Dissertation

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Erinc Hallacli, M.Sc.

Born in Ankara, Turkey

Date of the oral examination:

Architecture of Drosophila Dosage Compensation Complex

Referees:

Prof. Dr. Gabriele Petersen

Dr. Christian Haering

TABLE OF CONTENTS

ZUSAMMENFASSUNG

Männliche und weibliche Zellen heterogameter Organismen enthalten einen unterschiedlichen Satz von Geschlechtschromosomen. Die daraus resultierenden Unterschiede in der Genexpression werden oft durch epigenetische Mechanismen ausgeglichen, ein Phänomen bekannt als Dosiskompensation. Mittels einer molekularen Maschinerie genannt MSL Komplex erhöhen männliche Zellen in Drosophila die Genexpression ihres einzelnen X chromosoms ungefähr um den Faktor zwei. Dieser enzymatische Ribonucleoproteinkomplex, spezifisch gebildet in männlichen Zellen, bindet an das X Chromosom, ändert dessen Struktur hauptsächlich durch die Acetylierung von Histon H4 an Lysin 16, und ermöglicht dadurch eine verstärkte Transkription. Wie genau der MSL Komplex das X chromosom erkennt und die Transkription reguliert ist nicht im Detail verstanden. Mit Hilfe von röntgenkristallographischen Daten haben wir in dieser Arbeit Puntkmutationen in Msl1 generiert die zum Verlust einzelner Faktoren aus dem MSL Komplex führen, um so deren Beiträge zur X Chromosom Erkennung, RNA Integration und zum „spreading" des MSL Komplexes entlang des X Chromosoms zu studieren.

Der Fokus des ersten Teils dieser Arbeit liegt auf der PEHE region des Msl1 Proteins welche mit Mof und Msl3 interagiert. Wir haben Msl1 Punkmtutante erzeugt die nicht mehr an Mof oder Msl3 binden können. Wir zeigen dass sowohl der Verlust von Mof als auch von Msl3 aus dem MSL Komplex dessen „spreading" in den transkribierten Teil X chromosomaler Gene verhindert, wobei die Bindung von Msl1 an die Promoteren derselben Gene erhalten bleibt. Desweiteren beobachten wir qualitative Unterschiede zwischen verschiedenen „high affinity sites" (HAS), den initialen Bindestellen des MSL Komplexes, insoweit dass HAS an Promotoren Msl1 unabhängig von Mof und MSl3 rekrutieren können während die optimale Bindung von Msl1 an andere HAS von einer intakten PEHE Domäne abhängig ist.

Im zweiten teil dieser Arbeit haben wir die Interaktion zwischen MSl1 und Msl2 untersucht und zeigen dass Msl1 mit seiner coiled-cloil Domäne ein Homodimer bildet welches als Plattform für die Bindung an Msl2 dient. Diese Interaktion wird in Msl2

durch Helices um die RING Finger Domäne vermittelt. Wir zeigen dass der Msl2 RING Finger eine Ubiquitin-Ligase Aktivität aufweist und dass Msl1 ein *in vitro* Substrat für Msl2 vermittelte Ubiquitilierung ist. Durch Einbringung von Punktmutationen in Msl1 zeigen wir dass dessen Dimerisierung in männlichen und weiblichen Zellen unabhängig von Msl2 stattfindet. Msl1 Dimerisierung ist erforderlich für die Bindung an Msl2, roX2 Integration in den MSL Komplex, X Chromosom Erkennung, und „spreading" in den transkribierten Teil X chromosomaler Gene. Diese Ergebnisse zeigen deutlich dass die Funktion des MSL Komplexes von seiner Konfiguration als Dimer abhängig ist. Wir identifizieren die roX2 HAS als elementare HAS welche vom Msl1-Msl2 Dimer alleine erkannt wird. Desweiteren fanden wir das Msl1 Promotoren unabhängig von Dimerisierung/Msl3/Mof/Msl2 erkennt. Diese Interaktion findet auch an Autosomen statt und legt eine generelle Funktion von Msl1 an Genpromotoren in Drosophila nahe. Wir zeigen dass diese Promotoren auch von Msl1 aber nicht von Msl3 besetzt sind, was nahe legt dass Msl3 eine wichtige Rolle zur Unterscheidung des Promoter-Komplexes gegenüber dem klassischen MSL Komplex spielt. Um die Signifikanz der oben beschriebenen Aminosäuren in Msl1 weiter zu untermauern haben wir transgene Fliegen erzeugt welche die mutierten Formen von Msl1 vom selben genomischen Locus exprimieren. Wir zeigen dass diese Msl1 Mutanten nicht in der Lage sind die Männchen spezifische Lethalität in Abwesenheit von endogenem Msl1 zu beheben, und dass die Überexpression derselben Muanten in Anwesenheit von endogenem Msl1 ebenfalls zu männlicher Lethalität führt. Diese Ergebnisse bestätigen die zentrale Rolle der mutierten Aminosäurereste für die Funtkion von Msl1.

SUMMARY

In heterogametic organisms, male and female cells harbor structurally different sex chromosome pairs. The difference in the transcriptional output of these sex chromosomes is epigenetically balanced, a phenomenon dubbed as dosage compensation. In Drosophila, male cells up-regulate their one X chromosome roughly two times by the help of a molecular machine called MSL complex. This ribonucleoprotein enzymatic complex, specifically formed in male cells, localizes to the X-chromosome, changing its structure mostly by histone H4 Lysine 16 acetylation, thus enabling enhanced transcription. The details of how the complex finds the chromosome and how it regulates transcription are not thoroughly understood. In this thesis, with the help of X-ray crystallography, we derived point mutations on the scaffold Msl1 protein that create partial complexes to study the contribution of each subunit to X chromosome recognition, RNA integration and spreading alone the X chromosome.

In the first part of this thesis, we focused on the PEHE region of Msl1 protein that binds Mof and Msl3. We generated point mutants of Msl1 that cannot bind Mof or Msl3. We showed that loss of either Mof or Msl3 prevents spreading of the MSL complex on the body of the X-linked genes whereas Msl1 promoter binding remained unaffected. We observed qualitative differences between high affinity sites (HAS), initial binding platforms of MSL complex, and noticed that promoter located HAS can bind Msl1 independent of Mof and Msl3 whereas other HAS depend on the presence of intact PEHE module for optimal binding.

In the second part of the thesis, we examined the interaction of Msl1 and Msl2 and showed that Msl1 forms a homodimer through its coiled coil region and this homodimer creates a platform for Msl2 binding. Msl2 binding happens through helices surrounding the RING finger domain. We showed that Msl2 RING finger can function as a ubiquitin ligase, and Msl1 is an *in vitro* substrate of Msl2 ubiquitination. By point mutational analysis on Msl1 we showed that Msl1 forms a dimer independent of Msl2 in both male and female cells. Dimerization is required for Msl2 binding, roX2 RNA integration to the complex, X chromosome recognition and spreading along the body of X-linked genes. This clearly showed that functionality of MSL complex entirely depends on its dimeric

configuration. We identified roX2 HAS as an elementary HAS where its recognition only happens through Msl1-Msl2 dimer interface. Furthermore we discovered that Msl1 binds to promoters in a dimer/Msl3/Mof/Msl2 independent fashion. This binding occurs also at the autosomes and in both sexes suggesting a general function of Msl1 at promoters of Drosophila. We showed that promoters are also occupied by Msl2 but not by Msl3, indicating that Msl3 can have an important role for distinguishing promoter bound complex and canonical MSL complex. In order to support the *in vivo* importance of amino acid residues that had been point mutated, we generated transgenic flies that express Msl1 and its mutated forms from the identical genomic location. We showed Msl1 mutants are unable to rescue the *Msl1* null male lethality and also cause male specific lethality upon over-expression in wild type background confirming the importance of these mutated residues.

ACKNOWLEDGEMENTS

I would like to thank Asifa for giving me the opportunity work in her lab, setting up invaluable collaborations and giving me the freedom to do almost whatever I would like to do. I am grateful for the scientific support of my TAC members: Christian Haering, Lars Steinmetz and Gabriele Petersen.

