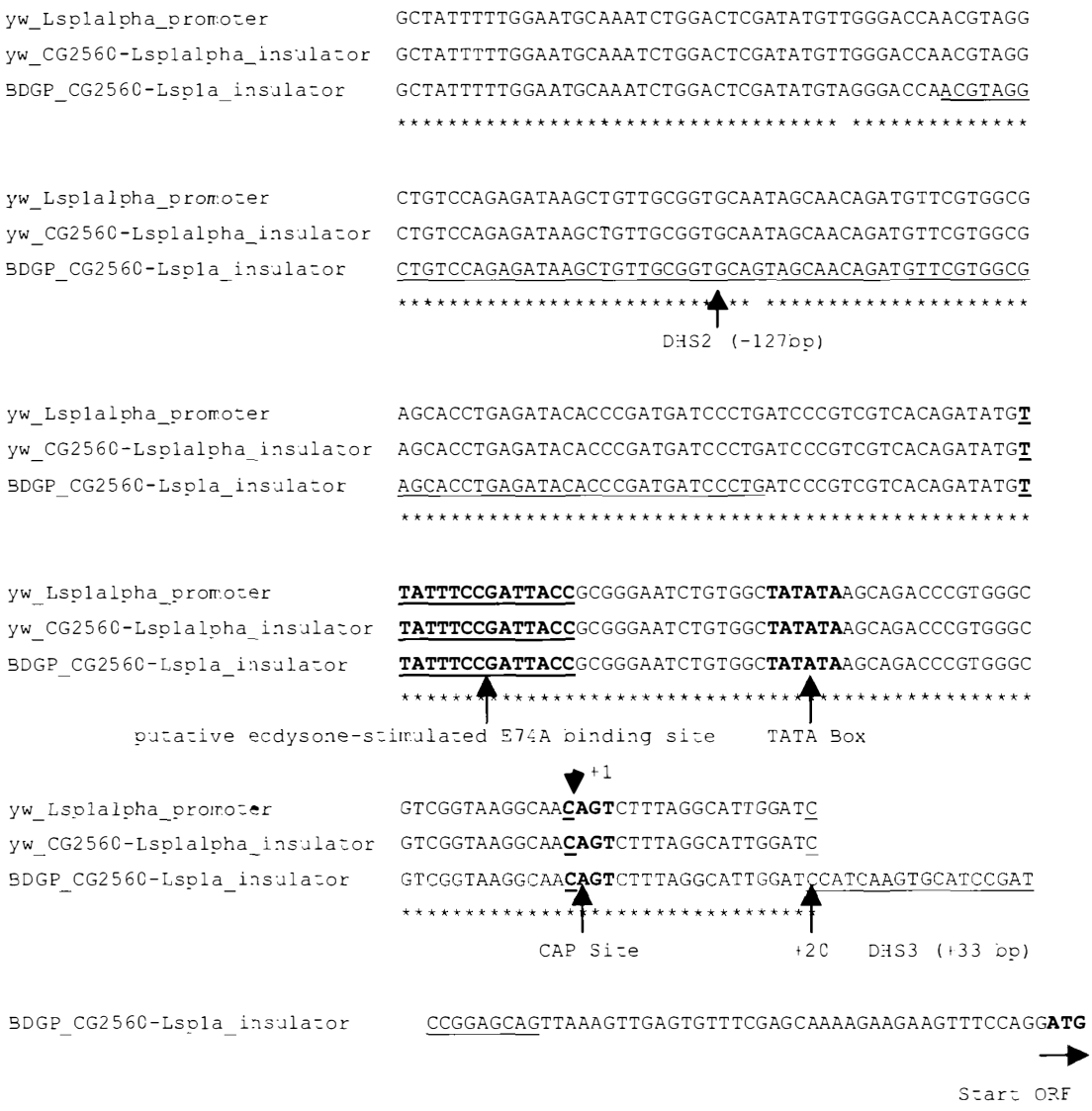
BDGP_CG2560-Lsp1a_insulator = *y2; cn bw sp* BDGP sequence from -863 bp to Translation Start Site.

Annotations and DNaseI Hypersensitive sites (DHS) mapped from (Jowett, 1985) and Genomatix Transcription Factor Binding Site Predictor.

CLUSTAL W (1.81) multiple sequence alignment

```
yw_Lsplalpha_promoter          --------------------------------------------------
yw_CG2560-Lsplalpha_insulator  CGATCAAAGTTTTAATGTGTTTTAAATGAGGAAAAATATGGATGGGGATA
BDGP_CG2560-Lspla_insulator    CGATCAAAGTTTTAATGTGTTTTAATTGAGGAATAATATGGATGGGGATA
                               ▲
                              -863 bp

yw_Lsplalpha_promoter          --------------------------------------------------
yw_CG2560-Lsplalpha_insulator  TTTAAAACTATTTTCTATTAATTTTTGGCTTTGTAATAAAATGTTATTGC
BDGP_CG2560-Lspla_insulator    TTTATATATATATT----------TTGGCTTTGTAATAAAATGTTATTGC

yw_Lsplalpha_promoter          --------------------------------------------------
yw_CG2560-Lsplalpha_insulator  AAGCAATTGTTGAAATTTTATTCAGCATCGTTACAAAAAAAAATTCGTTG
BDGP_CG2560-Lspla_insulator    AAGCAATTGTTGAAATTTTATTCAGCATCGTTACAAAAAAAAATTCGTTG

yw_Lsplalpha_promoter          --------------------------------------------------
yw_CG2560-Lsplalpha_insulator  TTTCAGAATAAAGTTAAGTTTTATTAACTGAAAATGAATAGTTTGTTTAC
BDGP_CG2560-Lspla_insulator    TTTCAGAATAAAGTTAAGTTTTATTAACTGAAAATGAATAGTTTGTTTAC

yw_Lsplalpha_promoter          --------------------------------------------------
yw_CG2560-Lsplalpha_insulator  TGTTTTATAAGCATGGCATATACAATTTTCCAACTAAGATACAGAATCTA
BDGP_CG2560-Lspla_insulator    TGTTTTATAAGCATGGCATATACAATTTTCCAACTTAGATACAGAATCTA

yw_Lsplalpha_promoter          ----------------------------------------GCAACAAACG
yw_CG2560-Lsplalpha_insulator  TTATTACAGATTATGATTTTGTTTGAATGCCTAAAAGCATGCAACAAACG
BDGP_CG2560-Lspla_insulator    TTATTACAGATTATGATTTTGTTCGAATGCCTAAAAGCATGCAACAAACG
                                                                ▲
                                                               -573 bp
```


yw_Lsplalpha_promoter TATGAAATTCATCTGCATCTACATATAAATATATAAATAAGAGCGATATT
yw_CG2560-Lsplalpha_insulator TATGAAATTCATCTGCATCTACATATAAATATATAAATAAGAGCGATATT
BDGP_CG2560-Lspla_insulator TATGAAATTCATCTGCATCTACATATAAATATATAAATTAAAGCGATATT
-529
-517
Consensus Box (Shermoen and Beckendorf, 1982)
yw_Lsplalpha_promoter TAATGAATTCTAGTTAAGACTCTTAGCAGCAGACAAGTTTGCTTGATGAA
yw_CG2560-Lsplalpha_insulator TAATGAATTCTAGTTAAGACTCTTAGCAGCAGACAAGTTTGCTTGATGAA
BDGP_CG2560-Lspla_insulator TAATGACTTCTAGTTAAGGCTCTTAGCAGCAGACAAGTTTGCTTGATAAA
EcoRI site in y w strain (bold)
DHS1 (-443 bp)
yw_Lsplalpha_promoter TTCGTTATGTACATGGAATTTTAAACTTCTTTAAAATCAGAATAGATGGA
yw_CG2560-Lsplalpha_insulator TTCGTTATGTACATGGAATTTTAAACTTCTTTAAAATCAGAATAGATGGA
BDGP_CG2560-Lspla_insulator TTCGTTATGTATATGGAATTTAAAACTTTTTTAAAAACGGAATAGAGTGG
EcoRI site in y w strain (bold)
yw_Lsplalpha_promoter TAGCCAAGTGGTCAAAGTTATGACACTGCTCATACGCCACGTTGTACACA
yw_CG2560-Lsplalpha_insulator TAGCCAAGTGGTCAAAGTTATGACACTGCTCATACGCCACGTTGTACACA
BDGP_CG2560-Lspla_insulator TAGCCAAATGGTCAAAGTTATGACATTGCTCATACGCCACGTTGTACACA
putative DSX Binding Site
yw_Lsplalpha_promoter CTTTTCTCAAAATAATGCCTGCCAATGCTTAACTTTAGTTCGTAATATGT
yw_CG2560-Lsplalpha_insulator CTTTTCTCAAAATAATGCCTGCCAATGCTTAACTTTAGTTCGTAATATGT
BDGP_CG2560-Lspla_insulator CTTTTCTCAAAATAATGCCTGCCAATGCTTAACTTTAGTTCGTAATATGT
yw_Lsplalpha_promoter ATATGATTGGTAATATCTCTATATTATATACATATAAATAATTAAAATTC
yw_CG2560-Lsplalpha_insulator ATATGATTGGTAATATCTCTATATTATATACATATAAATAATTAAAATTC
BDGP_CG2560-Lspla_insulator ATATGATTGGTAATATCTCTATATTATATACATATAAATAATTAAAATTC
yw_Lsplalpha_promoter AGAGCGATTGGAAGTTGTGCAGAAATTGCAGACAAAGAACCGAATCTAAC
yw_CG2560-Lsplalpha_insulator AAAGCGATTGGAAGTTGTGCAGAAATTGCAGACAAAGAACCGAATCTAAC
BDGP_CG2560-Lspla_insulator AAAGCGATTGAAACTTGTGCAGAAATTGCAGACAAAGAACCGAATCTAAC



Note: A fourth DHS is present in the ORF of *Lsp1α* at *ca.* +161 bp (Jowett, 1985).

