# A STUDY OF *CIS*-ACTING ELEMENTS REQUIRED FOR DOSAGE COMPENSATION IN *DROSOPHILA MELANOGASTER*

A thesis presented in partial fulfilment of the requirements
for the Degree of Master of Science in Genetics
at Massey University, Palmerston North,
New Zealand

**Rebecca Ann Henry**

**2000**

# ABSTRACT

Dosage compensation (the equalisation of X-linked gene products) occurs in *Drosophila melanogaster* by a two fold transcriptional up-regulation of X-linked gene expression in males. This involves the binding of five proteins, MSL-1, MSL-2, MSL-3, MLE, MOF, and potentially an RNA (*roX1* or *roX2*), to hundreds of sites along the male X chromosome. The *cis*-acting X-linked DNA sequences required for dosage compensation (called dosage compensation regulatory elements or DCREs) remain elusive, despite numerous attempts of identify them. An insulated reporter gene assay system has been developed to minimise problems previously encountered with identification of these elements. The reporter system consists of the constitutive *armadillo* promoter fused to the *lacZ* reporter gene (called *arm-lacZ*). This reporter construct is flanked by SCS/SCS' insulator elements to block potential repressive effects of an autosomal chromatin environment.

The role of the *roX* genes during dosage compensation was investigated. Initially both the *roX1* and *roX2* RNAs were expressed from within the *arm-lacZ* insulated system. Expression of either RNA lead to a significant increase in *lacZ* expression in males, although consistently less than two-fold. These results suggested that either the MSL complex was binding to the *roX* genes or the expression of the *roX* RNAs in *cis* lead to male-specific hypertranscription of *lacZ*. To test these possibilities *roX1* and *roX2* cDNAs were inserted into the *arm-lacZ* reporter. Insertion of either cDNA lead to a significant increase in *lacZ* expression in males, suggesting that the transcribed regions of the *roX* genes contain binding site(s) for the MSL complex. Interestingly the level of *lacZ* hypertranscription in males was significantly higher in homozygous *roX1* cDNA lines than homozygous *roX1* gene lines. This may indicate that too high a local concentration of *roX1* RNA has a dampening effect on the level of hypertranscription meditated by the MSL complex. In a set of experiments designed to identify the MSL binding site(s) in *roX1*, two regions of the cDNA sequence were amplified and inserted into the *arm-lacZ* system. One of these fragments, containing a proposed DNAseI hypersensitivity site and possible GAGA binding sites, increased *lacZ* expression in males, but to levels lower than the entire cDNA. This suggests there may be more than one MSL biding site in *roX1*.

A second method of dosage compensation is thought to occur in *Drosophila*, independently of the MSL proteins. The *arm-lacZ* insulated reporter system was used to investigate the hypothesis that some genes may be dosage compensated due to repression by *Sex-lethal (Sxl)* in females. Several genes have been found to contain three or more Sxl binding sites in their 3' UTRs, with some also carrying Sxl binding sites in the 5' UTR. Fragments from the *Sxl, Cut* and *Small Forked* genes, containing numerous Sxl binding sites from the 3' UTR, were inserted into the 3' UTR region of *arm-lacZ*. Males carrying autosomal insertions of the construct had on average 1.07 – 1.50 times the level of β-galactosidase in females. This suggests that some genes could be partially compensated through Sxl repression in females.

In addition to inserting 3' UTR fragments into *arm-lacZ,* a synthetic oligonucleotide containing a long Sxl binding site was inserted into the 5' region of an *arm-lacZ* construct already carrying the *Runt* 3' UTR fragment. Males carrying autosomal insertions of the construct had levels of β-galactosidase activity similar to those lines carrying autosomal insertions of the 3' UTR fragments alone. This suggests that other factors such as RNA binding proteins or RNA secondary structure may be required in order to obtain efficient translation repression by Sxl.

Finally three X-linked DNA fragments, from the 1C region, were inserted individually between the SCS' element and the *armadillo* promoter. If the X-linked fragment contained a DCRE then males carrying autosomal insertions of the construct would produce twice the β-galactosidase activity of females. However, males and females expressed the same levels of *lacZ*.

# ACKNOWLEDGEMENTS

I would first like to give a big thank you to my supervisor Dr. Max Scott for all his advice and encouragement even when nothing seemed to be going right. I really appreciate it.

My next big thanks goes to Anton, without whom I would never have got through all of this. You were a great support for me and so totally understanding when my constant excuse for not being able to go anywhere or do anything seemed to be that my flies needed watering!

To Mum and Dad for understanding that the orange-eyed flies were GOOD and when I found one it was cause for celebration. Thank you for putting up with my attempts to explain what I was studying. Your patience was admirable!

To Cate for helping to keep me sane in that first year and always going for a walk rain or shine when I needed a break from studying. To Sarah, Victoria, Mikey and Dougal for being good friends and not saying too much about me working with flies!

Thank you to my friends at Massey. To Beccy and Bek for letting me loiter in your lab and giving me someone to talk to. Beccy I really appreciate the help you gave me in the injection room. To Sheralee for teaching me new techniques, even when you were rushed off your feet. To Annika and so many others, thank you for your friendship and all the fun times we have had.

Thank you to everyone in the Fly Spot, past and present – Max, Lewis, Helen, Xuelei and Jörg. You all gave me time out of busy schedules to help when I needed it.