I would like to thank Jan Kadlec and Steven Cusack for their immense and important contributions in my PhD. It was a pleasure to work with you and I had grown a great deal of interest into protein structures.

Asifa's lab was almost like a home to me. It also has been an amalgam of lots of interesting people and their cultures and I very much enjoyed their company. However I must especially thank Ibrahim and Thomas. You have created an ultra funny, intellectual, educative, warm microenvironment and I take myself lucky to be a part of this small cosmos. I am indebted to Herbert for being the most helpful and polite person on the planet.

For четиринадесет billion years, the universe had gone through a lot of trouble to connect us, my love. Maybe this bond was even somehow formed immeasurably small time after the big взрив. So much energy was spent to calm down the eternal хаос so that I could realize this, or at least try to understand and appreciate it. Yet most of the time, I am trapped in deep holes confusion. Thank you for taking me out of those abysses. I hope that when everything goes apart in an increasing speed, we stay together, by the help of a yet undefined force.

For Boryana and my family

INTRODUCTION

INTRODUCTION

Heterogametic organisms with unequal numbers of sex chromosomes have to go through a process called dosage compensation to equilibrate their transcriptional output. Diverse solutions to the dosage problem evolved in different organisms during evolution. *Drosophila melanogaster* males transcriptionally up-regulate their single X chromosome roughly two times to compensate for the absence of an active homolog (Figure 1) [1], whereas in mammals, females inactivate one of the two X chromosomes [2]. Dosage compensation not only balances sex differences but has also been shown to equalize X to autosome ratios in mammals, *C. elegans* and *Drosophila* [3,4]. Dosage compensation mechanisms provide an excellent model for studying chromosome-wide transcription regulation through epigenetic mechanisms [5]. In *Drosophila melanogaster*, dosage compensation is regulated by male specific lethal (MSL) factors whose products are essential for male survival. These factors are collectively called the MSL complex or the dosage compensation complex (DCC).

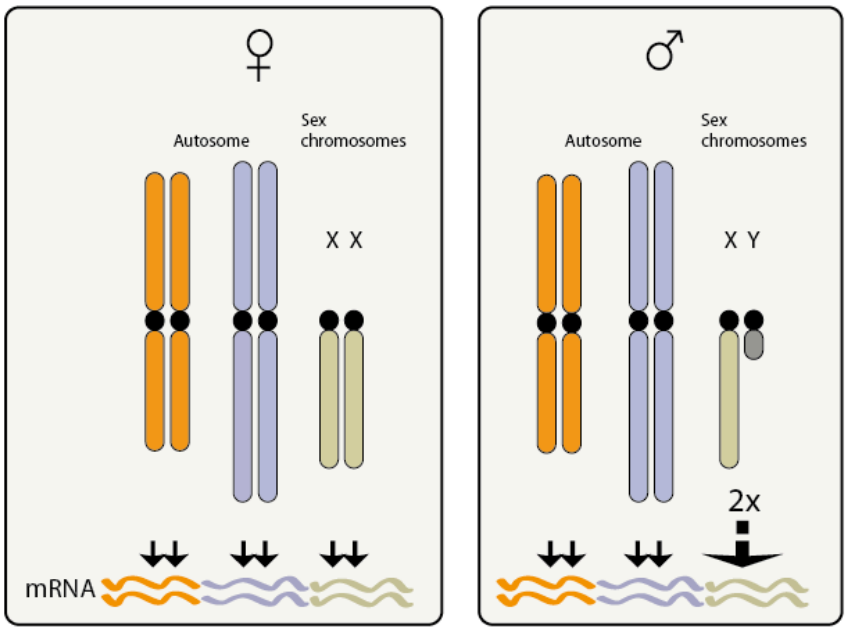


Figure 1. Dosage compensation in *Drosophila melanogaster*.

Female cells have two X chromosomes, whereas male cells has one X and one degenerated Y chromosome. The male X produces the same dose of RNA as the female to compensate for the absence of homologue.

COMPLEX COMPONENTS and ARCHITECTURE

Msl1 and Msl2

The MSL complex is a ribonucleoprotein complex that is composed of two long non-coding RNAs; roX1 or roX2 and at least five proteins namely Msl1, Msl2, Msl3, Mof and Mle (Figure 2). All protein-coding genes are transcribed in both sexes [6-11] however Msl2 mRNA translation is strictly inhibited by Sxl, master sex regulator, in females[12]. In the absence of Msl2, Msl1 protein is destabilized and presumably degraded because Msl1 is normally not detected at the protein level unless *msl2* is expressed ectopically in females[6]. Msl1 can be regarded as an assembly platform of the complex because it interacts with all other protein members, except for Mle[13]. Leucine zipper like motif at the amino (N) terminus interacts with Msl2[14] and carboxyl (C) terminus binds Mof and Msl3 [13]. Msl3 and Mof contact occurs on different parts of Msl1, Msl3 being close to the C terminus and Mof with PEHE domain[15]. Msl2 has a RING finger domain along with a cysteine rich motif at its C terminus[7]. The RING domain has two zinc finger clusters and mutations of polar residues chelating the first zinc ion have been shown to disrupt the interaction of Msl2 with Msl1[16]. Although the RING finger is conserved in many species, the novel combination of RING domain and cysteine cluster has been proposed to have an important contribution for the birth of '*msl2* like' genes and a driving force for the formation of compensasome [17]. Human MSL2 is shown to be a E3 ligase, making the third enzyme in the complex along with MOF and MLE, and responsible for a novel ubiquitination mark on histone H3K34 [18]. The importance of this mark for drosophila dosage compensation remains to be elucidated. Msl3 has a C-terminus MRG domain that mediates the interaction with Msl1[19]. MRG domains are highly conserved in MRG gene family and thought to be interaction platforms in large complexes that are usually chromatin related [20]. Interestingly the MRG domain of Msl3 is interrupted by non-conserved sequences and the importance of these Msl3 specific linkers are yet to be determined[19].

Mof and Msl3

The early observations of polytene squashes from the male larvae salivary glands revealed an interesting discovery that male X chromosome was enriched for a specific acetylation mark on the histone 4 lysine 16 (H4K16)[21]. Co-localization of this mark with known MSL members and its absence in MSL mutants predicted a histone acetyltransferase (HAT) enzyme in the complex[22]. Concordantly Mof, a member of MYST family of HATs (named for members MOZ, Ybf2/Sas3, Sas2, and Tip60), was shown to co-localize with MSL members and mark the male X chromosome together with H4K16 [11,23]. Mof has a peculiar C2HC zinc finger motif, which mediates substrate recognition [24], and its interaction with Msl1[15]. Although Mof binds to nucleosomes and its preferred substrate is H4[25,26], it can also acetylate Msl3 and Msl1[27], nevertheless integration into the complex shifts the substrate specificity strongly to H4[15]. Both Mof and Msl3 have a chromo-related domain but accumulating evidence indicate that they may have different substrate specificities. Solution structure of Mof chromo-related domain revealed that this fold has five beta strands forming a barrel shape, therefore named as chromo-barrel domain (CBD), which is structurally similar to Tudor domains[28]. This fold is quite different from chromodomains that bind lysine methylated histone tails through a hydrophobic cage formed by three conserved aromatic residues[29,30]. These residues are not found in Mof and in fact R387 of CBD may clash with superimposed methylated lysine H3 peptide[28]. Instead Mof chromo-barrel domain in the context of the full-length protein is important for RNA binding activity *in vivo* and *in vitro* and mutation of a conserved tyrosine disrupts this interaction [31]. Nucleic acid binding activity of a chromodomain is not unique to Mof but also well documented for dMi-2 protein of NuRD remodeling-HDAC complex in *Drosophila* [32]. Curiously, the CBD used for structure determination was not able to bind RNA by itself leading to the hypothesis that the fold may need other residues for RNA binding activity[28]. Supportive evidence for this prediction came from the structural elucidation of yeast homologue of Mof, Esa1 chromo-domain[33]. Esa1 chromo-domain folds similar to MOF CBD *in silico* and does not show nucleotide binding activity *in vitro* however structure of N terminally extended domain has an extra beta sheet which induces a loop in the barrel. This "knotted" barrel has a high affinity for RNA *in vitro* [33]. In the same vein, Mof chromo-barrel domain could adopt a slightly different form in the context of full protein such that it has a high affinity for RNA.