10.7 Alignment of protein sequences from *Drosophila* species

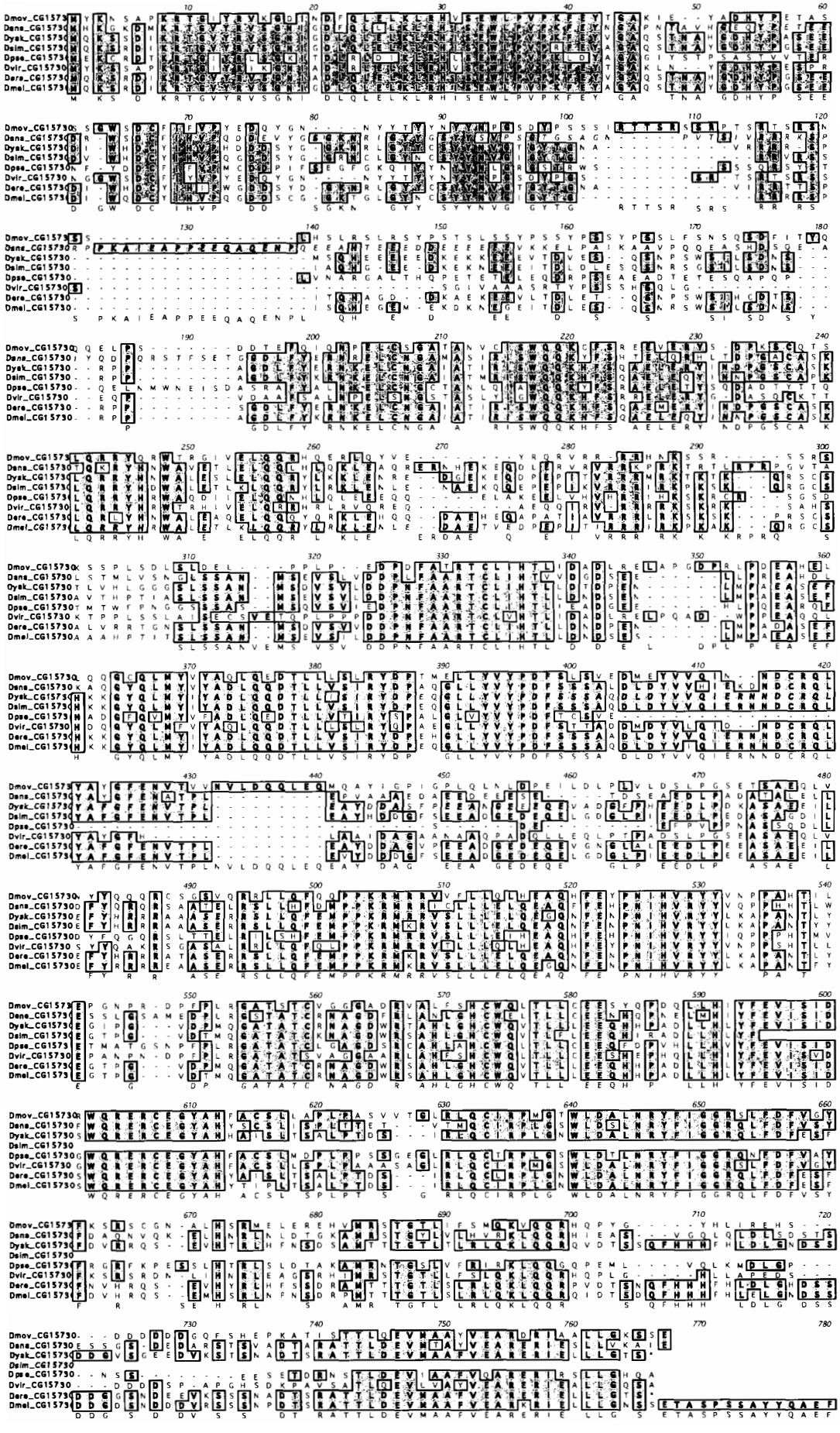
10.7.1 Alignment of putative CG15730 proteins from *Drosophila* species

ClustalW (v1.4) multiple sequence alignment (following page).

[8 Sequences Aligned; Alignment Score = 69664; Gaps Inserted = 121; Conserved Identities = 132; Pairwise Alignment Mode: Fast; Pairwise Alignment Parameters:ktup = 1, Gap Penalty = 3, Top Diagonals = 5, Window Size = 5; Multiple Alignment Parameters: Open Gap Penalty = 10.0, Extend Gap Penalty = 0.4, Delay Divergent = 40% , Gap Distance = 8, Similarity Matrix: blosum]

Key: *D. melanogaster* (Dmel), *D. yakuba* (Dyak), *D. erecta* (Dere), *D. simulans* (Dsim), *D. mojavensis* (Dmoj), *D. ananassae* (Dana), *D. pseudoobscura* (Dpse) and *D. virilis* (Dvir).

Grey shading indicates conserved residues. The *D. simulans* CG15730 sequence is not complete due to the partial nature of the genome data available.



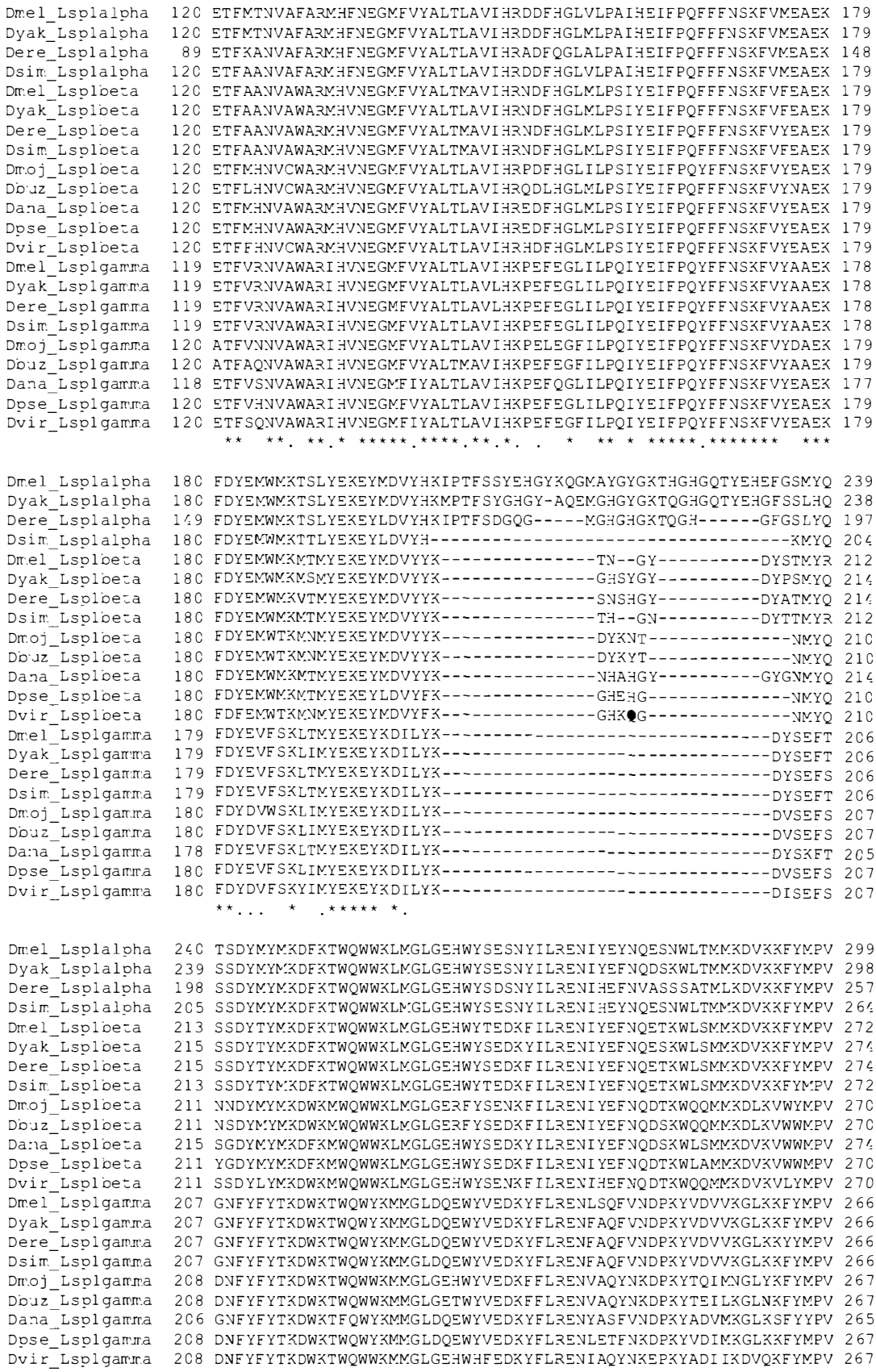
10.7.2 Alignment of putative LSP1 proteins from *Drosophila* species

ClustalW (v1.4) multiple sequence alignment

[22 Sequences Aligned; Alignment Score = 902940; Gaps Inserted = 89; Conserved Identities = 275; Pairwise Alignment Mode: Fast; Pairwise Alignment Parameters:ktup = 1, Gap Penalty = 3, Top Diagonals = 5, Window Size = 5; Multiple Alignment Parameters: Open Gap Penalty = 10.0, Extend Gap Penalty = 0.4, Delay Divergent = 40% , Gap Distance = 8, Similarity Matrix: blosum]