Thank you to Massey University for the Massey Masterate Scholarship, which made my postgrad years so much easier.

# ABBREVIATIONS

| | |
|---|---|
| β | beta |
| Δ | delta |
| λ | lambda |
| °C | degrees Celsius |
| ATP | adenine triphosphate |
| bp | base pairs |
| BSA | bovine serum albumin |
| cDNA | complementary DNA |
| DIG | digoxigenin |
| DNA | deoxyribose nucleic acid |
| DNAse | deoxyribonuclease |
| dNTPs | dinucleotide triphosphates |
| F | female |
| g | gram |
| L | litre |
| kb | kilobase pairs |
| μ | micro |
| m | milli |
| M | male or molar |
| mRNA | messenger RNA |
| nt | nucleotide pairs |
| OD | optical density |
| PCR | polymerase chain reaction |
| RNA | ribonucleic acid |
| RNase | ribonuclease |
| rpm | revolutions per minute |
| U | unit |
| UTR | untranslated region |
| UV | ultra violet |
| v/v | volume per volume |
| w/v | weight per volume |

# TABLE OF CONTENTS

***Errata***

The data values listed in tables within this thesis are correct, however due to calculation difficulties values quoted within the main body of text may vary slightly from those in the tables.

# LIST OF FIGURES

# LIST OF TABLES

# 1.0 INTRODUCTION

## 1.1 DOSAGE COMPENSATION - AN OVERVIEW

Dosage compensation is the mechanism by which the expression of X-linked genes is equalised between males with one X chromosome and females with two. Different organisms have evolved unique mechanisms to achieve dosage compensation. In mammals one female X chromosome is randomly inactivated to equal the expression of the single male X chromosome (Lyon, 1961). In *Caenorhabditis elegans* the expression of both female X chromosomes is down regulated to equal the expression of the single male X chromosome (Hsu and Meyer, 1993). Mammals and *C. elegans* are both examples of organisms where the male is the heterogametic sex. In organisms where the female is heterogametic (ZW), eg. birds and butterflies, dosage compensation has been shown not to occur (Baverstock *et al.*, 1982; Johnson and Turner, 1979).

## 1.2 DOSAGE COMPENSATION IN DROSOPHILA

### 1.2.1 Dosage Compensation Involves Transcriptional Up Regulation

In *Drosophila* dosage compensation is achieved by the hypertransactivation (transcriptional up-regulation) of genes on the single male X chromosome to equal the level of expression from two female X chromosomes. Early support for this statement came from Offermann's, (1936) observation in squashes of polytene chromosomes, from larval salivary glands, that the male X chromosome is wider and more diffuse in appearance than both female X chromosomes. This 'puffier' appearance indicates an increase in gene expression. Following this observation Mukherjee and Beermann, (1965) demonstrated that incorporation of tritiated ($^3$H) uridine into nascent salivary gland transcripts was significantly higher in the single unpaired male X chromosome than one female X chromosome. Results from these experiments pointed to an enhancing effect occurring on the male X chromosome rather than a repressing effect on the female.

### 1.2.2 Histone Acetylation, Transcriptional Activity and Dosage Compensation

The core particle of the nucleosome consists of four histones H2A, H2B, H3 and H4. Acetylation of the histones occurs at specific lysine residues in the N-terminal domain and is a ubiquitous post-translational modification found in all animal and plant species (Turner, 1991).

Histone acetylation has been stated as being potentially a major influence on transcription and DNA packaging through the cell cycle (Turner, 1991). Histones in actively transcribing genes are rapidly acetylated and deacetylated, which proposes a link between transcriptional activation and histone acetylation (reviewed by Turner, 1991). Neutralisation of positive charges by acetylation of histone H4 is thought to play a primary role in altering interactions between the DNA and histones, which may mediate enhanced binding of transcription factors to their DNA target sequences (Vettese-Dadey *et al.*, 1996).

Acetylation of ε-amino groups of lysine residues, present in the N-terminal domain of the core histones, is most strongly linked with transcriptional activity (Turner, 1991). Vettese-Dadey *et al.*, (1996) demonstrated that the highly acetylated histone H4 in nucleosome cores has the highest affinity for transcription factors USF and GAL-4H. Studies of the *Saccharomyces cerevisiae* mating type (MAT) locus (Johnson *et al.*, 1990) indicates that the repression of the silent mating loci requires histone acetylation. When lysine 16 is mutated to an arginine, which retains the positive charge, the regulation of the MAT locus is unaffected. But, when lysine 16 is mutated to a glutamine (a neutral amino acid mimicking acetylation) derepression of the locus occurs (Johnson *et al.*, 1990). A specific isoform of histone H4 acetylated at lysine 16 (H4Ac16) is also predominantly associated with the male X chromosome in *Drosophila* (Turner *et al.*, 1992). The acetylation of histone H4 on lysine 16 (H4Ac16) may play a role in loosening the chromatin structure and increasing the accessibility of transcription factors associated with the male X chromosome in *Drosophila* (Bone *et al.*, 1994).