Msl3 chromo-related domain is predicted to fold similar to chromo-barrel domain but unlike Mof, it contains the typical hydrophobic residues that forms the aromatic cage [28]. Msl3 can bind nucleosomes that are methylated on H3K36 and mutation of hydrophobic residues causes the loss of this interaction[34,35]. Besides, the chromodomain by itself can bind to a nucleosomal template[36]. Chromodomain structures of Eaf3 and MRG15, yeast and human homologue of drosophila Msl3 respectively, were solved and showed to bind methylated H3K36[37,38]. Nevertheless, Eaf3 interaction with this histone mark is a rather weak one and was suggested to take assistance from PhD finger for the optimal binding [39,40]. MRG15 binding to H3K36me3 is predicted to be different from canonical chromodomains because a beta strand in chromo-barrel preoccupies the histone peptide-binding groove[37]. Interestingly, Msl3 chromo-related domain can bind DNA *in vitro* suggesting that nucleosome interaction may occur partly through DNA[41]. Additionally, Msl3 localization to X-chromosome is lost upon RNAse treatment and Msl3 can bind roX RNA both *in vivo* and *in vitro*[19,27], therefore nucleic acid binding surfaces can be used to bind both RNA and DNA. Acetylated Msl3 cannot immunoprecipitate significant amounts of roX2 RNA; hence this posttranslational modification may be used as a regulatory switch for the substrate specificity of Msl3. Recently two groups independently solved the crystal structure drosophila Msl3 chromodomain and showed that it binds rather to H4K20 methyl marks (preferentially lower methylation status: mono and di)[42,43]. Interestingly H4K16 acetylation antagonizes with H4K20 methyl mark indicating a dynamic display of H4 peptide in the complex[42,43].

Mle

The third enzyme associated with the MSL complex is Mle. Mle bears a modified DEAD box motif, DEIH, which is one of the key signatures for RNA helicases[10]. Mutations in the DEAD box and the ATP binding pocket are lethal for flies [44,45]. Mle fulfills single stranded nucleic acid binding, double stranded nucleic acid binding, ATPase activity and homo-hetero duplex unwinding activities *in vitro* [45].

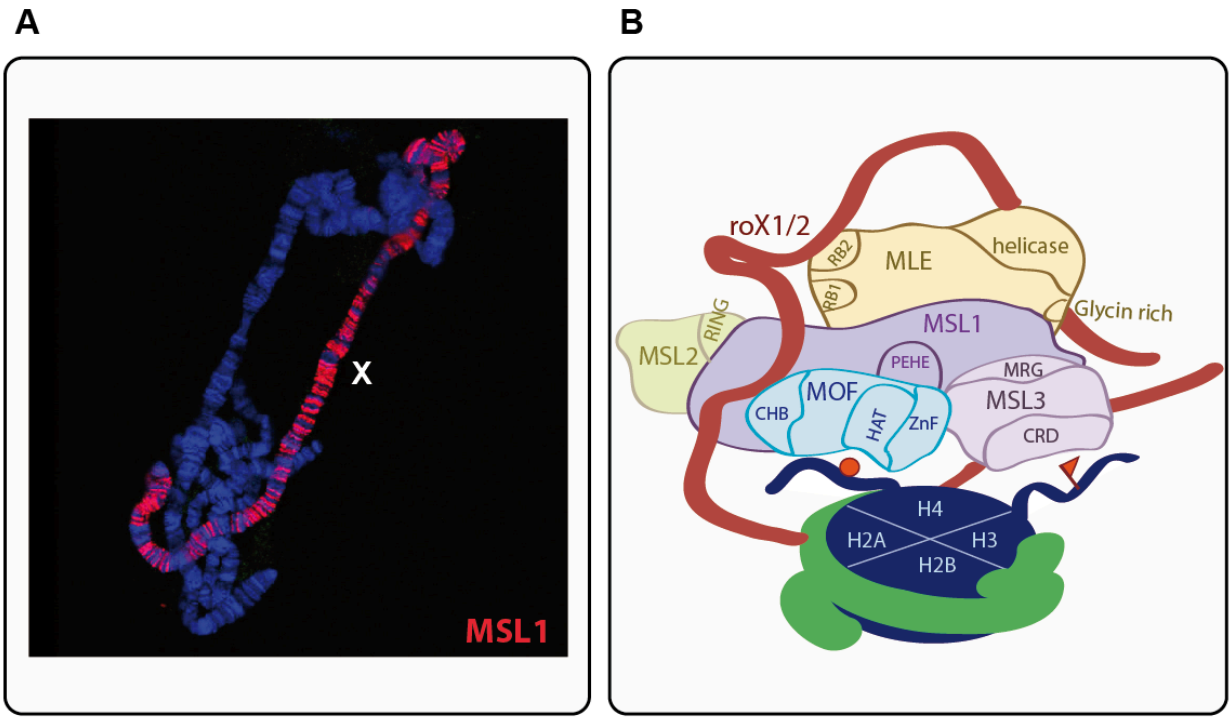


Figure 2. MSL complex and known interactions.

(A) MSL complex binds specifically to X chromosome in male cells. This is clearly seen in polytene squashes of salivary glands from third instar male larvae. DNA is stained with Hoechst and Msl1 is detected by specific antibody. **(B)** The MSL complex. Msl1 and Msl2 interact through RING domain of Msl2 and N terminus of Msl1. Mof chromobarrel (CHB) domain interacts with RNA. HAT (histone acetylase) domain acetylates H4K16 residue (represented by a red ball) and Zinc finger (ZnF) is important for the H4 specificity. Msl3 chromo-related domain (CRD) has been shown to bind to DNA and nucleosomes and been suggested to interact with tri-methylated H3 on K36 (H3K36me3 represented by red Flag). Msl3 and Mof bind Msl1 through ZnF and MRG domain, respectively. PEHE domain of Msl1 was shown to be crucial for Msl3 interaction. Mle has two RNA binding domains (RB1 and RB2) but only RB1 can bind RNA. Glycine rich region on the C terminus has a high affinity for RNAs. MLE could associate with the rest of the complex though RNA. The stoichiometry of the components and the mutual presence of roX RNAs are not known. The complex is not drawn to scale due to absence of any structural data therefore the figure must be seen as an artistic rendering of what is known so far.

It has two RNA binding domains in the N terminal region but only RB2 domain was shown to bind RNA[46]. RNA binding and deletion of C terminal glycine rich region increases the ATPase activity[46]. Therefore Mle may undergo continuous self-regulation through its own domains. Mle X chromosome localization is RNAse sensitive and it has a very salt susceptible, weak interaction with the rest of the complex, suggesting that it may bind MSL proteins through an RNA intermediate, presumably the roX RNAs [16,44]. However till today, direct Mle interaction with roX RNAs have not been shown.

roX RNAs

RoX RNAs were discovered in two different enhancer trap screens; one to look for sex specific expression in mushroom bodies of fly brains that causes dimorphic courtship behavior[47] and the other to look for differential expression of a reporter in mushroom bodies[48]. These RNAs are nuclear, male specific, expressed in all tissues of flies and co-localize with MSL proteins along the X chromosome in males[47-49]. Although their big difference in size and sequence, they are functionally redundant[49,50]. The only sequence similarity is a 30 bp sequence identity however deletion of this sequence has no phenotypic output[51]. Many evidence suggest that roX RNAs exert their functions through a yet unpredictable tertiary structure or at least not in a strict sequence dependent manner. Additional to the inter-redundancy of roXs, there is also high intra-redundancy in each roX [52,53]. Successive ten percent deletions of roX1 and series of small deletions in roX2 do not change the male viability except for a region near 3' end of roX1 that contains a predicted stem loop structure[52]. Moreover, roX RNAs from other *Drosophila* species can be integrated into *D.melonagester* MSL complex in spite of the low sequence homology; 31% in the example of *D.willistoni* roX2[53]. Recent findings suggest that there are evolutionary conserved "roX boxes" that may be the exchangeable functional units of roX RNAs [53]. Inarguably roX RNAs have physical contacts to the complex; they can be immunoprecipitated with MSL proteins and female expression of Msl2 causes stabilization of roX RNAs [26,31,47,54]. Although, Mle, Mof and Msl3 have the domains to bind roX RNAs, direct targets of roXs have not been determined but considering their size (roX1 3.7 kb and roX2 0.6 kb), it is plausible that they have several contact points. Interestingly, there is a time frame in the early hours of embryogenesis, where roX1 is transcribed in the absence of any detectable level of MSL components, but the rapid turnover of roX RNAs in the absence of the MSL complex

led to the question how this stability is achieved. It was found that in the absence of maternal Mle, roX1 RNA is hardly detectable, therefore maternal stores of Mle can contribute to early stabilization of roX1 [55].