Key: *D. melanogaster* (Dmel), *D. yakuba* (Dyak), *D. erecta* (Dere), *D. simulans* (Dsim), *D. mojavensis* (Dmoj), *D. buzzatii* (Dbuz), *D. ananassae* (Dana), *D. pseudoobscura* (Dpsc) and *D. virilis* (Dvir). * single fully conserved residue.

```
Dmel_Lsplalpha     1 MKFAIAFLACVAVVTATAYHKTHDIKVADKAFLMKQKFLFEIVYRVEDPLMFEEYIAMGK  60
Dyak_Lsplalpha     1 MKFAIAFLACVAVVAATGYHQTHDIKVADKDFLMKQKFLFEIVYRVEDPLMFEEYIAMGK  60
Dere_Lsplalpha     1                                 FLMKQKFLFEIVYRVEDPLMFEEHIATGK  29
Dsim_Lsplalpha     1 MKFAIAFLACVAVVTATGYHKTHDIKVADKAYLMKQKFLFEIVYRVEDPLMFEEYIAMGK  60
Dmel_Lsplbeta      1 MKIAIALLACLGLAAAASFHQTHEVKIADKAFLMKQKFLFEIVYRVEDPLMFEDHIKQGE  60
Dyak_Lsplbeta      1 MKIAIALLACLGLVAAASFHQTHEVKIADKAFLLKQKFLFEIVYRVEDPLMFEEYIKQGS  60
Dere_Lsplbeta      1 MKIAIALLACLGLVAAAGFHQTHEVKIADKAFLLKQKFLFEIVYRVEDPLMFEEYIKQGQ  60
Dsim_Lsplbeta      1 MKIAIALLACLGLAAAASFHQTHEVKIADKAFLMKQKFLFEIVYRVEDPLMFEDHIKQGE  60
Dmoj_Lsplbeta      1 MKIAIALLACLGLVAAASFHQTHEVKIADKEFLEKQKFLFEIVYRVEDPLMFEEYIKKGE  60
Dbuz_Lsplbeta      1 MKIAIALLACLGLVAAASFHQTHEVKIADKNFLKMQKILFEVVYRMDEPLMFEELIKMGE  60
Dana_Lsplbeta      1 MKVAIALLACLGLVAAASFHKQHEVKIADKSFLGMQKFLFEIVYRVEDPLMFEEYIKLGK  60
Dpse_Lsplbeta      1 MKIAIALLALVGLVAAASFHQTHEVKIADKGFLVKQKFLFEVVYRIEDPLMFEDIIKLGD  60
Dvir_Lsplbeta      1 MKIVIALLACLGLAAAASFHKTHEVKIADKAFLEKQKFLFEIVYRVEDPLMFEEYIKKGD  60
Dmel_Lsplgamma     1 MKLTLVILALVACVTAFSVPT-QKVKIADKNFLEKQKFLFEIVHRIDEPLMFEEWIKMGQ  59
Dyak_Lsplgamma     1 MKLTLVILALVGCVTAFSVPT-QQVKIADKTFLEKQKFLFEVVYRIDEPLMFEEWIKMGQ  59
Dere_Lsplgamma     1 MKLTLVILALVGCVTAFSVPT-QKVKIADKTFLEKQKFLFEIVYRIDEPLMFEEWIKMGQ  59
Dsim_Lsplgamma     1 MKLTLVILALVACVTAFSVPT-QKVKIADKNFLEKQKFLFEIVHRVDEPLMFEEWIKMGQ  59
Dmoj_Lsplgamma     1 MKLTLVLLALVGCVAAFSVPTSTKVKIADAEFLKKQQFFFEILYRVEDPLMFEEWIKMGQ  60
Dbuz_Lsplgamma     1 MKLTLVLLALVGCVAAFSVPTPTKVKVADAEFLKMQKFILEALYRVEEPFMFEEWIKIGQ  60
Dana_Lsplgamma     1 MKLTLVLLALVGCVAAFSVPT-PKVKIADKNFLEKQKFLLEITYRIEDPLMFEEWIKFGT  59
Dpse_Lsplgamma     1 MKLCLVLLALVGCVCAFSVPT-QKVKIADKDFLEKQKFLFEMVYRVEDPLMFEEWIKMGQ  59
Dvir_Lsplgamma     1 MKLTLVLLALVGCVAAFSVPTSTKVKIADQEFLQKQKFFFEIVYRVEDPLLFEEWIKLGQ  60
                                                  .*  *.   *   *...* .**. *  *

Dmel_Lsplalpha    61 QFYFDKEHYTH-FDLYMEKFFEAHKAHALLPKGEFFGALVKHHAKQARGLFNFFYYAKDW 119
Dyak_Lsplalpha    61 KFYFEQEHYTH-FDLYMEKFYEAYKAHALLPKGEFFGALVKHHAKQARGLFNFFYYAKDW 119
Dere_Lsplalpha    30 QFHFEKQHYTH-FDLYMEKFHEAHKAHALLPKGEFFGALVKHHAKQARGLFNYFYYARDW  88
Dsim_Lsplalpha    61 QFFFEKEHYTH-FDLYMEKFFEAHKAHALLPKGEFFGALVKHHAKQARGLFNFFYYAKDW 119
Dmel_Lsplbeta     61 KFYFEESYYTH-YDMYMKKFFEAYKAHALLPKGEFFGALVMSHAKQARGLFNFFYYAKDW 119
Dyak_Lsplbeta     61 KFYFEESYYTH-YDLYMKKFFEAYKAHALLPKGEFFGALVMSHAKQARGLFNFFYYAKDW 119
Dere_Lsplbeta     61 KFYFEESYYTH-YDLYMQKFFEAYKAHALLPKGEFFGALVMSHAKQARGLFNFFYYAKDW 119
Dsim_Lsplbeta     61 KFFFEESYYTH-YDMYMKKFFVADKAHALLPKGEFFGALVMSHAKQARGLFNFFYYAKDW 119
Dmoj_Lsplbeta     61 KFYFEQSYYTK-YDKYMMKFYEGYKAHSLLPKGEFFGELVKTHAMQARGLFNFFYYAKDF 119
Dbuz_Lsplbeta     61 KFYFDQSYYTH-YDIYMKKFYEAYKTHSLLPKGEFFGQLVKAHAKQARGLFNFFYYAKDV 119
Dana_Lsplbeta     61 AFDFKEPNYTH-FDKYMKAFFEAYKVHALLPKGEFFGAAVMSHAKQARGLFNFFYYAKDW 119
Dpse_Lsplbeta     61 KFYFDESYYTH-YDLYMKKFFEAYKMHSLLPKGEFFGQLVKSHAKQARGLFNFFYYAKDW 119
Dvir_Lsplbeta     61 KFYFDESYYTH-YDIYMKKFYEAYKAHSLLPKGEFFGQLVKYHAKQARGLFNFFYYAKDF 119
Dmel_Lsplgamma    60 KLITDKAQYET-FDFYMEKLWESYKLGALLPKGEFFGALVKTHHKQAYGLFNFFYYAKDW 118
Dyak_Lsplgamma    60 KLITDKAQYET-FDFYMEKFWESYKLGVLLPKGEFFGALVKTHHKQAYGLFNFFYYAKDW 118
Dere_Lsplgamma    60 KLITDKAQYET-FDFYMEKFWESYKLGALLPKGEFFGALVKTHHKQAYGLFNFFYYAKDW 118
Dsim_Lsplgamma    60 KLITDKAQYET-FDFYMEKFWESYKLGALLPKGEFFGALVKTHHKQAYGLFNFFYYAKDW 118
Dmoj_Lsplgamma    61 KLVVDKAQYKE-YDFYMEKFYESYKLGALLPKGEFFGALVKTHYKQAYGLFNFFYYAKDW 119
Dbuz_Lsplgamma    61 KLVADKGQYKT-YDFYMEKFYEAYKLGALLPKGEFFGALVKTHHLQAYGLFNFFYYAKDW 119
Dana_Lsplgamma    60 KFIYDKSFYE--YDYYMEKFYESYKLGALLPKGEFFGALVKTHMKQAYGLFNFFYYAKDW 117
Dpse_Lsplgamma    60 KLIVDKAQYKQEFDFYMEKFWESYKLGALLPKGEFFGSLVKTHHKQAYGLFNFFYYANDW 119
Dvir_Lsplgamma    61 KLVVDKSLYTT-FDFYMEKFWESYKLGALLPKGEFFGALVKTHHKQAWGLFNFFYYAKDW 119
                            *   .* **       *   *********  *  *  ** ****.**** *
```