## 1.3 THE TRANS-ACTING MALE SPECIFIC LETHALS

A simple model for dosage compensation in *Drosophila* would predict that increases in X-linked gene transcription result from the action of *trans*-acting factors upon target *cis*-acting sequences localised to the X chromosome (Palmer *et al.*, 1993). It has been rationalised that mutations inactivating regulatory genes responsible for dosage compensation could result in sex specific lethality (Lucchesi and Manning, 1987). A mutation that prevents normal compensation could cause the death of an individual with a single X chromosome due to a deficiency of X-linked gene products. Belote and Lucchesi, (1980a) carried out a large screen for ethyl methanesulfonate (EMS) induced sex-specific lethals on the 2$^{nd}$ and 3$^{rd}$ chromosomes. Three male-specific lethal mutations *male-specific lethal-1 (msl-1), male-specific lethal-2 (msl-2) and maleless (mle)* were discovered. Temperature sensitive mutants of *mle* had previously been isolated from natural populations of *D. melanogaster* (Fukunaga *et al.*, 1975; Golubovsky and Ivanov, 1972). These three genes plus the subsequently discovered *male-specific lethal-3 (msl-3)* (Lucchesi *et al.*, 1982) and *males-absent on the first (mof)* (Hilfiker *et al.*, 1997) have been collectively named the *male-specific lethals* or *msls.* Males mutant in any of these genes exhibit prolonged posthatching development and eventually die during the late larval or early pupal stages (Belote, 1983). These mutations have been shown to have no discernible effect on the viability and development of females (Belote and Lucchesi, 1980a).

Males homozygous for *msl-1, msl-2* or *mle* show a significant reduction in X-linked enzyme activities, while the levels of autosomal enzymes are not affected (Belote and Lucchesi, 1980b).

The MSL proteins bind to hundreds of sites along the entire length of the male X chromosome (Kuroda *et al.*, 1991; Palmer *et al.*, 1993). Immunolocalisation experiments show that the MSLs bind to the same sites along the X chromosome (Bone *et al.*, 1994), and the native X chromosome binding of any MSL protein requires the wildtype function of the other four MSLs (Bashaw and Baker, 1995; Gorman *et al.*, 1993; Gu *et al.*, 1998; Hilfiker *et al.*, 1994; Kelley *et al.*, 1995; Palmer *et al.*, 1994,). This dependent binding suggests that the MSL proteins form a heteromultimeric complex.

### 1.3.1 Maleless

As discussed above the *msl* genes have all recently been cloned. The *mle* gene (Kuroda *et al.*, 1991) encodes a polypeptide containing several short motifs characteristic of a superfamily of DNA and RNA helicases. MLE shows the highest sequence homology to a subfamily of RNA helicases containing DEAH box motifs (Schwer and Guthrie, 1991; Nakajima *et al.*, 1997). MLE shares 50% identity with human RNA helicase A (RHA) which mediates the interaction of CBP (CREB Binding Protein) with RNA polymerase II (Nakajima *et al.*, 1997). It has been proposed that the recruitment of CBP complexes may promote local unwinding of promoter DNA via RHA and allow access of transcriptional apparatus (Nakajima *et al.*, 1997). A study by Lee *et al.*, (1997) showed that MLE possesses NTPase and both RNA and DNA helicase activities and that these activities are essential functions of MLE for dosage compensation. Preliminary studies by Nakajima and Montminy (unpublished data cited by Nakajima *et al.*, 1997) have observed MLE associating with a 250 kDa CBP with histone acetylase activity. The unpublished data along with evidence that MLE appears to co-localise with acetylated histone H4 (Bone *et al.*, 1994), and has NTPase/helicase activity (Lee *et al.*, 1997) suggests that MLE may be involved in initiation of transcription, perhaps via chromatin remodelling of X-linked genes.

### 1.3.2 Male-Specific Lethal-1

The cloning and characterisation of the *msl-1* gene (Palmer *et al.*, 1993) showed that the MSL-1 protein is not closely related to any proteins in the current databases. It does however contain acidic regions in the N-terminus consisting of two extended aspartate and glutamate clusters, characteristic of proteins involved in chromatin modelling and transcription (Palmer *et al.*, 1993). The acidic regions of these proteins may provide a region of interaction with histones to mediate nucleosome assembly or release and thereby promote changes in chromatin structure and transcription (Palmer *et al.*, 1994; Turner *et al.*, 1992). MSL-1 protein is present in *mle* and *msl-3* mutant larvae, but is undetectable in *msl-2* mutant male larvae (Palmer *et al.*, 1994). This finding plus other genetic tests carried out by Palmer *et al.*, (1994) suggests that *msl-2* expression positively regulates the translation or stability of MSL-1 in males. MSL-1 also contains regions rich in proline, serine, threonine and glutamic acid which are residues (PEST sequences) associated with rapidly degraded proteins (Palmer *et al.*, 1993).