EARLY EVENTS of MSL ASSEMBLY

What do females do?

The MSL complex formation is strictly inhibited in female cells, which is achieved by the master sex regulator, Sxl protein. Sxl is expressed only in females. In males, Sxl pre-mRNA is spliced in a way that a premature stop codon is retained and the mRNA is degraded[56]. Many observations clearly showed *msl2* mRNA as the direct target of Sxl. *Msl2* mRNA has poly U stretches, binding sequences of Sxl, in both of its UTRs. In females these binding sites are retained however in males a 133 bp intron in 5'UTR, containing two of are spliced out[57]. Additionally transgenic constructs lacking poly U stretches enabled expression of Msl2 protein in females[6]. Sxl modulates female specific intron retention by interfering with U2AF65 and U2AF35 snRNP interaction on the 3' splice site [58] and U1 recognition of the 5' splice site [59]. *Msl2* mRNA translation inhibition occurs a by a dual mechanism, one conducted through 3'UTR binding and inhibition of 43S recruitment and the second one through 5' UTR binding and prevention of 43S pre-initiation complex scanning [12,60-62]. 3'UTR control of Sxl requires a co repressor, Unr, which is present in both male and female cytoplasm but is specifically recruited to the *msl2* mRNA 3'UTR by Sxl in females[60,63,64]. Interestingly, Unr was found to have an important role in male dosage compensation[65]. Overexpression of Unr causes a preferential male lethality and loss of the MSL complex from X chromosome. Unr also immnoprecipitates roX1 and roX2 however if this is a direct or indirect interaction has not been shown [65].

Somatic versus germ line

Most of the available data on dosage compensation comes from observations of somatic cells, therefore much less is known how the male germ line deals with the dose problem. Mle had been known to function in spermatogenesis and consistently can be detected in male germ line cells[66]. Interestingly Msl1, Msl2 and Msl3 are not observed in these cells and Mle, along with H4K16 acetylation, is not concentrated on X but rather scattered throughout the genome[66,67]. Nevertheless, expression microarray analysis

showed that male germ line do compensate for the imbalance despite the absence of the MSL complex although the number of escaping genes are higher than those on the somatic cells[68]. The possibility of an MSL independent dosage compensation mechanism was pointed out before because *msl* mutant males can complete embryogenesis, survive up to third instar larvae - early pupae and additionally some genes are compensated in the absence of MSL, like *runt*[69]. This mechanism could be the result of a complex buffering system inherent in genetic networks or another uncharacterized protein complex may function in the early dosage compensation[70]. Although the nature and the timing of an MSL independent mechanism is elusive, it appears that MSL mediated activation begins at blastoderm, coinciding with zygotic transcription start in embryos[55,71,72]. In males with homozygous *msl1, mle* or *msl3* mutant mothers, the onset of MSL detection on X chromosome is delayed; supporting the idea that maternal contribution may help balancing the low level of zygotic expression in initial stages of the MSL complex establishment[67,71]. Initiation of dosage compensation relies on the expression of one of the roX RNAs[55]. In males, roX1 RNA transcription starts in the early stages of blastoderm (2 hours After Egg Laying, AEL) and Msl2 localization to nuclei foci follows after. When roX1 is absent, Msl2 localization to the nuclear foci can only be seen after roX2 expression, which is nearly 6 hours AEL. This may indicate that roX transcription may guide the MSL complex to the X chromosome.

TARGETING THE MALE X CHROMOSOME

Single gene versus high throughput analyses

The advances of new technologies such as expression-arrays, high resolution tiling arrays and new generation sequencing technologies coupled with biochemical methods gave a totally new pace in understanding the mechanism of dosage compensation in *Drosophila*. MSL-chromatin interactions had often been carried out in polytene squashes of salivary glands by immunofluoresence but the resolution of this technique is very low, therefore one big leap in the field occurred when MSL components were mapped in high resolution throughout the *Drosophila* genome by ChIP on chip method[73-78]. Although the immunoprecipitated proteins, cell type and embryonic stage are different in each case, common themes arose. First, not all genes on X are bound by MSL complex and also there are a few autosomal sites that are clearly bound. The MSL complex members are mostly found on genes rather than intergenic sequences and

when the binding profiles are averaged, a clear enrichment in the body and towards the end of the genes are observed [74,76] (Figure 3). Most target genes seem to be actively expressed however there is no correlation between the expression level and MSL abundance [75,76]. Although most MSL target genes are actively transcribed, transcription *per se* is not sufficient to explain MSL binding because many genes that are bound by elongating form of RNAPII and canonical elongation factors are devoid of the MSL complex [74,75]. Also, Msl1 binding profiles of 4-6 hour embryos and third instar larvae salivary gland are fairly similar[75], supporting the notion that most compensated genes are selected early during development and bind irrespective of developmental changes[79].