Dmel_Lsplalpha 120 ETFMTNVAFARMHFNEGMFVYALTLAVIHRDDFHGLVLPAIHEIFPQFFFNSKFVMEAEK 179
Dyak_Lsplalpha 120 ETFMTNVAFARMHFNEGMFVYALTLAVIHRDDFHGLMLPAIHEIFPQFFFNSKFVMEAEK 179
Dere_Lsplalpha 89 ETFKANVAFARMHFNEGMFVYALTLAVIHRADFQGLALPAIHEIFPQFFFNSKFVMEAEK 148
Dsim_Lsplalpha 120 ETFAANVAFARMHFNEGMFVYALTLAVIHRDDFHGLVLPAIHEIFPQFFFNSKFVMEAEK 179
Dmel_Lsplbeta 120 ETFAANVAWARMHVNEGMFVYALTMAVIHRNDFHGLMLPSIYEIFPQFFFNSKFVFEAEK 179
Dyak_Lsplbeta 120 ETFAANVAWARMHVNEGMFVYALTLAVIHRNDFHGLMLPSIYEIFPQFFFNSKFVFEAEK 179
Dere_Lsplbeta 120 ETFAANVAWARMHVNEGMFVYALTMAVIHRNDFHGLMLPSIYEIFPQFFFNSKFVYEAEK 179
Dsim_Lsplbeta 120 ETFAANVAWARMHVNEGMFVYALTMAVIHRNDFHGLMLPSIYEIFPQFFFNSKFVFEAEK 179
Dmoj_Lsplbeta 120 ETFMHNVCWARMHVNEGMFVYALTLAVIHRPDFHGLILPSIYEIFPQYFFNSKFVYEAEK 179
Dbuz_Lsplbeta 120 ETFLHNVCWARMHVNEGMFVYALTLAVIHRQDLHGLMLPSIYEIFPQFFFNSKFVYNAEK 179
Dana_Lsplbeta 120 ETFMHNVAWARMHVNEGMFVYALTLAVIHREDFHGLMLPSIYEIFPQFFFNSKFVYEAEK 179
Dpse_Lsplbeta 120 ETFMHNVAWARMHVNEGMFVYALTLAVIHREDFHGLMLPSIYEIFPQFFFNSKFVYEAEK 179
Dvir_Lsplbeta 120 ETFFHNVCWARMHVNEGMFVYALTLAVIHRHDFHGLMLPSIYEIFPQFFFNSKFVYEAEK 179
Dmel_Lsplgamma 119 ETFVRNVAWARIHVNEGMFVYALTLAVIHKPEFEGLILPQIYEIFPQYFFNSKFVYAAEK 178
Dyak_Lsplgamma 119 ETFVRNVAWARIHVNEGMFVYALTLAVLHKPEFEGLILPQIYEIFPQYFFNSKFVYAAEK 178
Dere_Lsplgamma 119 ETFVRNVAWARIHVNEGMFVYALTLAVLHKPEFEGLILPQIYEIFPQYFFNSKFVYAAEK 178
Dsim_Lsplgamma 119 ETFVRNVAWARIHVNEGMFVYALTLAVIHKPEFEGLILPQIYEIFPQYFFNSKFVYAAEK 178
Dmoj_Lsplgamma 120 ATFVNNVAWARIHVNEGMFVYALTLAVIHKPELEGFILPQIYEIFPQYFFNSKFVYDAEK 179
Dbuz_Lsplgamma 120 ATFAQNVAWARIHVNEGMFVYALTMAVIHKPEFEGFILPQIYEIFPQYFFNSKFVYAAEK 179
Dana_Lsplgamma 118 ETFVSNVAWARIHVNEGMFIYALTLAVIHKPEFQGLILPQIYEIFPQYFFNSKFVYEAEK 177
Dpse_Lsplgamma 120 ETFVHNVAWARIHVNEGMFVYALTLAVIHKPEFEGLILPQIYEIFPQYFFNSKFVYEAEK 179
Dvir_Lsplgamma 120 ETFSQNVAWARIHVNEGMFIYALTLAVIHKPEFEGFILPQIYEIFPQYFFNSKFVYEAEK 179
** **. **.* *****.****.**.*. . * ** * *****.******* ***
Dmel_Lsplalpha 180 FDYEMWMKTSLYEKEYMDVYHKIPTFSSYEHGYKQGMAYGYGKTHGHGQTYEHEFGSMYQ 239
Dyak_Lsplalpha 180 FDYEMWMKTSLYEKEYMDVYHKMPTFSYGHGY-AQEMGHGYGKTQGHGQTYEHGFSSLHQ 238
Dere_Lsplalpha 149 FDYEMWMKTSLYEKEYLDVYHKIPTFSDGQG-----MGHGHGKTQGH------GFGSLYQ 197
Dsim_Lsplalpha 180 FDYEMWMKTTLYEKEYLDVYH----------KMYQ 204
Dmel_Lsplbeta 180 FDYEMWMKMTMYEKEYMDVYYK----------TN--GY----------DYSTMYR 212
Dyak_Lsplbeta 180 FDYEMWMKMSMYEKEYMDVYYK----------GHSYGY----------DYPSMYQ 214
Dere_Lsplbeta 180 FDYEMWMKVTMYEKEYMDVYYK----------SNSHGY----------DYATMYQ 214
Dsim_Lsplbeta 180 FDYEMWMKMTMYEKEYMDVYYK----------TH--GN----------DYTTMYR 212
Dmoj_Lsplbeta 180 FDYEMWTKMNMYEKEYMDVYYK----------DYKNT----------NMYQ 210
Dbuz_Lsplbeta 180 FDYEMWTKMNMYEKEYMDVYYK----------DYKYT----------NMYQ 210
Dana_Lsplbeta 180 FDYEMWMKMTMYEKEYMDVYYK----------NHAHGY----------GYGNMYQ 214
Dpse_Lsplbeta 180 FDYEMWMKMTMYEKEYLDVYFK----------GHEHG----------NMYQ 210
Dvir_Lsplbeta 180 FDFEMWTKMNMYEKEYMDVYFK----------GHK●G----------NMYQ 210
Dmel_Lsplgamma 179 FDYEVFSKLTMYEKEYKDILYK----------DYSEFT 206
Dyak_Lsplgamma 179 FDYEVFSKLIMYEKEYKDILYK----------DYSEFT 206
Dere_Lsplgamma 179 FDYEVFSKLTMYEKEYKDILYK----------DYSEFS 206
Dsim_Lsplgamma 179 FDYEVFSKLTMYEKEYKDILYK----------DYSEFT 206
Dmoj_Lsplgamma 180 FDYDVWSKLIMYEKEYKDILYK----------DVSEFS 207
Dbuz_Lsplgamma 180 FDYDVFSKLIMYEKEYKDILYK----------DVSEFS 207
Dana_Lsplgamma 178 FDYEVFSKLTMYEKEYKDILYK----------DYSKFT 205
Dpse_Lsplgamma 180 FDYEVFSKLIMYEKEYKDILYK----------DVSEFS 207
Dvir_Lsplgamma 180 FDYDVFSKYIMYEKEYKDILYK----------DISEFS 207
... * .*** *.
Dmel_Lsplalpha 240 TSDYMYMKDFKTWQWWKLMGLGEHWYSESNYILRENIYEYNQESNWLTMMKDVKKFYMPV 299
Dyak_Lsplalpha 239 SSDYMYMKDFKTWQWWKLMGLGEHWYSESNYILRENIYEFNQDSKWLTMMKDVKKFYMPV 298
Dere_Lsplalpha 198 SSDYMYMKDFKTWQWWKLMGLGEHWYSDSNYILRENIHEFNVASSSATMLKDVKKFYMPV 257
Dsim_Lsplalpha 205 SSDYMYMKDFKTWQWWKLMGLGEHWYSESNYILRENIHEYNQESNWLTMMKDVKKFYMPV 264
Dmel_Lsplbeta 213 SSDYTYMKDFKTWQWWKLMGLGEHWYTEDKFILRENIYEFNQETKWLSMMKDVKKFYMPV 272
Dyak_Lsplbeta 215 SSDYTYMKDFKTWQWWKLMGLGEHWYSEDKYILRENIYEFNQESKWLSMMKDVKKFYMPV 274
Dere_Lsplbeta 215 SSDYTYMKDFKTWQWWKLMGLGEHWYSEDKFILRENIYEFNQETKWLSMMKDVKKFYMPV 274
Dsim_Lsplbeta 213 SSDYTYMKDFKTWQWWKLMGLGEHWYTEDKFILRENIYEFNQETKWLSMMKDVKKFYMPV 272
Dmoj_Lsplbeta 211 NNDYMYMKDWKMWQWWKLMGLGERFYSENKFILRENIYEFNQDTKWQQMMKDLKVWYMPV 270
Dbuz_Lsplbeta 211 NSDYMYMKDWKMWQWWKLMGLGERFYSEDKFILRENIYEFNQDSKWQQMMKDLKVWWMPV 270
Dana_Lsplbeta 215 SGDYMYMKDFKMWQWWKLMGLGEHWYSEDKYILRENIYEFNQDSKWLSMMKDVKVWWMPV 274
Dpse_Lsplbeta 211 YGDYMYMKDFKMWQWWKLMGLGEHWYSEDKFILRENIYEFNQDTKWLAMMKDVKVWWMPV 270
Dvir_Lsplbeta 211 SSDYLYMKDWKMWQWWKLMGLGEHWYSENKFILRENIHEFNQDTKWQQMMKDVKVLYMPV 270
Dmel_Lsplgamma 207 GNFYFYTKDWKTWQWYKMMGLDQEWYVEDKYFLRENLSQFVNDPKYVDVVKGLKKFYMPV 266
Dyak_Lsplgamma 207 GNFYFYTKDWKTWQWYKMMGLDQEWYVEDKYFLRENFAQFVNDPKYVDVVKGLKKFYMPV 266
Dere_Lsplgamma 207 GNFYFYTKDWKTWQWYKMMGLDQEWYVEDKYFLRENFAQFVNDPKYVDVVKGLKKYYMPV 266
Dsim_Lsplgamma 207 GNFYFYTKDWKTWQWYKMMGLDQEWYVEDKYFLRENFAQFVNDPKYVDVVKGLKKFYMPV 266
Dmoj_Lsplgamma 208 DNFYFYTKDWKTWQWWKMMGLGEHWYVEDKFFLRENVAQYNKDPKYTQIMNGLYKFYMPV 267
Dbuz_Lsplgamma 208 DNFYFYTKDWKTWQWWKMMGLGETWYVEDKFFLRENVAQYNKDPKYTEILKGLNKFYMPV 267
Dana_Lsplgamma 206 GNFYFYTKDWKTFQWYKMMGLDQEWYVEDKYFLRENYASFVNDPKYADVMKGLKSFYYPV 265
Dpse_Lsplgamma 208 DNFYFYTKDWKTWQWYKMMGLDQEWYVEDKYFLRENLETFNKDPKYVDIMKGLKKFYMPV 267
Dvir_Lsplgamma 208 DNFYFYTKDWKTWQWWKMMGLGEHWHFEDKYFLRENIAQYNKEPKYADIIKDVQKFYMPV 267
* * ** * ** *.*** . . . **** . .. . **

```
Dmel_Lsplalpha  300 DYSRDLYLYNEESKLSYFTEDLGWNSYWYYLNMDYSFFLDGKTFGLQNDRRGEWWLYNVH 359
Dyak_Lsplalpha  299 DYSRDLHLYNEESKLSYFTEDLGWNTYWYYLNMDYSFFLDGKTFGLQNDRRGEWWLYNVH 358
Dere_Lsplalpha  258 DYSRDLHLYNEESKLSYFTEDLGWNSYWYYLNMDYSFFLDGKTFGLQNDRRGEWWLYNVH 317
Dsim_Lsplalpha  265 DYSRDLYLYNEETKLSYFTEDLGWNTYWYYLNMDYSFFLDGKTFGLQNDRRGEWWLYNVH 324
Dmel_Lsplbeta   273 DYTRDLNLYNKESKLSYFTEDLGWNAYWYYLNMDYSFFLDGNTFDLKNDRRGEWWLYNVH 332
Dyak_Lsplbeta   275 DYSRDLNIYNEESKLSYFTEDLGWNAYWYYLNMDYSFFLDGNTFDLKSDRRGEWWLYNVH 334
Dere_Lsplbeta   275 DYSRDLNIYNSESKLSYFTEDLGWNAYWYYLNMDYSFFLDGNTFDLKSDRRGEWWLYNVH 334
Dsim_Lsplbeta   273 DYTRDLNLYNKESKLSYFTEDLGWNAYWYYLNMDYSFFLDGNTFDLKNDRRGEWWLYNVH 332
Dmoj_Lsplbeta   271 DYSRDLNMYNEESKLSYFTEDLGWNAYWYYLNLDYSFFLDGNTFGLKNDRRGELWLYTVH 330
Dbuz_Lsplbeta   271 DYSRDLNMYNEESKLSYFTEDLGWNAYWYYLNLDYSFFLDGNTFGLKNDRRGELWLYTVH 330
Dana_Lsplbeta   275 DYTRDLNMYNEESKLSYFTEDLGWNSYWYYLNMDYSFFLDGTTFDLKNDRRGEWWLYNVH 334
Dpse_Lsplbeta   271 DYTRDLDLYNEESKLSYFTEDLGWNAYWYYLNMDYSFFLDGTTFDLKNDRRGEWWLYNVH 330
Dvir_Lsplbeta   271 DYSRDLDFYNEESKLSYFTEDLGWNAYWYYFNLDYSFFLDGNTFGLKNDRRGELWLYTVH 330
Dmel_Lsplgamma  267 DYTRDIDFFNDETKMTYFTEDLGWNAYWYYLNMDYAFFLNGKQFGLDKDRRGEYWIYNVQ 326
Dyak_Lsplgamma  267 DYTRDIDFFNDETKMSYFTEDLGWNAYWYYLNMDYAFFLNGKQFGLDKDRRGEYWIYNVQ 326
Dere_Lsplgamma  267 DYTRDIYFFNDETKMTYFTEDLGWNAYWYYLNMDYAFFLNGKQFGLDKDRRGEYWIYNVQ 326
Dsim_Lsplgamma  267 DYTRDIDFFNDETKMTYFTEDLGWNAYWYYLNMDYAFFLNGKQFGLDKDRRGEYWIYSVN 326
Dmoj_Lsplgamma  268 DYTRDIDFFNAESKLSYFTEDLGWNAYWYYLNMDYAFFLDGKQFGLDKDRRGEWWIYNVQ 327
Dbuz_Lsplgamma  268 DYTRDIEFFNAESKLSYFTEDLGWNAYWYYLNLDYAFFLDGKQFGLDKDRRGEWWIYNVQ 327
Dana_Lsplgamma  266 DYTRDIEFFNDQSKLSYFTEDLGWNAYWYYLNMDYAFFLDGKKFGLDKDRRGEWWIYNVQ 325
Dpse_Lsplgamma  268 DYTRDIDFFNAESKLSYFTEDLGWNAYWYYLNMDYAFFLNGKEFGLDKDRRGEWWIYNVQ 327
Dvir_Lsplgamma  268 DYTRDIDFFNAESKLSYFTEDLGWNSYWYYLNLDYAFFLDGKQFGLDKDRRGEYWIYNVQ 327
                        **.**.  .* ..*..*********.**** *.**.*** *  * *  ***** *.* *

Dmel_Lsplalpha  360 QLLSRYHMERLSHGLGEIPQFSWFHQIEMGYDPQLIYYNGIGYSYRKNYYELETYGNFEM 419
Dyak_Lsplalpha  359 QLLSRYHMERLSHGLGEIPQFSWYHQIEMGYDPQMIYYNGIGYSYRKNYYELETYGNFEM 418