### 1.3.3 Male-Specific Lethal-2

MSL-1 and MSL-2 have been shown to co-immunoprecipitate from male nuclear extracts (Kelley *et al.*, 1995). MSL-2 (Zhou *et al.*, 1995), contains a RING finger, which is a $C_3HC_4$ zinc finger (Lovering *et al.*, 1993). Using a two-hybrid system Copps *et al.*, (1998) found that the RING finger domain of MSL-2 binds MSL-1. When residues clustered around the first zinc-binding site of the RING finger domain in MSL-2 were mutated interaction with MSL-1 was lost. In addition to the RING finger motif the MSL-2 protein also contains a positively and a negatively charged amino acid residue cluster and a coiled coil domain that may be involved in protein-protein interactions (Zhou *et al.*, 1995). Zhou *et al.*, (1995) hypothesise that MSL-2 may be a transcription regulator, with the positively charged amino acid cluster contributing to a DNA binding domain and the negatively charged cluster being part of a transcription *trans*-activator domain. Copps *et al.*, (1998) propose that the RING finger domain interaction with MSL-1, through the first zinc-binding site, may be an important prerequisite for subsequent protein-protein interactions and that the second zinc-binding site may have a second, but as yet unidentified activity.

MSL-1 and MSL-2 appear to form a core complex within the MSL complex. When either is removed through mutation the remaining MSL proteins fail to bind any site along the X chromosome (Gorman *et al.*, 1993; Gorman *et al.*, 1995; Palmer *et al.*, 1994; Lyman *et al.*, 1997). Conversely if MSL-3, MLE or MOF are removed MSL-1 and MSL-2 remain bound to 30 - 40 'high affinity' binding sites along the X chromosome (Palmer *et al.*, 1994; Gorman *et al.*, 1995; Gu *et al.*, 1998; Lyman *et al.*, 1997)

### 1.3.4 Male-Specific Lethal-3

Cloning and characterisation of the *msl-3* gene found that it encodes a novel protein (Gorman *et al.*, 1995). MSL-3 contains two chromatin organisation modifier (chromo) domains that are 30 - 50 amino acid domains conserved in several eukaryotic chromatin-binding proteins such as *Drosophila* hetereochromatin protein 1 (HP1) and Polycomb (PC) (Koonin *et al.*, 1995). Chromodomains have been implicated in the delivery of both positive and negative transcription regulators to chromatin targets.

### 1.3.5 Males-Absent on the First

Experimental results have identified an additional gene *males-absent on the first (mof)*, which encodes a putative histone acetyl transferase thought to be crucial for dosage compensation. This fifth male lethal gene was isolated by screening the X chromosome of *Drosophila melanogaster* for EMS-induced mutations, to identify genes carrying mutations that cause male specific lethality. Males mutant for *mof* die at the third instar larval stage of development, MSL-1, MSL-2 and MLE association with the X chromosome is reduced and the X-specific isoform of H4Ac16 is absent (Hilfiker *et al.*, 1997).

The *mof* mRNA encodes an 827 amino acid protein that contains a 250 amino acid domain common to many acetyl transferases and is shown to be required for binding of acetyl coenzyme A. This domain is found in proteins known to acetylate histones, such as histone acetyl transferase 1 of yeast (Kleff *et al.*, 1995) and histone acetyl transferase A of *Tetrahymena* (Brownell *et al.*, 1996). The mutation of *mof* results from a substitution of Gly691 (the most conserved residue in the 250 amino acid motif) for glutamic acid. This mutation leads to the absence of H4Ac16 on the male X chromosome and a male lethal phenotype. Lu *et al.*, (1996) showed that the mutation of the corresponding glycine to an aspartate, in the human spermidine/spermine acetyl transferase, abolishes enzyme activity. Recently it has also been demonstrated that MOF co-localises with the MSL complex on the male X chromosome using loss-of-function mutations (Gu *et al.*, 1998).

Immunolocalisation experiments have shown that all of the MSL proteins bind to hundreds of specific sites along the male X chromosome. Each of the MSLs is produced in both sexes except for MSL-2, which is absent in females (Zhou *et al.*, 1995). Henikoff and Meneely, (1993) suggest that MLE could catalyse the movement of the MSL complex along the nascent RNA. In *msl-1, msl-2* or *msl-3* mutant backgrounds MLE does not bind the X chromosome, but is still present, indicating that MSL-1, MSL-2 and MSL-3 are required for binding, but not for regulation of X chromosome expression (Gorman *et al.*, 1993). MOF, the fifth MSL protein, encodes a putative histone acetylase. Mutational studies provide strong evidence that MOF has histone acetyl transferase activity and is responsible for the histone acetylation involved in male specific hypertranscription of X-linked genes (Hilfiker *et al.*, 1997).

### 1.3.6 Histone Acetylation and MSL Localisation

The pattern of H4Ac16 distribution on the X chromosome is very similar to that of the MSLs suggesting a link between the signals required for localising these proteins to the male X chromosome (Bone *et al.*, 1994). Bone *et al.*, (1994) also observed that presence of this H4 isoform on the X chromosome requires the wildtype function of the *msl* genes. This suggests that the mechanism of dosage compensation involves histone acetylation through association with the MSL proteins.

## 1.4 NON-CODING RNA INVOLVEMENT IN DOSAGE COMPENSATION

Two new genes *roX1* and *roX2* (RNA on the X chromosome) have been isolated using an enhancer detector screen for β-galactosidase activity in the mushroom bodies of the *Drosophila* brain (Amrein and Axel, 1997; Meller *et al.*, 1997). Both genes are X-linked and each encodes an RNA without a significant open reading frame (ORF). Their expression is confined to the nucleus of male flies, which suggests that they may encode non-coding RNAs (Amrein and Axel, 1997; Meller *et al.*, 1997). Expression of *roX1* and *roX2* is dependent on the MSL complex (dosage compensation machinery) as neither of the genes are expressed in flies mutant for any of the *msls.* Additionally, expression of a *msl-2* transgene in females induces the expression of both *roX1* and *roX2* RNA (Amrein and Axel, 1997; Meller *et al.*, 1997). *In situ* hybridisation of *roX1* probes to late third-instar male larvae salivary gland X chromosomes displays a subcellular localisation of *roX1* RNA very similar to the localisation of the MSL complex binding the X chromosome (Amrein and Axel, 1997; Meller *et al.*, 1997).