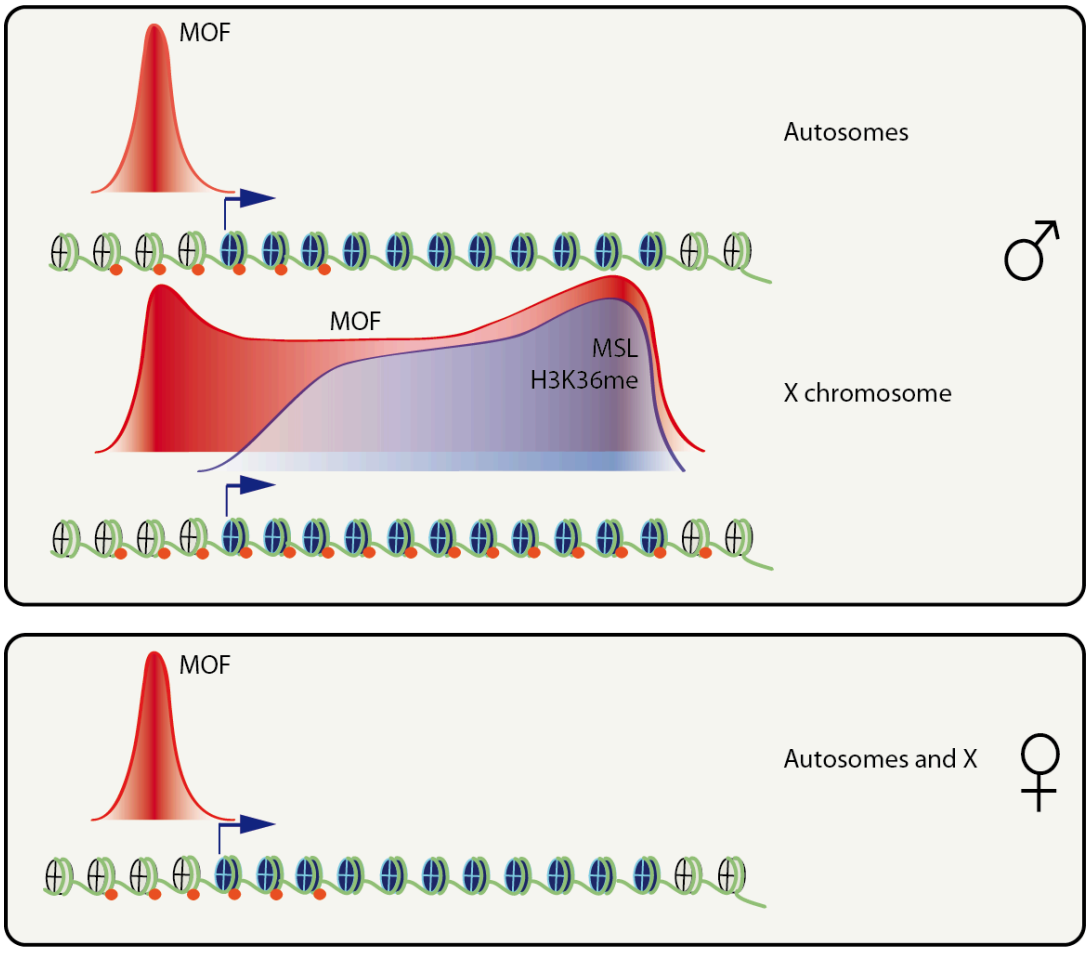


Figure 3. Global profiles of MSL components and associated histone marks averaged on a single transcriptional unit. In male cells, MSL components are enriched in the body of the genes peaking at the end along with the H3K36me3 mark.

Mof shows a bimodal distribution peaking at promoters and at the 3'end of the genes. Mof peak at the promoters are independent of MSL complex and also found in female cells and autosomal genes in males. H4K16 acetylation is on the whole body of the gene in Mof bound genes. Red balls indicate histone acetylation and transcriptional unit nucleosomes are depicted in blue.

Realization of 3' enrichment of the MSL complex members stimulated investigation of associations to other well-known 3' enriched epigenetic marks. Genome-wide histone modifications from yeast and humans showed that Set2 dependent H3K36me3 is a conserved 3' bias epigenetic mark associated with active genes[80,81]. Mapping of H3K36me3 on X chromosome revealed that more than 90% of MSL targets are also enriched with this mark and there is a high correlation of MSL and H3K36me3 position on the gene[35,82]. Interestingly, there seems to be a context dependent crosstalk between H3K36me3 and H4K16ac because reduction of Hypb, the enzyme required for final methylation state of H3K36 in *Drosophila*, causes a reduction of H4K16 acetylation mark on X-linked genes but not on autosomes[82]. Although H3K36me3 enrichment near the end of the genes and its interaction with Msl3 chromodomain is the current dogma, recently found H4K20me interaction with Msl3 chromodomain may yet reveal other epigenetic marks for the 3' enrichment of MSL complex.

Mof seems to have a unique status in the complex. Promoter binding of Mof is distributed throughout the whole genome in both males and females in an MSL independent manner whereas 3' enrichment is restricted to the X chromosome in males and is MSL dependent [73]. This led to the hypothesis that Mof plays an important role on the promoters of both sexes, and the MSL complex members binds this HAT to skew its location towards the end of the genes for specific acetylation and up-regulation of male X chromosome[73]. Morever, purifications of Mof led to the discovery of a novel transcriptional regulator complex, called NSL complex, which is partly responsible for the promoter targeting of Mof [83,84].

High affinity sites

One of the most obvious questions of dosage compensation is how the MSL complex recognizes the X chromosome specifically. An intriguing feature of the MSL complex is that Msl1 and Msl2 are able to bind 30-40 bands in polytene squashes in the absence of Msl3, Mof or Mle, which have been named as Chromatin Entry Sites (CES)[85,86]. Msl1 and Msl2 behave as the core of the complex because they are strictly dependent on each other for stability and they can localize to X without the other partners[6,7,86] . Surprisingly, *roX1* and *roX2* sites were the first CES to be mapped, due to their ability to recruit MSL components upon translocation to an autosomal site[54,87]. This feature of roX genes does not depend on their transcription but on a DNAse hypersensitive site (DHS) that can bind the MSL complex[51,88]. Initially, CES were thought to be the only sites for early MSL binding however, investigation of large X to autosome translocations showed that any segment of X was able to recruit MSL complexes even if they do not possess a previously mapped CES[89,90]. Moreover, translocated genomic segments from autosomes to X were devoid of MSL complexes[90]. Identification of other Msl1 binding fragments by ChIP showed that only a subset of these fragments are able to recruit Msl1/2 when moved to autosomes, and the rest can do so only in the presence of over expressed Msl1/2 [91]. Therefore, X chromosome seems to have a gradient of potential to recruit the MSL complex, named as the "affinity model"[89]. Some sequences can recruit MSL complex independent of any apparent targeting determinant, called High Affinity Sites (HAS) and other sequences, Low Affinity Sites (LAS), can only do so by the help of other mechanisms. The cipher of high affinity sites had remained a mystery due to absence of advance sequence algorithms and low number of mapped sites. But high throughput experiments discovered important clues about this phenomenon. High-resolution binding profiles of Msl1 and Msl2 in the absence of other MSL components revealed more than 130 Chromatin Entry sites (CES) or High Affinity sites (HAS) and a GA rich motif named as MRE motif (MSL Recognition) in these sites[77,78]. This motif is slightly enriched on X chromosome and autosomal transposition of a minimal CES, containing as few as three MRE elements was able to recruit MSL, and up-regulate the upstream reporter gene [77]. Since there are thousands of similar motifs scattered around the *Drosophila* genome, the choice of X chromosome still remains an unsolved issue. However, H3 depletion around the CES site indicates that accessibility could be an important player[77,78].

The importance of roX RNAs for the initial targeting have been shown in a number of cases[92]. MSL protein complex *per se* have a weak affinity towards the chromatin but this is greatly enhanced with the presence and/or integration of roX RNAs[50,93,94]. In the absence of both roX RNAs, partial MSL complex can be located on a few X chromosome loci, autosomal loci and chromocenter[50,93-95]. These sites have the intact complexes because all MSL proteins and H4K16 acetylation are seen coincidently[50,94]. Moreover, overexpression of Msl1 and Msl2 can rescue male lethality to some extent[94]. Nevertheless, there are opposing reports claiming that even though roX mutants are so severe that allows no detection by any means, they may still contribute to targeting and dosage compensation[96,97]. Therefore the role of roX RNAs in targeting is still an open issue.

Are the presences of entry sites on *roX* genes a coincidence? Compelling evidence suggest that the complex forms on the site of roX transcription in a co-transcriptional manner[94,98]. It is plausible that roX genes have acquired an entry site to establish the complex formation more efficiently and fast. On the other hand, *roX1* DHS was shown to play a role in *roX1* transcription activation in males and Msl2 was shown to be important for this role as an independent task from the MSL complex[99,100]. Similarly *roX2* gene was found to have elements that bind Mle and regulates its transcription[101]. Therefore, a complex regulatory network that contains components of the MSL complex may fine tune roX transcription and eventually formation of the MSL complex.

Beyond the high affinity sites

How is the complex located further from the high affinity sites? An intriguing observation upon roX gene translocation to autosomes was the spreading of the complex from the site of insertion in cis[54,87]. This spreading depends on the site of insertion, amount of the MSL complex, and presence of the competing roX transgene[95]. Because other large X to autosome translocations did not show any cis spreading, roX situation was pointed to be a unique feature of roX genes due to their function as the site of complex formation[89]. Mof enzymatic activity is required for localization to low affinity sites but it is not known if this is due to its canonical histone acetylation activity or another protein acetylation event that may help maturation of the MSL complex[27,102]. Mle helicase activity is also found to be important for LAS localization[103]. Since Mle is

required for the roX association into the complex[54], the phenotype can be a downstream result of an incomplete complex that is not able to spread further.