Dere_Lsplalpha  318 QLLSRYHMERLSHGLGEIPQFSWFHQIEMGYDPQLIYYNGIGYSYRKNYYELETYGNFEM 377
Dsim_Lsplalpha  325 QLLSRYHMERLSHGLGEIPQFSWFHQIEMGYDPQMIYYNGIGYSYRKNYYELETYGNFEM 384
Dmel_Lsplbeta   333 QLLSRYYMERLSHGFGEIPEFSWYQQIEMGYDPQLIYYNGIGYSFRKNYYEMETYANYDM 392
Dyak_Lsplbeta   335 QLLSRYYMERLSQGFGEIPEFSWYHQIEMGYDPQMIYYSGIGYSYRKNYYEMDTYANYDM 394
Dere_Lsplbeta   335 QLLSRYYMERLSHGFGEIPEFSWYHQIEMGYDPQMIYYNGIGYSYRKNYYEMDTYANYDM 394
Dsim_Lsplbeta   333 QLLSRYYMERLSHGFGEIPEFSWYHQIEMGYDPQMIYYNGIGYSFRKNYYEMETYANYDM 392
Dmoj_Lsplbeta   331 QLLSRYYMERLSHGYGEIPEFSWYSQIEMGYDPQLIYYNGIGYSYRKDYYELETYGNFEM 390
Dbuz_Lsplbeta   331 QLLSRYYMERLSHGYGEIAEFSWYTQIEMGYDPQLIYYNGIGYSYRKNYYELETYGNFEM 390
Dana_Lsplbeta   335 QLLSRYYMERLSHGFGAIPEFSWYHQIEMGYDPQMIYYNGIGYSYRKNYYEMETYGNFEM 394
Dpse_Lsplbeta   331 QLLSRYYMERLSHGFGDIPEFSWYHQIEMGYDPQMIYYNGIGFSYRKNYYEMETYGNFEM 390
Dvir_Lsplbeta   331 QLLSRYYMERLSHGYGDIPEFSWYHQIEMGYDPQLIYYNGIGYSYRKNYYEMETYGNFEM 390
Dmel_Lsplgamma  327 QILARYYQERLANGFGEIPEFFWYKQIEYGYDPQLIYYNGIGYSYRKNYYDFYTYGKFEM 386
Dyak_Lsplgamma  327 QILARYYQERLANGFGEIPEFFWYKQIEYGYDPQLIYYNGIGYSYRKNYYDFYTYGNFEM 386
Dere_Lsplgamma  327 QILARYYQERLANGFGEIPEFFWYKQIEYGYDPQLIYYNGIGYSYRKNYYEFYTYGQFEM 386
Dsim_Lsplgamma  327 RSLARYYQERLANGFGEIPDFFWYKQIEYGYDPQLIYYNGIGYSYRKNYYDFYTYGKFEM 386
Dmoj_Lsplgamma  328 QILARYYQERLANGFGEIPEYFWYKQVEYGYDPQLIYYNGIGYSYRKNYFEYQTYGKFDM 387
Dbuz_Lsplgamma  328 QILARYYQERLANGFGEIPEYFWYKQVEYGYDPQLIYYNGIGYSYRKNYFEYQTYGKFDM 387
Dana_Lsplgamma  326 QILARYYQERLANGFGEIPEFFWYKQIEYGYDPQLIYYNGVGYSYRKNYYDFYTYGNFEM 385
Dpse_Lsplgamma  328 QILARYYQERLANGYGDIPEFFWYKQIEYGYDPQLIYYNGVGYSYRKNYYDYQTYGNFEM 387
Dvir_Lsplgamma  328 QILARYYQERLANGFGEIPEFFWYKQIEYGYDPQLIYYNGVGYSYRKNYFEYQTYGKFDI 387
                        . *.**  ***. * * *  . *. *.* *****.*** *.*.*.** *..  **  ...

Dmel_Lsplalpha  420 LDKITGFQQRIQNIVELGYYQTTDGHMIDLRKPESIEIIGNMLQGNVDAIDNIFFQFWYM 479
Dyak_Lsplalpha  419 LDKITDFQRRIHNIIELGYYQTTDGHMIDLRKPESIEIIGNMLQGNIDAIDNMFFQFWYM 478
Dere_Lsplalpha  378 LDKITGFQQRIHSIIDSGYYQTTDGHMIDLRQPESIEIIGNMLQGNVDAMDNMFFQFWYM 437
Dsim_Lsplalpha  385 LDTITG---------------------------------------------------- 390
Dmel_Lsplbeta   393 LDKITGFLKRIHNIVEMGYYKTADGHTIDLRKPEAIEFIGNMLQGNVDAMDKMFYQFWYM 452
Dyak_Lsplbeta   395 LDKITGFQKRIHNIVELGYYKTADGHTIDLRKPEAIEFIGNMLQGNVDAMDKMFYQFWYM 454
Dere_Lsplbeta   395 LDKITGFLKRIHNIVELGYYKSADGHIIDLRKPEAIEFIGNMLQGNVDAMDKMFYQFWYM 454
Dsim_Lsplbeta   393 LDKITGFLKRIHNIVELGYYKTADGHTIDLRKPEAIEFIGNMLQGNVDAMDKMFYQFWYM 452
Dmoj_Lsplbeta   391 LDKITGFMTRVQNIIDMGYYKTADGHMIDLRKPESVEYIGNMMQGNIDSMDKMFYQFWYM 450
Dbuz_Lsplbeta   391 LDKITGFMKRIHNIIDMGYYKTADGHMIDLRKPESIEYIGNMMQGNIDAMDKMFYQFWYM 450
Dana_Lsplbeta   395 LDKITSFMKRIQTIVDMGYYKTMDGHMIDLRKPESIEYIGNMMQGNVDAMDKMFFQFWYM 454
Dpse_Lsplbeta   391 LDKITGFMKRIQDIIELGYYKTTDGHTIDLRKPESIEFIGNLMQGNIDAMDKMFYQFWYM 450
Dvir_Lsplbeta   391 LDKITGFLKRIHNIVETGYYKTADGHMVDLRKPESIEYIGNILQGNVDSMDKMFYQFWYM 450
Dmel_Lsplgamma  387 YSQIQNFFSRVYKVLETGFYKTADGQVFDLHKPEAIKIVGNYLQGNADTFDKYFFNYYYL 446
Dyak_Lsplgamma  387 YNQIQNFFSRIYKVLETGFYKTADGQVFDLHKPEAIKIVGNYLQGNADTFDKYFFNYYYM 446
Dere_Lsplgamma  387 YNQIQNFFSRIYKALETGFYKTVDGQVFDLHKPEAIQIVGNYLQGNADTFDKYFFNYYYM 446
Dsim_Lsplgamma  387 YNQIQNFFSRVYKVLETGFYKTADGQVFDLHKPEAIKIVGNYLQGNADTFDKYFFNYYYL 446
Dmoj_Lsplgamma  388 LYQIQNFFDRINNVIETGYYKTADGVVIDFRKPEAIQFIGNYLQGNVDTYDKYFFNYYYV 447
Dbuz_Lsplgamma  388 LYQIQSFFDRIYDVIETGYYKTADGVVIDLRKPESIKFIGNYLQGNVDTYDKYFFNYYYV 447
Dana_Lsplgamma  386 FSKIQNFFGKVYKFLETGYYKTADGQVIDFHKPESVKFIGNYLQGNADTYDKYFFNFYYM 445
Dpse_Lsplgamma  388 FNQIQNFFSRVYKVLETGFYKTADGVVIDLRKPEAIKFIGNYLQGNVDTYDKYFFNYFYM 447
Dvir_Lsplgamma  388 LYQIQNFFDRIYDVLDTGYFKTADGVVIDLHKPESIKFIGNYLQGNVDTYDKYFFNYWYV 447
                          *
```

```
Dmel_Lsplalpha  480 LAHMYFADTHYYQMEVYPNVMLNFETMMRDPMFYMFYKSIAQVYFQFMHHLPKYTKEQLL 539
Dyak_Lsplalpha  479 LAHMYFADTHYYQMEVYPNVMLNFETMMRDTMFYMFYKSIAQVYFQFMHHLPKYTKEQLL 538
Dere_Lsplalpha  438 LAHMYFADTHYYQMEVYPNVMLNFETMMRDPMFYMFYKSIAQVYFQFMHHLPKYTKEQLL 497
Dsim_Lsplalpha  391 ----------FQMEVYPNVMLNFETMMRDPMFYMFYKSIAQVYFQFMHHLPKYTKEQLL 439
Dmel_Lsplbeta   453 LAHMYFADADYYQMDVYPNVMLNFETMMRDPMYYMFYKSIAQVYFQFMHYLPKYTKEQLL 512
Dyak_Lsplbeta   455 LAHMYFADADYHQMDVYPNVMLNFETMMRDPMYYMFYKSIAQVYFQFMHYLPKYTKEQLL 514
Dere_Lsplbeta   455 LAHMYFADADYAQMDVYPNVMLNFETMMRDPMYYMFYKSIAQVYFQFMHYLPKYTKEQLL 514
Dsim_Lsplbeta   453 LAHMYFADADYHQMDVYPNVMLNFETMMRDPMYYMFYKSIAQVYFQFMHYLPKYTKEQLL 512
Dmoj_Lsplbeta   451 FAHMYFADADYYQMEVYPNVMLNFETMMRDPMYYMFYKSIAQIYFKFMHYLPKYTKEQLM 510
Dbuz_Lsplbeta   451 FAHMYFADADYYQMEVYPNVMLNFETMMRDPMYYMFYKSIAQIYFKFMHYLPKYTKEQLM 510
Dana_Lsplbeta   455 LAHMYFADVDFHQMEVYPNVMLNFETMMRDPMYYMFYKSIVQVYFQFMHYLPKYTKEQLL 514
Dpse_Lsplbeta   451 LAHMYFGDVDYHQMEVYPNVMLNFETMMRDPMYYMFYKSIASIYFKFMHYLPKYTKEQLL 510
Dvir_Lsplbeta   451 FAHMYFADTDYYQMEVYPNVMLNFETMMRDPMYYMFYKSIAQIYFKFMHYLPKYTKEQLL 510
Dmel_Lsplgamma  447 LAHMYFADVDYNDMEVFPNVFLNFETMLRDPFFYTFYKKFTDVFYTFKYYLKPYTQKDLF 506
Dyak_Lsplgamma  447 LAHMYFADVDYTDMEVFPNIFLNFETMLRDPFFYTFYKKFTDVFYTFKYYLKPYTQKDLF 506
Dere_Lsplgamma  447 LAHMYFADVDYTDMEVFPNIFLNFETMLRDPFFYTFYKKFTDVFYTFKYYLKPYTQKDLF 506
Dsim_Lsplgamma  447 LAHMYFADVDYTDMEVFPNIFLNFETMLRDPFFYTFYKKFTDVFYTFKYYLKPYTQKDLF 506
Dmoj_Lsplgamma  448 LSHMYFADVDYFDFEVFPNVFLNFETMLRDPFFYSFYKKFTDVFYKFKFYLAPYTKNDLT 507
Dbuz_Lsplgamma  448 LSHMYFADVDYFDYEVFPNVFLNFETMMRDPFFYSFYKKFADVFYKFKFYLAPYTKKDLT 507
Dana_Lsplgamma  446 YSHMYFADVDYNDMDVFPNVFLNYETMLRDPFFYTFYKKFTDVFYQFKYFLGPYTQKDLF 505
Dpse_Lsplgamma  448 LSHMYFADVDFYDFEVFPNVFLNFETMMRDPFFYTFYKKFTDVFYQFKYYLKPYTQKDLV 507
Dvir_Lsplgamma  448 LSHMYFADVDYYDFEVFPNVFLNFETMLRDPFFYSFYKKFTDVFYKFKFFLGPYTKKDLY 507
                                .*.**. **.***.**  .* ***    ... *   *  **.  *

Dmel_Lsplalpha  540 MPGVTLKHVEVSELVTYFDLVDFDVTNMLNGKMVFHEGQFLWDKSLFARQMRLNHKPFSY 599
Dyak_Lsplalpha  539 MPGVTLKHVEVSELVTYFDMVDFDVTNMLNGKMIFHEGQFLWDKSLFARQMRLNHKPFSY 598
Dere_Lsplalpha  498 MPGVTLKHVEVSELVTYFDLVDFDVTNMLNGKMIFHEGQFLWDKSLFARQMRLNHKRFTY 557
Dsim_Lsplalpha  440 MPGVTLKHVEVSELVTYFDLVDFDVTNMLNGKMVFHEGQFVWDKSLFARQMRLNHKPFSY 499
Dmel_Lsplbeta   513 MPGVTMKNVEVSDLTTYFDLVDFDVTNMLNDKMIFQDGKFVWDMSLFARQMRLNHKPFTY 572
Dyak_Lsplbeta   515 MPGVTVKHVEVSDLTTYFDLVDFDVTSMLNDKMIFQDGKFVWDKSLFARQMRLNHKPFTY 574
Dere_Lsplbeta   515 MPGVTMKHVEVSDLTTYFDLVDFDVTSMLNDKMIFQDGKFVWDKSLFARQMRLNHKPFTY 574
Dsim_Lsplbeta   513 MPGVTMKHVEVSDLTTYFDLVDFDVTNMLNDKMIFQDGKFVWDMSLFARQMRLNHKPFTY 572
Dmoj_Lsplbeta   511 MPGVNLKNVEVSELVTYFDLVDFDVTNLMNDKMVFHDGKFVWDKSLYARQMRLNHKPFTY 570
Dbuz_Lsplbeta   511 MPGVTLKNVEVSELVTYFDLVDFDVTNXMNEKMVFHDGKFVWDKSLYARQMRLNHKPFSY 570
Dana_Lsplbeta   515 MPGVSMKHVEVSELTTYFDLVDFDVTNMLNDKMVFHDGKFVWDKSLYARQMRLNHKPFTY 574
Dpse_Lsplbeta   511 MPGVTMKHVEVSDLTTYFDLVDFDVTNMLNDKMIFHDGKFVWDKSLYARQMRLNHKPFTY 570
Dvir_Lsplbeta   511 MPGVTLKHVEVSELVTYFDLVDFDVTNLMNDKMVFHEGKFVWDKSLYARQMRLNHKPFTY 570
Dmel_Lsplgamma  507 YEGITIKDVSVSKLVTYYDIVDFDVTNLLNDKMTFVDGQYIWDKALLARQARLNHKPFNF 566
Dyak_Lsplgamma  507 YEGITIKDVSVSKLVTYYDIVDFDVTNLLNDKMTFVDGQYVWDKALLARQARLNHKPFNF 566
Dere_Lsplgamma  507 YEGITIKDVSVSKLVTYYDIVDFDVTNLLNDKMTFVDGQYVWDKALLARQARLNHKPFNF 566
Dsim_Lsplgamma  507 YEGITIKDVSVSKLVTYYDIVDFDVTNLLNDKMTFVDGQYIWDKALLARQARLNHKPFNF 566
Dmoj_Lsplgamma  508 YEGFKFTDVAVSKLETYFDFVDFDVTNLLNDKMTFVDGKFVWDKTLLARQARLNHKPFNF 567
Dbuz_Lsplgamma  508 YEGIQFKDVTVSKLETYFDFVDFDVTNLLNDKMTFVDGKFVWDKTLLARQARLNHKPFNF 567
Dana_Lsplgamma  506 YEGVTIKDVTVSKLVTYYDIVDFDVTNLLNDKMTFVDGQFVWDKALLARQARLNHKPFEF 565
Dpse_Lsplgamma  508 YEGIKIKDVSVSKLVTYYDIVDFDVTTLLNDKMTFVDGEFVWDKALLARQARLNHKPFDF 567