Disruption of *roX1* produces no obvious phenotype, lethality, or developmental delay, which rules out *roX1* as an essential component of the dosage compensation complex. However the disrupted *roX1* mutant (the $roX1^{ex6}$ mutation removes the 5' half of the *roX1* gene and produces no stable RNA (Kelley *et al.*, 1999)) was used to show that *roX1* RNA could spread in *trans.* Kelley *et al.*, (1999) inserted a DNA fragment, containing the *roX1* gene, into either the second or third chromosome by P element mediated transformation. Males homozygous or hemizygous for the null $roX1^{ex6}$ mutation, but carrying one copy of the *roX1* transgene, were used for RNA *in situ* hybridisation to polytene chromosomes. These *in situ* experiments showed the

autosomally encoded RNA coating the entire X chromosome. This indicates that the *roX1* RNA still bound to the X despite being produced on another chromosome.

Kuroda *et al.*, (1991) proposed an RNA component of the dosage compensation system based on the observations that MLE contains RNA binding domains and is released from the X chromosome by RNase digestion (Richter *et al.*, 1996). However male flies mutant for *roX1* exhibit normal MLE binding and are fully dosage compensated (Meller *et al.*, 1997).

Franke and Baker, (1999) genetically produced a mutant with simultaneous loss of both *roX1* and *roX2*, which abolished binding of the MSL complex to the male X chromosome. They suggest this is a strong indication that the *roX* RNAs are integral components of a dosage compensation nucleoprotein complex and is consistent with the earlier proposal of Meller *et al.*, (1997) that there is a family of non-homologous and redundant genes including *roX1* and *roX2* that can compensate for the loss of one of its members. They also propose that *roX1* and its family members associate along the entire X chromosome to help change chromatin conformation and achieve hypertranscription, perhaps by associating with the MSLs, histone acetyl transferase or other chromatin constituents.

Comparisons have been made between the *roX* RNA and *Xist* RNA that coats the inactive mammalian X chromosome. *Xist* encodes a non-coding RNA expressed from the X-inactivation centre of the inactive X chromosome in mammals and is thought to 'spread' (in *cis*) along one of the female X chromosomes, remodelling chromatin to form a transcriptionally inactive Barr body (Lee *et al.*, 1996). There are some similarities between *roX1* and *Xist*; they are both nuclear and localised to a structurally modified X chromosome undergoing dosage compensation (Meller *et al.*, 1997).

## 1.5 REGULATION OF DOSAGE COMPENSATION

### 1.5.1 *Sex-Lethal*

The process of dosage compensation is one of several controlled by the 'master regulatory gene' *Sex-lethal (Sxl)*. Initially dosage compensation is controlled by the expression patterns of *Sxl,* which in turn is controlled by the ratio of X chromosomes (X) to autosomes (A). An X:A ratio of 1.0 (2X:2A) results in female development and a X:A ratio of 0.5 (1X:2A) results in male development. In *Drosophila* the X:A ratio acts to switch the *Sxl* gene into either the female mode which represents ON (functional) or the male mode which represents OFF (non-functional) (Cline, 1978). The X:A ratio itself is assessed by 'counting' genes, referred to as numerators and denominators (reviewed by Parkhurst and Meneely, 1994). These proteins are members of the helix-loop-helix (HLH) family of transcription factors (Parkhurst *et al.*, 1990). The numerators are a group of X chromosomal genes that behave as feminising elements because they increase the probability of activating *Sxl* expression. Lowering the number of numerators results in female lethality due to the lack of activated *Sxl,* whereas raising the number results in male lethality because *Sxl* is activated. Denominators are autosomally encoded genes acting as antagonists to the numerators by competing with numerators to form heterodimers. The heterodimers formed activate *Sxl* at the level of transcription (Keyes *et al.,* 1992).

The initial activation of *Sxl* results in production of *Sxl* mRNA transcripts from the early 'establishment' promoter $P_E$ in females. These early *Sxl* mRNA protein products specify the production of active female-specific transcripts from the late 'maintenance' promoter $P_L$ (Bell *et al.*, 1991) and thereby establish an autoregulatory feedback loop. Transcripts of *Sxl* are also produced in males from $P_L$, but these are truncated and inactive and maintained by default (Bell *et al.,* 1991; Keyes *et al.,* 1992).

In males a functional Sxl protein is missing, therefore male differentiation and dosage compensation occurs. In females, active Sxl protein acts upon the mRNA of *transformer* (*tra*), the next gene in the pathway. Sxl binding to *tra* RNA blocks a splice acceptor site, resulting in another female-specific splicing pattern occurring. The functional Tra protein is only produced in females and is involved in somatic sex

determination. *In vitro* studies by Samuels *et al.*, (1994) showed that Sxl protein binds to poly uridine (polyU) tracts in mRNA that consist of eight or more Us or $AU_7$.