Enrichment of H3K36 methylation on actively transcribed genes prompted investigation of Msl3 as it contains a chromo-related domain that could be a good candidate that can bind to this mark. By ChIP-chip analysis, it was seen that Msl3 chromodomain mutants, including the deletion mutant were enriched around the high affinity sites suggesting the requirement of intact chromodomain for spreading[34]. In view of these results, Sural et al, proposes a two-step model, in which sequence dependent initial targeting to X chromosome is established on MRE containing chromatin entry sites and is followed by H3K36me3/chromodomain mediated spreading, analogous to heterochromatin spreading. Surprisingly, transgenic flies carrying the chromodomain mutants have a range of phenotypes ranging from fully viable, developmental delay and severe affect on male[34,36]. This already suggests that chromodomain mediated spreading from high affinity sites cannot be the sole mechanism for the MSL binding pattern throughout the X chromosome. Transcription was noticed as a good candidate inferred from the high correlation between active transcriptional state and MSL binding. Indeed when *mof* gene, a low affinity site, is translocated to an autosomal site, it can recruit Msl1 only in the presence of endogenous or exogenous promoter[104].When the promoter is absent, blocked or reversed, Msl1 cannot bind. Effect of transcription extends beyond the *mof* gene because blocking of RNA polymerase II by alpha amanitin decreases the occupancy MSL components on X-linked genes[104]. Binding of Msl1 after a strong activation of X linked genes by a Gal4 induced promoter had also been observed before[105]. The passage of RNA Polymerase II may either expose targeting sequences that are normally hidden or it can change the chromatin marks such that the gene becomes a better target for the MSL complex. All these models are not mutually exclusive and genes may have evolved different strategies to recruit MSL depending on their need to compensate, their inherent affinity towards MSL or their plasticity during the development. One attractive possibility could be that transcription in combination with MSL proteins that recognize chromatin marks on active genes could facilitate spreading along the X chromosome.

EVOLUTIONARY CONSIDERATIONS

Dosage compensation arises as an inevitable consequence of sex chromosome evolution. In *Drosophila* species, the dose problem begins with the evolution of Y chromosome[106]. Although the exact nature of Y chromosome formation is under debate[106], the current model predicts the random acquisition of a male determining gene on an autosome and prevention of recombination of that locus[107]. This leads to a strong tendency to accumulate degenerative mutations like transposition, duplication and finally the heterochromatinization of the whole chromosome [107]. Therefore the male cell nucleus is forced to balance this hemizygocity by formation of a novel complex acting on the X chromosome. There are plethora of evidence that once the MSL complex evolved, it was co-opted in other *Drosophila* species, which have a different sex chromosome history [108]. In *D. pseudoobscura*, a fusion event between an autosomal chromosome (Muller D element) and original X chromosome (Muller A element) led to formation of a metacentric X chromosome and the similarization of the both arms in terms of sequence identity[109]. Interestingly, the autosomal homolog of D element is lost in males and the newly translocated hemizygous arm is bound by MSL complex and acetylated on H4K16 [110]. On the contrary, in *D. americana americana*, a similar fusion event occurred but the males kept the autosomal homologue (Neo-Y) chromosome[111]. Since the two homologues can still recombine, there is no sign of degeneration on the neo-Y and no MSL binding on the X chromosome [110].

Perhaps the best tool to study the evolution of sex chromosomes and dosage compensation is the neo-sex chromosomes of *D.miranda*. *D.miranda* is closely related to *D. Pseudoobscura* and has the same metacentric X chromosome, which is fully dosage compensated. But in addition, a Robertsonian translocation of an autosome (Muller C element) to the Y chromosome generated a neo-Y chromosome [112]. The fusion event is thought to occur about 2 million years ago and the homologue pair is kept in the male cells (neo-X chromosome) [113]. After the fusion event, neo-Y chromosome had undergone extensive random degeneration, including retrotransposition, duplication and nonsense mutations but most loci are still intact [114]. Amazingly, the neo-X chromosome recruits MSL complex and acetylates H4K16 to the loci that are degenerating in the neo-Y homologue [108,110] (Figure 4).

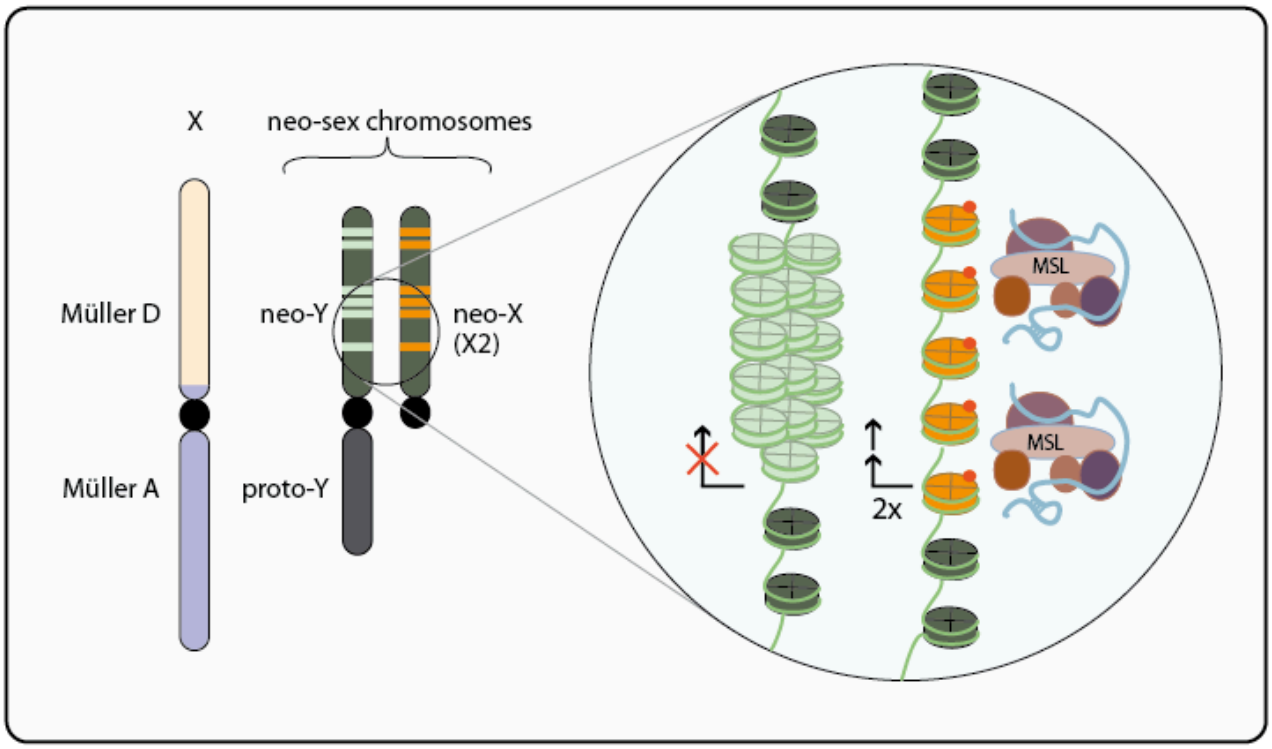


Figure 4. Dosage compensation in *Drosophila miranda*.

In male cells the proto X chromosome, (which is indicated by Muller A+Muller D) is compensated normally as in D.*melanogaster*. A translocation event of an autosomal arm onto the proto-Y chromosome created a neo-Y chromosome, which is in the process of degeneration. The degenerated loci are indicated as light colored bars. The autosomal homolog of neo-Y, also called neo-X or X2, shows upregulation at the loci in which there is degeneration on neo-Y. The upregulated loci are indicated in orange color and they correspond to similar loci as in the neo-Y. A hypothetical magnifier to one of these loci is shown. The degenerated loci on the neo-Y go under heterochromatinization due to retrotransposition and/or other means of molecular events leading to hemizygosity. In the homologous region on neo-X chromosome, the MSL complexes are recruited and upregulate the genes for dosage compensation.