Dvir_Lsplgamma  508 YEGITIKDVSVSKLTTYFDFVDFDVTNLLNDKMTFVDGKFVWDKTLLARQARLNHKPFNF 567
                      *     * ** * **.* ******  .* ** * .* ..** .* *** ***** * .

Dmel_Lsplalpha  600 TYTIDSARDEKVVIRAFLGPKFDEYGRMISLTDNRMNFMEIDEFTYTLKTGSNLITRKST 659
Dyak_Lsplalpha  599 TYTIDSARDEKVVIRAFLGPKFDEFGRMISLTDNRFNFMEIDEFTYSLKTGSNLITRKSN 658
Dere_Lsplalpha  558 TYTIDSARDEQVVIRAFLGPKFDEFGRMISLADNRLNFMEIDEFTYTLKTGSNLITRKST 617
Dsim_Lsplalpha  500 TYTIDSARDEKVVIRAFLGPKFDEYGRMISLADNRMNFMEIDEFTYTLKTGSNLITRKST 559
Dmel_Lsplbeta   573 TYTIESEKVEKVVIRAFLGPKFDEFGKVISLAENRMNFMEIDEFYYELKAGTNKITRKSS 632
Dyak_Lsplbeta   575 TYTIESEKVEKVVIRAFLGPKFDEFGKVISLAENRMNFMEIDEFYYELKAGTNMVTRKSS 634
Dere_Lsplbeta   575 TYTIESEKVEKVVIRAFLGPKFDEFGKVISLAENRMNFMEIDEFYYELKAGTNMVTRKSS 634
Dsim_Lsplbeta   573 TYTIESEKVEKVVIRAFLGPKFDEFGKVISLAENRMNFMEIDEFYYELKAGTNMITRKSS 632
Dmoj_Lsplbeta   571 TYTIESDKPEKVVLRAFIGPKFDEFGRMISLTENRMNFMEIDEFYYELKAGTNMITRKSN 630
Dbuz_Lsplbeta   571 TYTVESDKPEKVVVRAFLGPKYDEFGRMISLTENRMNFMEIDEFYYELKAGTNMITRKSN 630
Dana_Lsplbeta   575 TYTIESDKVEKVVVRAFLGPKFDEFGKMISLTENRMNFMEIDEFFYELKAGTNMIKRNSN 634
Dpse_Lsplbeta   571 TYTIESDKVEKVVVRAFLGPKFDEFGKMISLTENRMNFMEVDEFYYDLKAGSNVITRKSN 630
Dvir_Lsplbeta   571 TYTVESDKPEKVVVRAFLGPKFDEFGRVISLAENRMNFMEVDEFYYELKAGSNMITRNSK 630
Dmel_Lsplgamma  567 EFTIDSDKVQKGVVRVFLGPKFDEYGRVIPLDYNRKNFVQIDSFVYPFIAGTNTIKRSSK 626
Dyak_Lsplgamma  567 EFTIESDKVQKGVVRVFLGPKFDEYGRVIPLDYNRKNFVQIDSFVYPFIAGTNTIKRSSK 626
Dere_Lsplgamma  567 EFTIESEKVQKGVVRVFLGPKFDEYGRVLPLDYNRKNFVQIDSFVYPFIAGTNTIKRSSK 626
Dsim_Lsplgamma  567 EFTIESDKVQKGVVRVFLGPKFDEYGRVIPLDYNRKNFVQIDSFVYPFIAGTNTIKRTSK 626
Dmoj_Lsplgamma  568 EFNIVSDKVQKGVVRVFLGPKFDEYGRVIPLKYNSQNFVQIDSFVYPFVAGQNTIKRSSK 627
Dbuz_Lsplgamma  568 EFVIESDKVQKGVVRIFLGPKFDEYGRVIPLKYNSQNFVQIDSFVYPFVAGQNTIKRSSK 627
Dana_Lsplgamma  566 EFTIESTQAQKGVVRVFLGPKFDEYGQVIPLNYNRKNFVQIDSFVYPFVAGVNTFKRSSK 625
Dpse_Lsplgamma  568 EFTIDSDKVQKGVVRVFLGPKFDEYGRVIPLVYNSKNFVQIDSFVYPFVAGSNTIKRSSK 627
Dvir_Lsplgamma  568 EFTIDSAKEQKGVVRVFLGPKFDEFGRVIPLKYNGQNFVQIDSFVYPFVAGQNTIKRSSK 627
                      . . * . .. *.* *.***.**.*... *  *  **...* * *   .* *   * *
```

```
Dmel_Lsplalpha  660 DFAWTVKDRTTYTELYYYTMMAFDGKYDYPLDLTEPHCGFPDRLVLPMGWKKGMPMQMFF 719
Dyak_Lsplalpha  659 DFAWTVKDRTTYTELYYYTMMAFDGKYDYPMDLTEPHCGFPDRLVLPMGWKKGMPMQMFF 718
Dere_Lsplalpha  618 DFAWTVKDRTTYTELYYHTMMAIDGKYDYPMDLTEPHCGFPDRLVLPMGWKKGMPMQMFF 677
Dsim_Lsplalpha  560 DFAWTVKDRTTYTELYYYTMMAFDGKYDYPLDLTEPHCGYPDRLVLPMGWKKGMPMQMFF 619
Dmel_Lsplbeta   633 EFYWTVKDRTTYTELYYYTMMAFDGKYDFPLDISEPHCGFPDRLVLPMGWQKGMPMQMFF 692
Dyak_Lsplbeta   635 EFYWTVKDRTTYSELYYYMMMAFDGKYDFPLDISEPHCGFPDRLVLPMGWQKGMPMQMFF 694
Dere_Lsplbeta   635 EFYWTVKDRTTYTELYYYMMMAFDGKYDFPLDISEPHCGFPDRLVLPMGWQKGMPMQMFF 694
Dsim_Lsplbeta   633 EFYWTVKDRTTYTELYYYTMMAFDGKYDFPLDISEPHCGFPDRLVLPMGWQKGMPMQMFF 692
Dmoj_Lsplbeta   631 DFYWTVKDRTTYSELYYYVMMAFDGKYDFPLDISEAHCGFPDRLILPMGWKKGMPMQMFF 690
Dbuz_Lsplbeta   631 DFYWTVKDRTTYSELYYYVMMAFDGKYDFPLDISEAHCGFPDRLILPMGWKKGMPMQMFF 690
Dana_Lsplbeta   635 EFYWTVKDRTTYTELYYYLMMAFDGKYDFPLDISEPHCGFPDRLVLPMGWKKGMPMQMFF 694
Dpse_Lsplbeta   631 DFYWTVKDRTTYTELYYYVMMAFDGKYDFPLDISEPHCGFPDRLVLPMGWKKGMPMQMFF 690
Dvir_Lsplbeta   631 DFFWTVKDRTSYSELYYYTMMAFDGKYKFPLDISEPHCGFPDRLILPMGWKKGMPMQMFF 690
Dmel_Lsplgamma  627 EFSWTAEDRITYTELYKYVMLASEGKYDFPLDISEPHNAFPDRLVLPKGWEQGMPMQFYF 686
Dyak_Lsplgamma  627 EFSWTAEDRITYTELYKYVMLASEGKYDFPLDISEPHNAFPDRLVLPKGWEQGMPMQFYF 686
Dere_Lsplgamma  627 EFSWTAEDRITYTELYKYVMLASEGKYDFPLDISEPHNAFPDRLVLPKGWEQGMPMQFYF 686
Dsim_Lsplgamma  627 EFSWTAEDRITYTELYKYVMLASEGKYDFPLDISEPHNAFPDRLVLPKGWEQGMPMQFYF 686
Dmoj_Lsplgamma  628 EFTWTAEDRITYTELYKYVMLASEGTYDFPLDLSEPHNAFPDRLVLPRGSEEGMPMQFYF 687
Dbuz_Lsplgamma  628 EFTWTAEDRITYTELYKYVMLASEGTYDFPLDVSEPHNAFPDRLVLPRGWEEGMPMQFYF 687
Dana_Lsplgamma  626 EFSWTSEDRITYTELYKYVMLASEGKYDFPLDISEPHNAFPDRLVLPKGWEQGMPMQFYY 685
Dpse_Lsplgamma  628 EFSWTAEDRTTYTELYKYVMLATEGKYDFPLDISEPHSAFPDRLVLPKGWESGMPMQFYF 687
Dvir_Lsplgamma  628 EFTYTAEDRATYTELYKYVQLASEGKYKFPLDISEPHNAFPDRLVLPRGWEQGLPMQFYF 687
                        .*  *  ** .*.***    .* .* * .*.*..* *  .****.** *   *.*** ..

Dmel_Lsplalpha  720 MVVPYMAP-QHEQFSTFDYTYSCGIGSGARHVDSLPFGYPFDREINEYEFHVPNMYFKDV 778
Dyak_Lsplalpha  719 MVVPYMAP-QHEQFTTFDYTYSCGIGSGARHVDSLPFGYPFDREIDEYEFYVPNMYFKDV 777
Dere_Lsplalpha  678 MVVPYMAP-KHEQFSTFDYSYSCGIGSGARHVDSLPFGYPFDREINEYEFYVPNMYFKDV 736
Dsim_Lsplalpha  620 MVVPYMAP-QHEQFSTFDYTYSCGIGSGARHVDSLPFGYPFDREINEYEFHVPNMYFKDV 678
Dmel_Lsplbeta   693 MVVPYVAP-AHEQFSTFDYTYSCGIGSGARYVDSLPFGYPFDRAIDEYEFFVPNMYFKDV 751
Dyak_Lsplbeta   695 MVVPYVAP-AHEQFSTFDYTYSCGIGSGARFVDSMPFGYPFDREIDEYEFFVPNMYFKDV 753
Dere_Lsplbeta   695 MVVPYVAP-AHEQFSTFDYTYSCGIGSGARFVDSMPFGYPFDREIDEYEFFVPNMYFKDV 753
Dsim_Lsplbeta   693 MVVPYVAP-AHEQFSTFDYTYSCGIGSGARYVDSLPFGYPFDREIDEYEFFVPNMYFKDV 751
Dmoj_Lsplbeta   691 MVMPYSAP-KHEQFSTFDYTYSCGIGSGARFIDDMPFGYPFDREIDEYEFYVPNMYFKDV 749
Dbuz_Lsplbeta   691 MVMPYNAPTKHEQFSTFDYTYSCGIGSGARFIDDMPFGYPFDREIDEYEFYVPNMYFKDV 750
Dana_Lsplbeta   695 MVMPYSAP-THEQFSTFDYTYSCGIGSGARFVDNMPFGYPFDREIDEYEFFVPNMYFKDV 753
Dpse_Lsplbeta   691 MVMPYKAP-SHEQFSTFDYTYSCGIGSGARFVDSMPFGYPFDRQIDEYEFFVPNMYFKDV 749
Dvir_Lsplbeta   691 MVMPYTAPSTHEQFSNFDYTYSCGIGSGARFIDDLPFGYPFDREIDEYEFFVPNMYFKDV 750
Dmel_Lsplgamma  687 FVSPFAET--YEQFSNFDYTYSSGVGSGTRFVDTKPFGYPFDRQIDESDFFVPNGFFKDV 744
Dyak_Lsplgamma  687 FVSPFAEE--YEQFSNFDYTYSSGVGSGTRFVDTKPFGYPFDRQIDESDFFVPNGFFKDV 744
Dere_Lsplgamma  687 FVSPFAEE--YEQFSNFDYTYSSGVGSGTRFVDTKPFGYPFDRQIDETDFFVPNGFFKDV 744
Dsim_Lsplgamma  687 FVSPFAET--YEQFSNFDYTYSSGVGSG                                 712
Dmoj_Lsplgamma  688 FVSPYTEN--YEQFSNFDYTYQSGVGSGTRFVDSKPFGYPFDRSAEEYSYFVPNAFFKDV 745
Dbuz_Lsplgamma  688 FVSPYTED--YEQFSNFDYTYQSGVGSGTRFVDSKPFGYPFDRSADEYSYFVPNAYFKDV 745
Dana_Lsplgamma  686 FVSPYAEE--YEQFSTFDYTYASGVGSGTRFVDSKPFGYPFDRQIDEYDFFVPNGFFQDV 743
Dpse_Lsplgamma  688 FVSPYTAD--YEQFSNFDYTYSSGVGSGTRYVDSKPFGYPFDRQIDEYTFFVPNGYFKDV 745
Dvir_Lsplgamma  688 FVSPYTED--YEQFSNFDYAYQAGVGSGTRFVDSKPFGYPFDRSVDEYSFFVPNGYFKDV 745
                     * *.      ***..***.* .*.***

Dmel_Lsplalpha  779 TIYHADTMEKYYNYKEYTNYGHFDYSFFNDYYTKYFKL  816
Dyak_Lsplalpha  778 TIYHADAMEKYYNYKEYSNYGHFDYTFFSDYYTKYFKL  815
Dere_Lsplalpha  737 TIYHADTMEKYHNYKEYSNYGHFDYSFFNDYYTKNFKL  774
Dsim_Lsplalpha  679 TIYHADTMEKYYNYKEYSNYGHFDYSFFNDYYTKYFKL  716
Dmel_Lsplbeta   752 SIYHADTMEPYYKYKSYSNYGHFDYTFFNDYYTKYFKF  789
Dyak_Lsplbeta   754 SIYHADTMEPYYKYKSYSNYGHFDYAFFNDYYTKYFKF* 792
Dere_Lsplbeta   754 SIYHADTMQPYYKYKSYSNYGHFDYAFFNDYYTKYFKF* 792
Dsim_Lsplbeta   752 SIYHADTMEPYYKYKSYSNYGHFDYAFFNDYYTKYFKF  789
Dmoj_Lsplbeta   750 PIYHIDTMEPYYKYKGYTNYGQFDYTFFRDYYTKYFKH  787
Dbuz_Lsplbeta   751 PIYHVDTMEPYYKYKGYTNYGQFDYTFFRDYYTKYFKF  788
Dana_Lsplbeta   754 KIYHADTMEPYYKYKGYSNYGHFDYTFFSDYYTKYFKY* 792
Dpse_Lsplbeta   750 TIFHTDTMEPYYKYKGYTNYGHFDYTFFRDYYTKYFKL* 788
Dvir_Lsplbeta   751 PIYHVDTMEPYYKYKGYTNYGHFDYTFFHDYYTKHFKF* 789
Dmel_Lsplgamma  745 KVYYVDTFAKYFEKK-YTQFGTFDYSIEY           772
Dyak_Lsplgamma  745 KVYYVDTFAKYFEKK-YAQFGSFDYSIEY           772
Dere_Lsplgamma  745 KVYYVDTFAKYFEKK-YAQFGTFDYSIEY*          773
Dsim_Lsplgamma  713                                       712
Dmoj_Lsplgamma  746 KIYYVDQFAKYFEKK-YDNYGTFDYSIVY*          774
Dbuz_Lsplgamma  746 KIYYVDQFAKYFEQK-YANYGTFDYSIVY           773
Dana_Lsplgamma  744 KIYYVDTFAKYFEKK-YPKFGTFDYSIEY*          772
Dpse_Lsplgamma  746 KIFFVDTFAKYFEKK-YAQFGTFDYSYFEF*         775
Dvir_Lsplgamma  746 KIYYVDTFAKYFEKK-YENYGVFDYSVEY*          774
```