### 1.5.2 *Sxl* Regulation of Dosage Compensation

*Sxl* loss-of-function mutations cause female lethality and gain-of-function mutations cause male lethality (Cline, 1978). Zhou *et al.*, (1995) demonstrated that the primary target of Sxl during dosage compensation is *msl-2*. The *msl-2* transcript is present in both males and females, with the same ORF, but the MSL-2 protein is present only in males (Bashaw and Baker, 1995; Kelley *et al.*, 1995; Zhou *et al.*, 1995). A small intron in the 5' UTR (untranslated region) of the *msl-2* transcript is spliced out in males and retained in females (Bashaw and Baker, 1995; Kelley *et al.*, 1995; Zhou *et al.*, 1995). Within this intron are poly(U) runs that resemble the Sxl binding sites found in *Sxl* and *tra* and are therefore spliced out in males. Four more Sxl binding sites are present in the 3' UTR that is retained in both sexes. Mutations of Sxl binding sites in either the 5' or the 3' regions result in ectopic expression of MSL-2 protein in females. This indicates that the sites in both the 3' and 5' UTRs are required for appropriate regulation of *msl-2* translation (Bashaw and Baker, 1997; Kelley *et al.*, 1997). Bashaw and Baker, (1997) suggest the possibility that *Sxl* binding at both ends of the *msl-2* transcript changes the structure of the RNA by circularisation and therefore prevents access of translational machinery.

### 1.5.3 A Second MSL Independent Method of Dosage Compensation

A second method of dosage compensation has been suggested in *Drosophila* that is independent of the *msls*. The first evidence for this second method was the observation by Cline, (1978) that females homozygous for a loss-of-function mutation for *Sxl* were not rescued if also homozygous for mutations in *msl-2, msl-1* or *mle*. *msl-3* is yet untested, but assumed to have the same phenotype due to the co-dependence of the *msls*. These findings suggest that *Sxl* and the *msls* may act on different loci to direct dosage compensation.

Dosage compensation of the X-linked *Runt* gene has been shown to be dependent on *Sxl*, but independent of the *msls* (Gergen, 1987; Bernstein and Cline, 1994). Wildtype *Runt* is required for the normal segmentation of *Drosophila* embryos (Gergen and Wieshaus, 1986) and is active at the blastoderm stage. Gergen, (1987) studied the

dosage compensation of *Runt* at the blastoderm stage of development and found that *Runt* expression was not affected by mutations in *msl-1, msl-2* and *mle*. Examination of the *Runt* gene revealed three *Sxl* binding sites in the 3' UTR (Kelley *et al.*, 1995). *Sxl* expression from the early *Sxl* promoter ($Sxl_E$) occurs at the same stage of development as *Runt* expression, which supports the idea that early dosage compensation begins at mid-stage four and that *Runt* expression is *Sxl* dependent and *msl* independent (Bernstein and Cline, 1994; Gergen, 1987). The regulation of *Runt* by *Sxl* is probably due to repression of expression in females as indicated by only female-specific lethal alleles of *Sxl* affecting dosage compensation at the blastoderm stage (Gergen, 1987).

Two models have been suggested for the relationship between MSL dependent and *Sxl* mediated dosage compensation. The first involves *Sxl* controlling 'early' dosage compensation during embryogenesis, while the MSLs mediate 'late' dosage compensation during the larval and pupal stages. Evidence for this model is that *msl* mutant males complete embryogenesis, but die as late larvae or pupae (Belote and Lucchesi, 1980a; Fukunaga *et al.*, 1975) and *Runt* is expressed before MSLs become functional (Gergen, 1987). *Sxl* and *Runt* expression is detected at mid-stage four of embryogenesis, but MSL binding to the X chromosome does not occur until the end of stage five (blastoderm stage). The *Sxl* 'early' dosage compensation process may have evolved to satisfy a need for dosage compensation before the MSLs become functional and therefore the two systems are operating sequentially (Franke *et al.*, 1996).

The second model is that *Sxl* and MSL mediated dosage compensation pathways act in parallel (Rastelli *et al.*, 1995) during development on separate sets of genes (Kelley *et al.*, 1995). Recent data suggests that *Sxl* may reduce the stability or translation of a subset of X-linked transcripts in females (Kelley *et al.*, 1995). Kelley *et al.*, (1995) suggest that this second dosage compensation system may upregulate X-linked genes in males, while a subset of X-linked genes are down regulated in females. A computer search scanning all available 3' UTRs of *Drosophila* genes produced 21 genes containing three or more 3' poly(U) sites. 20 of these genes are on the X chromosome. The only autosomal gene found was *msl-1* (*msl-2* is also autosomal) (Kelley *et al.*, 1995). Kelley *et al.*, (1995) proposed that *Sxl* directly regulates dosage compensation of many genes through their 3' UTRs. Bernstein and Cline, (1994) suggest that *Sxl* mediated dosage compensation is not limited to embryonic development through studies

of partial loss-of-function *Sxl* mutants, while Rastelli *et al.*, (1995) suggest that *msl*-dependent dosage compensation is not limited to larval development.