Interestingly, the core promoters of the *lcp1-4* genes that are upregulated in the neo-X in response to degeneration in neo-Y show no apparent significant sequence alteration that could lead to two-fold upregulation [115]. Moreover, significant sequence variation of the neo-X chromosome from the old X shows that multiple selective sweeps of cis-acting regulatory regions did not occur [116]. These results show that MSL complex recruitment does not require a strict gene-by-gene basis cis-acting sequence evolution.

Since dosage compensation in *Drosophila* is an old problem and various subgroups use the same complex to cover up hemizygousity, MSL components are expected to be under stabilizing (purifying) selection. But recent experiments show that even two closely related *Drosophila* species, *D. melanogaster* and *D. simulans* that diverged 2.5 million years ago have highly asymmetric rapid evolution of MSL genes [117]. For instance the Mof acetylation site on Msl3 in *D.melanogaster* is unique in the *Drosophila* species [118]. It is possible that there are other selective forces that are acting on the MSL complex of *D.melanogaster* like the male killing bacteria *S.poulsonii* [119]. MSL proteins could be evolving away from recognition by these bacteria. Nevertheless since the protein complexes tend to co-evolve; the whole complex could be trying to fine tune to escape from selection while also trying to keep its essential function. Curiously, Mof has been shown to bind LTR retrotransposons in *D.melanogaster* and inhibit their transposition[120]. Inhibition of retrotransposition could also be a strong selective force on this subgroup. These recent findings can provide an explanation why it is difficult to find a consensus sequence for the MSL binding among other species.

MSL-LIKE PROTEINS IN OTHER ORGANISMS

The protein components of the MSL complex of *D. melanogaster* have clear homologues from yeast to mammals, except that Msl1 and Msl2 are not found in yeast[17]. Yeast NuA4 HAT complex contains Esa1 and Eaf3, homologues of Mof and Msl3 respectively[121]. However, stringent sequence analysis showed that yeast NuA4 complex is not the direct ancestor of compensasome in *Drosophila*, rather a novel complex arose with the concomitant evolution of Msl1 and Msl2[17]. Although humans have entirely different strategy for the dosage compensation problem, the MSL complex is kept in mammals as well[83,122,123]. However it seems that human MSL complex has evolved other functions in DNA damage response pathway and inhibition of tumor genesis[122,123]. Recently it was shown that human MSL2 is responsible for the mono-ubiquitination of p53 and

subsequent extra nuclear localization[124]. Even though mammals shows X-inactivation in female nucleus, the active homologue of X is two-fold up regulated to reach the stoichiometry of autosomal gene expression level[3,4,125,126]. The role of the MSL complex in this upregulation is a tempting hypothesis.

HOW GENERAL FACTORS MAY IMPINGE ON THE MECHANISM

One of the first issues addressed by high throughput experiments was the global regulation of expression on X by the MSL complex. A significant amount of earlier data proposed an alternative mechanism named inverse dose model for the X versus autosome balance[127-129]. This model suggests that X chromosome inherently possesses sequences that recruit transcription factors for a roughly two fold up-regulation and the MSL complex functions to titrate Mof from autosomes to X to inhibit over-expression of autosomes by overriding the effects of hyper-upregulation of X[130]. RNAi against the components of the MSL complex and stringent normalization analysis of expression arrays showed that the MSL complex was indeed acting to up-regulate X-linked genes to the autosomal levels arguing against the inverse dosage model[131,132]. Moreover artificial recruitment of the MSL complex upstream of a reporter gene can cause its up regulation and roX autosomal transgenes can overcome silent heterochromatin[133,134].

Specific enrichment of X chromosome by H4K16 acetylation led to the belief that this canonical activation mark could be largely responsible for the up regulation. Indeed Mof can activate transcription *in vivo* and *in vitro*[25]. Moreover H4K16 acetylation can decondense 30 nm chromatin fiber *in vitro*[135]. A simple prediction was that opening the chromatin might enable loading more polymerase RNA Polymerase II onto compensated genes. However, polymerase profiles show that there are not more polymerases on compensated genes than non-compensated ones[74]. Recent findings of new components related to dosage compensation tell us that the story may be much more complicated.

Identification of nuclear pore components, Nup153 and Mtor, in MSL purifications, and their effect on X linked gene expression suggest a link between dosage compensation and nuclear architecture[83]. Numerous findings indicate that position of a gene in the nuclear volume could affect its transcriptional status [136-138]. Although nuclear periphery was long known accepted as a repressive zone and a host for heterochromatin, nuclear

pore complexes (NPC) can be a docking site for an induced gene [139,140]. Close approximation of the X-linked genes to NPC may create a transcription competent domain/environment and may even provide up regulation of genes that are not bound by MSL but still dosage compensated. Interestingly, human interphase chromosomes are found to be associated with lamins in domains that are clearly demarcated by insulators showing that the genome can indeed be organized in discrete structures under the nuclear envelope[141].

Another protein found to be associated with MSL complex is the Jil-1 kinase. Jil-1 can co-immunoprecipitate with MSL components and it is enriched on male X chromosome, although it is also distributed on other chromosomes[142]. Jil-1 is the main kinase that is responsible for H3S10 phosphorylation[143]. Although this mark was known to be a mitotic marker, it is also enriched in euchromatic regions and can antagonize heterochromatin spreading[143,144]. Recently Jil-1 was shown to be an important activator in many genes in *Drosophila* and can relieve the promoter proximal pausing of RNAPII, which is thought to be a checkpoint after the initiation of transcription[145,146]. Conceptually, selective recruitment of Jil-1 kinase by MSL to the X-linked genes may relieve this pausing more than autosomes and female X; helping twice as much transcription on male X. Albeit this attractive hypothesis, an opposing experiment demonstrated that RNA Polymerase II mediated transcription is independent of H3S10 phosphorylation and Jil-1 kinase affects transcription through maintaining the structural integrity of the chromosomes[147].

It seems that the male X chromosome is generally more sensitive to perturbations related to proteins that are responsible for general chromatin morphology. Two of these proteins are NURF, a chromatin remodeler, and Su(var)3-7, a protein responsible for heterochromatin formation by the help of Hp1[148,149]. NURF is the founding member of ISWI family of remodelers and it contains ISWI protein as the catalytic subunit that enables sliding of nucleosomes[150]. Normally ISWI is not enriched on the male X chromosome or its mutations do not cause mislocalization of the MSL proteins, or the acetylation. Nevertheless, the male X chromosome looks much decondensed and broader in its absence and a functional MSL is required for this phenotype[149,151]. ISWI protein is also found in other complexes however this effect is related to NURF remodeler because aberrant phenotype of male X is repeated in Nurf301 mutations, the

main scaffold in NURF complex[152]. Recently it was found that roX null mutation could suppress the puffy appearance coming from the NURF mutations. Additionally, NURF can repress roX2 transcription in females[153]. Similar to ISWI, Su(var)3-7 mutation causes male X chromosome decondensation, which can be suppressed by null *mle* mutation[148,154]. These antagonistic relations suggest that chromatin opening is not unchecked but actually scrutinized by various complexes to maintain a sufficiently open- not more than necessary- state of chromatin. Actually, an analog system can be seen in a smaller scale on actively transcribing genes. Active genes have an increasing H3K36 di and trimethylation on the body of their genes and this mark is recognized by an HDAC complex, Rpd3S, which inhibits spurious transcription that may come from cryptic promoters[155]. Interestingly, Rpd3S, the histone deacetylase and Set2, the enzyme required for H3K36 methylation play a role in dosage compensation[27,34,82]. Curiously, the components of the exosome, Dis3 and Rrp6, also copurify with MSL proteins[83] suggesting that RNA degradation may be also be coupled to the system. In this sense exosome may degrade, antisense or cryptic transcripts that were generated uncontrolled due to open chromatin structure. Another fail-safe mechanism could be mediated by Supercoiling factor (Scf), of which genetic interaction with MSL has been shown[156]. Scf was hypothesized to help decreasing the helical torsion that may have generated during chromatin remodeling however its role in dosage compensation is not determined yet [156].