10.8 Raw real-time RT-PCR data

Gene	Sample	Crossing Point[1]	Mean[2]
		21.00	
		20.97	
	fat body cDNA	20.93	20.97 ± 0.04
		20.60	
		20.56	
	whole larval cDNA	20.70	20.62 ± 0.08
	fat body RNA	>41.00	
	whole larval RNA	37.80	
rp49	negative	37.58	
		21.69	
		21.86	
	fat body cDNA	21.88	21.81 ± 0.12
		23.03	
		22.95	
	whole larval cDNA	22.95	22.98 ± 0.05
	fat body RNA	35.98	
	whole larval RNA	35.10	
Lsp1α	negative	35.21	
		23.91	
		24.49	
	fat body cDNA	24.58	24.33 ± 0.41
		25.09	
		25.29	
	whole larval cDNA	25.20	25.19 ± 0.11
	fat body RNA	38.95	
	whole larval RNA	38.74	
Lsd-2	negative	>41.00	
		31.25	
		32.21	
	fat body cDNA	31.16	31.54 ± 0.66
		26.66	
		26.75	
	whole larval cDNA	26.54	26.65 ± 0.12
	fat body RNA	37.79	
	whole larval RNA	26.63	
CG2560	negative	40.82	
CG2556		22.06	
		22.43	
	fat body cDNA	22.47	22.32 ± 0.26

[1] The cycle number at which the fluorescence level of each sample reaches a set threshold level, *i.e.* greater than background level.