## 1.6 *CIS*-ACTING ELEMENTS CONTROLLING DOSAGE COMPENSATION

Relatively little is known about the *cis*-acting sequence characteristics of the X chromosome which identify it as a target for dosage compensation regulators (ie. MSLs and Sxl). These dosage compensation regulatory elements (DCREs) are thought to be distributed throughout the X chromosome. Evidence suggests that DCREs exert their efforts locally on individual genes or small groups of genes. When fragments from the X chromosome are transposed to an autosome the X-linked genes within the fragment remain dosage compensated (Ghosh *et al.*, 1989; Hazelrigg *et al.*, 1984; Krumm *et al.*, 1985; Levis *et al.*, 1985; McNabb and Beckendorf, 1986; Pirrotta *et al.*, 1985; Spradling and Rubin, 1983). Also when cloned X-linked genes are translocated to autosomal sites they remain at least partially dosage compensated (reviewed by Baker *et al.*, 1994; Lucchesi and Manning, 1987).

Conversely when autosomal fragments are translocated to X chromosome sites the genes within the fragment remain non-compensated. But, when cloned autosomal genes are translocated to the X chromosome they are compensated in males (Baker *et al.*, 1994). These observations suggest that dosage compensation in *Drosophila* is controlled by *cis*-acting sequences both distant and close to the genes. Supporting this suggestion is the observation that not all genes on the X chromosome are dosage compensated. These non-compensated genes can be in close proximity to genes that are.

Support for the hypothesis that *cis*-acting elements confer transcriptional upregulation only onto nearby sequences is the finding that *LSP-1α* is an X-linked gene but is not compensated. *LSP-1α* codes the alpha subunit of larval serum protein-1. *LSP-1α* has a transcription unit named *L12* immediately adjacent to it that is compensated (Ghosh *et al.*, 1989). Females exhibit twice the amount of gene product found in males (Brock and Roberts, 1982; Roberts and Evans-Roberts, 1972). This phenomenon could be explained by assuming that the *LSP-1α* gene has only relatively recently been

translocated to the X chromosome. Ghosh *et al.*, (1989) determined that *LSP-1α* is inherently capable of dosage compensation by relocating the *LSP-1α* gene to ectopic X chromosome sites. The results of this experiment showed steady state levels in males (one dose) are equivalent to females (two doses).

### 1.6.1 DCREs Are Still Unidentified

For the last ten years the search for *cis*-acting sequences involved in dosage compensation has been fruitless. Two X-linked genes (*white* and *Sgs-4*) have been extensively studied using genetic and molecular techniques in an attempt to localise the DCREs. The studies predominantly involved inserting X-linked transgenes, which contained progressive deletions, into autosomes to isolate a possible consensus sequence for dosage compensation.

Levis *et al.*, (1985) analysed the *cis*-acting sequences involved in regulating the *white* gene. Varying lengths of both 3' and 5' flanking sequences were deleted from the *white* gene. Flanking sequences 420 bp upstream and 160 bp downstream of the gene were found to be sufficient for dosage compensation to occur. Pirrotta *et al.*, (1985) further delimited the required sequence to 200 bp upstream of the gene.

As sequences in the *white* gene are gradually removed from the 5' end a progressive decline in dosage compensation is observed (Qian and Pirrotta, 1995). Qian and Pirrotta, (1995) concluded that *cis*-acting DCREs consist of multiple elements present near and within the promoter and some within the coding region of the gene. Despite these observations no DCRE consensus sequence has been identified.

Transformation experiments involving the *Sgs-4* gene demonstrated that 840 bp upstream and 130 bp downstream of the gene are sufficient for proper activity and regulation when relocated to autosomal sites (McNabb and Beckendorf, 1986). Sequence comparisons between compensated and non-compensated alleles failed to show any base substitutions specific to the non-compensated alleles (Hofmann and Korge, 1987).

### 1.6.2 Mono and Dinucleotide Repeats Correlate with Dosage Compensation

Evidence has been presented that suggests the X chromosome has unique structural features that may be related to dosage compensation. Two dinucleotide repeats $(CA/GT)_n$ and $(CT/GA)_n$ and one mononucleotide repeat $(C/G)_n$ have been reported to be found at twice the level on the X chromosome as on autosomes (Huijser *et al.*, 1987; Lowenhaupt *et al.*, 1989; Pardue *et al.*, 1987). Chromosomal arms from autosomes translocated to the X chromosome acquire the ability to dosage compensate in several *Drosophila* species. The newly translocated arm also gains a higher density of $(CA/GT)_n$ similar to the other X chromosomes. The pattern of $(CA/GT)_n$ sequences shows several correlations with general chromosomal functions such as dosage compensation (Pardue *et al.*, 1987). Pardue *et al.*, (1987) suggests that the acquisition of dosage compensation ability and higher density of $(CA/GT)_n$ repeats reflects a relationship between the two processes.

These repeats are all able to adopt a non B form of DNA when subjected to negative supercoiling *in vitro* and may be involved in the adoption or maintenance of a decondensed X chromatin structure required for dosage compensation (Lowenhaupt *et al.*, 1989). Other than their enrichment on the X chromosome there is no evidence that these repeats are involved in dosage compensation as the repeats are also found on autosomes at significant levels.

### 1.6.3 Why Have DCREs Not Been Identified?

A major limitation of previous attempts to identify DCREs is that X-linked genes on autosomes are only partially compensated.