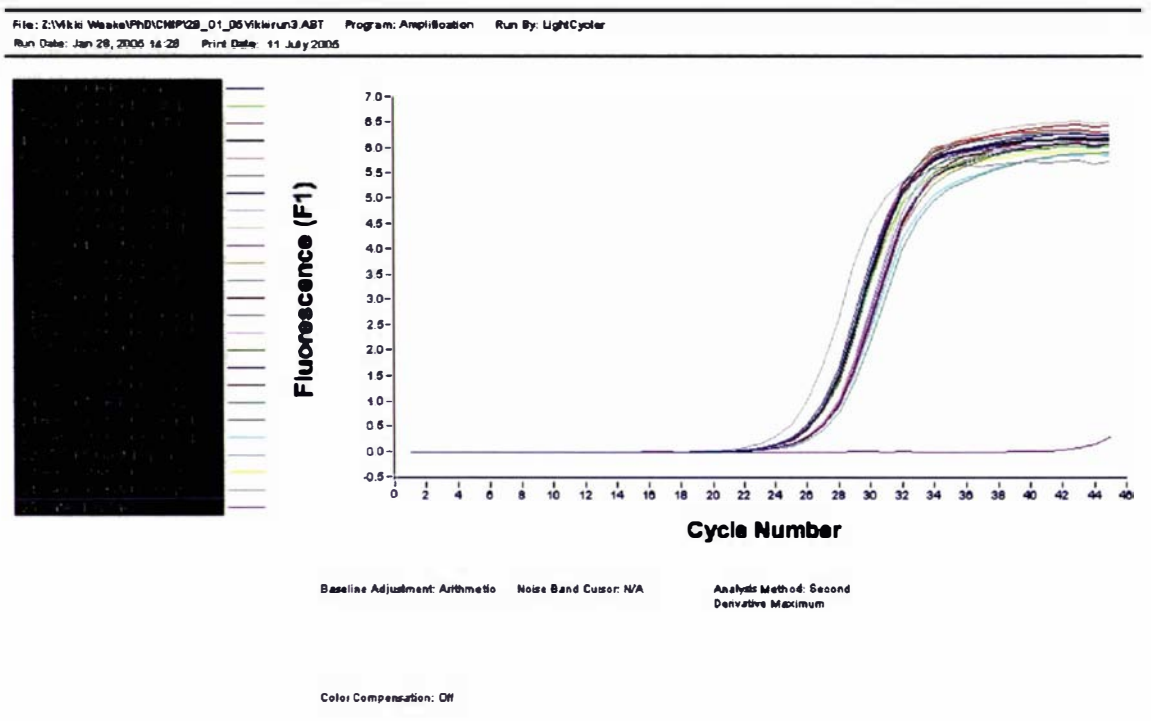
[2] 95% confidence intervals are indicated for 3 replicates of each PCR reaction.

		22.02	
		22.23	
	whole larval cDNA	22.35	22.20 ± 0.19
	fat body RNA	35.79	
	whole larval RNA	26.01	
	negative	36.12	
		25.01	
		24.88	
	fat body cDNA	24.87	24.92 ± 0.09
		24.55	
		24.58	
	whole larval cDNA	24.57	24.57 ± 0.02
	fat body RNA	32.45	
	whole larval RNA	24.43	
GPDH	negative	32.09	

10.9 Raw chromatin immunoprecipitation data

10.9.1 Example graph showing fluorescence intensity over the PCR reaction

Data are shown for the Lspla-2 primer set. The negative control is represented by the flat line. This trace is typical of those obtained in these experiments.



10.9.2 Chromatin immunoprecipitation crossing points for each PCR reaction

		Pgd-5		Pgd-10		Lspla-1		Lspla-2		Lsd-2	
Fly Line		CP[1]	mean	CP	mean	CP	mean	CP	mean	CP	mean
P-lacZ:9B4	ChIP[2]	~~37.6~~[3]	24.20	22.84	22.97	26.63	26.71	26.44	26.44	24.26	24.45
		24.21	±	22.98	±	26.81	±	26.32	±	24.56	±
		24.19	0.02[4]	23.09	0.14	26.68	0.11	26.56	0.14	24.52	0.18
	input	25.98	25.89	25.31	25.61	26.19	26.09	26.33	26.42	25.63	25.72
		25.82	±	25.86	±	26.01	±	26.49	±	25.72	±
		25.87	0.09	25.65	0.31	26.07	0.10	26.43	0.09	25.81	0.10
	ChIP	25.08	25.13	23.46	23.76	27.13	27.21	27.10	27.03	25.24	25.18
		25.17	±	23.79	±	27.20	±	26.97	±	25.13	●
		25.13	0.05	24.03	0.32	27.30	0.10	27.02	0.07	25.16	0.06
	input	26.14	26.18	25.56	26.05	26.00	26.16	26.05	26.23	25.67	25.71
		25.90	±	26.07	±	26.07	±	26.48	±	25.84	●
		26.50	0.34	26.52	0.54	26.42	0.25	26.17	0.25	25.62	0.13
P-lacZ:19E7	ChIP	24.96	25.02	22.96	23.54	27.64	27.67	27.14	27.18	25.33	25.23
		25.03	±	23.77	±	27.76	±	~~24.96~~	±	25.15	±
		25.08	0.07	23.88	0.57	27.60	0.09	27.22	0.06	25.22	0.10
	input	25.90	25.86	25.96	25.38	26.24	26.24	26.23	26.32	25.71	25.78
		25.81	±	25.21	±	26.28	±	26.43	±	25.81	±
		25.88	0.05	24.96	0.59	26.20	0.05	26.29	0.12	25.82	0.07
	ChIP	25.05	25.06	23.27	23.55	27.70	28.00	27.31	27.15	24.94	25.13
		25.12	±	23.66	±	28.10	±	27.07	±	25.23	±
		25.01	0.06	23.72	0.28	28.21	0.30	27.08	0.15	25.22	0.19
	input	26.25	26.31	25.93	25.94	27.19	26.79	26.97	27.03	26.55	26.52
		26.44	●	26.05	●	26.46	●	27.09	±	26.52	±
		26.24	0.13	25.84	0.12	26.73	0.42	27.03	0.07	26.49	0.03
	neg[5]	35.81		37.05		>41.00		36.20		29.76	
y w	ChIP	25.15	25.47	23.85	24.01	28.30	28.41	27.06	26.73	24.69	24.69
		25.65	±	23.78	±	28.27	●	26.84	±	24.69	●
		25.61	0.31	24.39	0.38	28.65	0.24	26.30	0.44	24.69	0.00
	input	26.20	26.14	25.94	25.98	25.81	25.82	26.06	26.08	25.14	25.29
		26.02	±	25.91	●	25.83	●	26.12	●	25.33	±
		26.21	0.12	26.10	0.12	25.82	0.01	26.06	0.04	25.41	0.16
	neg	37.89		36.96		37.14		36.77		36.92	
	ChIP	25.24	25.25	24.27	24.52	28.88	29.02	27.45	27.36	25.30	25.23
		25.44	±	24.46	±	29.44	±	27.33	±	25.24	●
		25.07	0.21	24.83	0.32	28.73	0.42	27.30	0.09	25.15	0.09
	input	25.59	25.63	25.78	25.97	26.69	26.77	26.19	26.35	25.30	25.43
		25.58	±	25.99	±	26.87	±	26.43	●	25.49	±
		25.71	0.08	26.14	0.20	26.75	0.10	26.44	0.16	25.51	0.13
	neg	37.07		35.59		38.21		36.77		36.56	

[1] Crossing Point: The cycle number at which the fluorescence level of each sample reaches a set threshold level, *i.e.* greater than background level.

[2] Immunoprecipitated DNA.

[3] Discarded values are indicated by double strike-through lines.

[4] 95% confidence intervals are indicated for 3 replicates of each PCR reaction.

[5] Negative control (water).

		lacZ		GPDH		CG2556-3'UTR		CG2556-5'UTR	
Fly Line		CP	mean	CP	mean	CP	mean	CP	mean
P-lacZ:9B4	ChIP	24.31 24.45 22.75	24.38 ± 0.11	24.57 24.45 24.48	24.50 ± 0.07	25.69 25.75 25.60	25.68 ● 0.09		
	input	25.32 25.50 25.54	25.45 ± 0.13	23.80 23.57 23.73	23.70 ± 0.13	26.16 25.93 26.20	26.10 ± 0.16		
	ChIP	24.81 24.88 24.83	24.84 ± 0.04	24.95 24.80 24.71	24.82 ± 0.14	25.93 26.76 26.95	26.55 ± 0.61		
	input	25.56 25.52 25.31	25.46 ± 0.15	23.83 23.71 23.82	23.79 ± 0.08	26.12 26.53 25.76	26.14 ± 0.44		
P-lacZ:19E7	ChIP	24.16 24.47 24.45	24.36 ± 0.20	24.74 24.85 24.60	24.73 ● 0.14			26.74 ~~28.05~~ 26.66	26.70 ± 0.06
	input	25.25 25.45 25.42	25.37 ± 0.12	23.59 23.73 23.75	23.69 ± 0.10			26.04 25.94 25.57	25.85 ± 0.28
	ChIP	24.21 24.33 24.26	24.27 ± 0.07	24.68 24.64 24.81	24.71 ± 0.10			26.90 26.90 ~~25.95~~	26.90 ± 0.00
	input	25.87 25.80 25.68	25.78 ± 0.11	24.33 24.26 24.41	24.33 ± 0.08			~~25.99~~ 26.65 26.54	26.60 ± 0.09
	neg	30.22		37.06				35.85	
y w	ChIP			24.43 24.43 24.24	24.37 ± 0.12				
	input			23.44 23.45 23.32	23.40 ± 0.08				
	neg			32.48					
	ChIP			24.91 24.86 24.80	24.86 ± 0.06				
	input			23.25 23.25 23.18	23.23 ± 0.05				
	neg			33.05					