All studies (excluding *white* studies) used Northern Blots or RNase Protection Assays to quantitate gene expression levels in males and females. These methods experience technical difficulties when quantitating two fold differences in expression. Studies of the *white* gene used spectrophotometric eye pigment assays that must take into account the non-linearity of the pigmentation response to gene dose (Qian and Pirrotta, 1995). Early indications of this non-linearity were demonstrated when transgenes at autosomal sites exhibited full dosage compensation - males with one dose produce twice as much pigment as one dose females. However females with two copies of the white transgene

have a two to three fold higher level of pigmentation than males with one (Hazelrigg *et al.*, 1984; Levis *et al.*, 1985; Pirrotta *et al.*, 1985).

Hypotheses have been suggested to account for partial dosage compensation when X-linked transgenes are translocated to autosomes. Qian and Pirrotta, (1995) suggest that the requirement for a certain amount of DCREs associated with the gene is not being met and/or that the autosomal chromatin environment (more condensed than X chromosomes) is having an inhibitory effect upon the transgene. The use of insulator elements to flank the transgene supports this hypothesis (Roseman *et al.*, 1995).

## 1.7 A NEW APPROACH TO IDENTIFYING DCRES

Fitzsimons *et al.*, (1999) developed a reporter gene assay that can be used to screen X chromosomal DNA for DCREs. The components of this assay are the *E. coli lacZ* gene under the control of the constitutive promoter from the *armadillo* gene (this fusion is referred to as *arm-lacZ*). The *arm* promoter was chosen because it is constitutive and active in all tissues and all stages of development (Vincent *et al.*, 1994) in both males and females. *arm-lacZ* was flanked by SCS and SCS' insulator elements (specialised chromatin structures). SCS and SCS' sequences act as domain boundaries (Udvardy and Schedl, 1993). Domain boundaries establish a domain of independent gene activity by protecting against regulatory effects of surrounding chromosomal DNA. It has been found that *arm-lacZ* can respond to DCREs when on the X chromosome - one copy in males is expressed at twice the level of one copy in females.

A limitation in studying X-linked genes is that the coding region must remain intact enough so its product can be assayed for. Using this newly developed assay system allows X-linked sequences to be subdivided as a reporter gene is detected in the assay rather than the gene product. Fitzsimons *et al.*, (1999) placed portions of DNA from the *D. melanogaster* X chromosome immediately upstream of the *arm* promoter. The hypothesis for these experiments was that any X-linked sequence containing DCREs would confer dosage compensation onto *arm-lacZ* in males and thereby produce twice the *lacZ* activity in males over females. As yet the DCREs remain unidentified.

## 1.8 RESEARCH OBJECTIVES

This study has three main objectives. The first objective is to investigate the effect of *roX* genes on dosage compensation regulated by isolated DCREs. The second is to continue with the study begun by Fitzsimons *et al.*, (1999) and isolate the DCREs involved in dosage compensation. However our study will focus particularly on regions of the X chromosome (eg. 1C) known to contain 'high affinity' binding sites for the MSL-1/MSL-2 core complex (Lyman *et al.*, 1997). The final objective is to investigate further the possibility of *Sxl* regulating a second dosage compensation pathway throughout development.

### 1.8.1 Specific Objectives

Previous studies have been carried out to develop a new reporter gene assay system that can be used to screen X chromosomal DNA for DCREs. Fitzsimons *et al.*, (1999) developed and used this assay system on many constructs containing X-linked fragments of DNA in the attempt to isolate the elusive DCREs. This study will utilise the *arm-lacZ* assay system developed by Fitzsimons *et al.*, (1999) to investigate the role of various X-linked DNA fragments in dosage compensation.

The initial aim was to test if the *roX* genes are needed to be present in *cis* in order for a fragment containing a DCRE to cause a male specific increase in *lacZ* expression. It was found that the *roX* genes alone caused elevated *lacZ* expression ie. the *roX* genes contained DCREs. Consequently the initial objective was modified to test if *roX* cDNAs and fragments of *roX* genes contained DCREs.

The second aim of this study was to determine if DNA fragments from the tip of the X chromosome, in particular the region that shows "high affinity" binding with the MSL-1/MSL-2 core complex. The assay system will also be used to look for DCREs in these X-linked DNA fragments. The presence of DCREs would be confirmed by an increase in *lacZ* expression in males

The third aim was to investigate the role of *Sxl* in dosage compensation in females throughout development. The study aimed to determine if insertion of 3' UTR fragments from other X-linked genes (*Sxl, Small Forked, and Cut*) would cause a

decrease in female specific expression of *lacZ*. These 3' UTR fragments contained 3 or more Sxl binding sites and were inserted into the 3' UTR of *arm-lacZ*.

The fourth and final aim also looked at Sxl involvement in dosage compensation. Experiments with *msl-2* showed the Sxl binding sites were required in both the 3' and 5' UTR to get complete repression of translation. Previously Fitzsimons *et al.*, (1999) showed insertion of a *Runt* 3' UTR fragment into the *arm-lacZ* 3' UTR caused a modest decrease in female *lacZ* expression. This study aimed to determine if an additional Sxl site in the 5' UTR of the *arm-lacZ* construct, carrying the *Runt* 3' UTR fragment, would result in a more dramatic decrease in female *lacZ* expression.