All these observations suggest that capabilities of MSL reach far beyond than expected before. Not only it behaves as a HAT complex but also acts as a mediator that fine tunes two fold upregulation by approaching to nuclear pore, cross talking with chromatin remodelers, heterochromatin proteins, and RNA degradation machines. Although the dazzling discoveries brought by powerful genetics, biochemistry and high throughput approaches, the new findings bring about their own mysteries, which eventually motivated us to find the results presented in this thesis.

RESULTS

RESULTS

Structure of the mammalian MSL1-MSL3 complex

The MSL3 construct we used corresponds to the hMSL3 isoform c (residues 167-517). It contains the predicted MRG domain but lacks the upstream chromo-barrel domain (Figure 5A). Compared to the sequence of the known structure of the MRG domain of MRG15 (PDB entry 2AQL), the human and *Drosophila* MSL3 domains contain two poorly conserved insertions with no predicted secondary structure elements (residues 223-250 and 290-441 in hMSL3). In order to obtain diffracting crystals, the longer insertion was removed and replaced with an 8-residue linker. MSL3 (167-289,442-517) was co-expressed with the C-terminal fragment of the mammalian MSL1 PEHE region (545-597) in bacteria and the structure of the complex was determined at 3Å resolution.

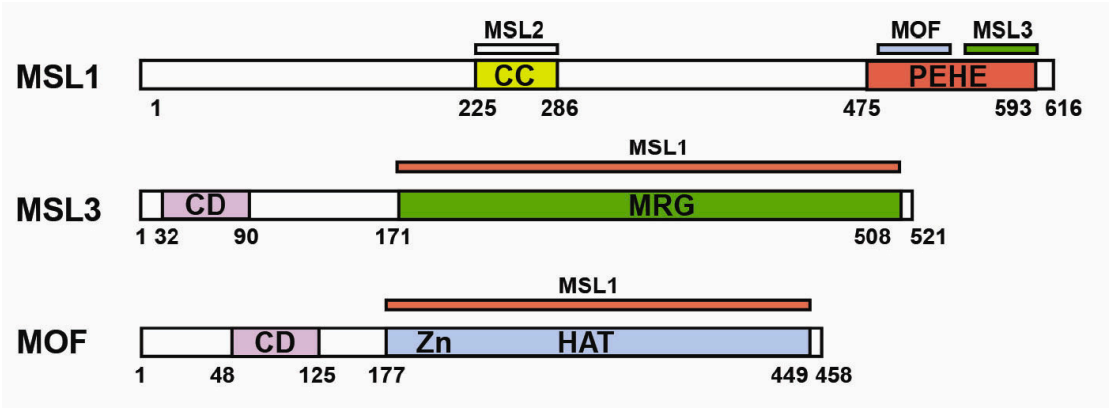
The MSL3 MRG domain structure consists essentially of a compact bundle of six helices (Figure 5B). The core made of two helical hairpins that are orthogonal to each other (α2/α3 and α5/α6) is flanked by the N- and C-terminal loops and helices α1 and α4. The 151-residue region deleted from the MSL3 construct links helices α4 and α5 (Figure 5B).

The MSL1 region interacting with MSL3 wraps around its MRG domain as an extended chain (Figures 5B), which is probably intrinsically unstructured in the absence of MSL3. MSL1 forms numerous hydrophobic as well as several charged contacts with MSL3. The crucial interacting residues of MSL1 are four highly conserved phenylalanines (Phe556, 557, 577 and 589) that insert into different hydrophobic pockets on MSL3 (Figure 6A). Essentially all MSL3 and MSL1 residues involved in the interaction are well conserved among species (Figures 6A).

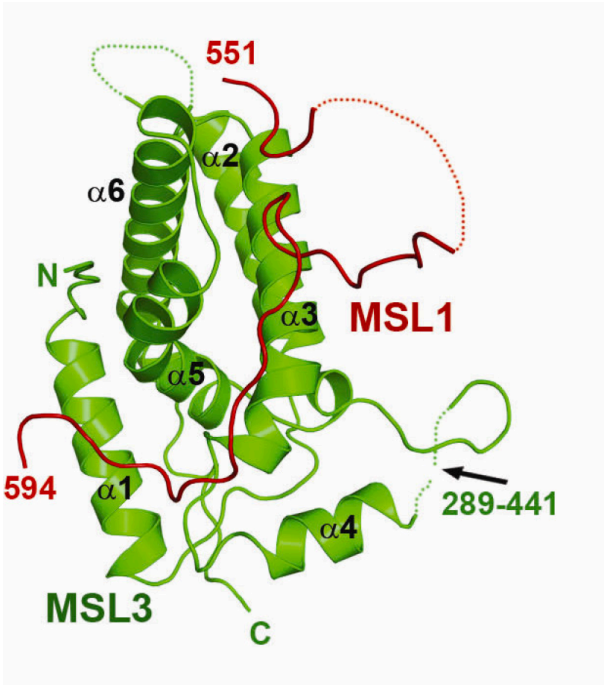
Structure of the mammalian MSL1-MOF complex

The HAT domain of human MOF (174-458) was co-expressed in bacteria with the N-terminal part of the MSL1 PEHE region (470-540). The complex was co-crystallized with acetyl-CoA, and its structure was determined by X-ray crystallography at 2.8 Å resolution (Figure 5C).

A


MSL2
MOF
MSL3
MSL1
CC
PEHE
1
225
286
475
593
616
MSL1
MSL3
CD
MRG
1
32
90
171
508
521
MSL1
MOF
CD
Zn
HAT
1
48
125
177
449
458

B


551
α2
α6
N
α3
MSL1
α5
594
α1
289-441
α4
MSL3
C

C

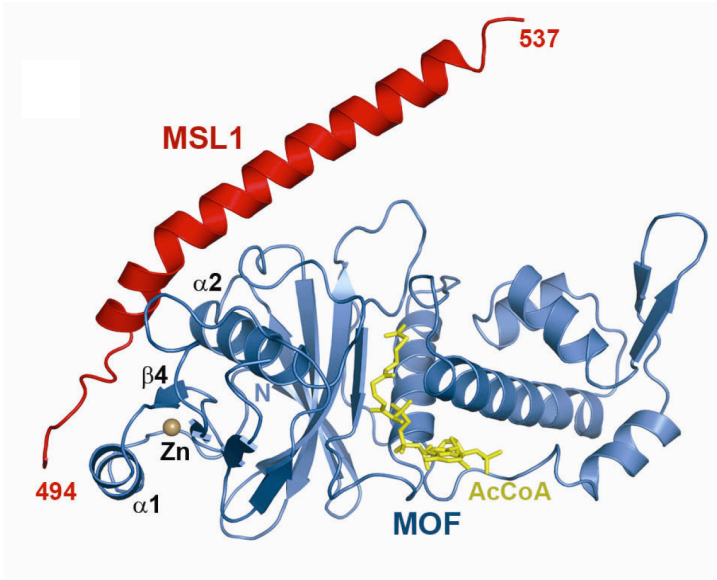

537
MSL1
α2
β4
N
Zn
494
α1
AcCoA
MOF

Figure 5. Crystal structures of the mammalian MSL1-MSL3 and MSL1-MOF sub-complexes. **(A)** Schematic representation of the domain structure of mouse MSL1 (which is essentially identical to hMSL1), hMSL3 and hMOF. Domain colors correspond to the ribbon diagram in B and C. The red, blue and green bars indicate MSL1, MOF and MSL3 interacting regions, respectively, as defined in this work and in Scott *et al*., 2000. CC, coiled coil; CD, chromo-barrel domain. **(B)** Ribbon representation of the complex between MSL1 and MSL3. The MSL3 MRG domain (residues 167-288, 442-517) is shown in green and its secondary structures are labeled. The disordered regions in MSL1 and MSL3 are shown as dots. The arrow indicates the place where residues 289-441 were deleted and replaced by an 8 amino acid linker. **(C)** Ribbon diagram of the mammalian MSL1/MOF/AcCoA complex. The HAT domain of MOF (residues 174-458) is shown in blue. The MOF secondary structures interacting with MSL1 are labeled. This figure is provided by Jan Kadlec.

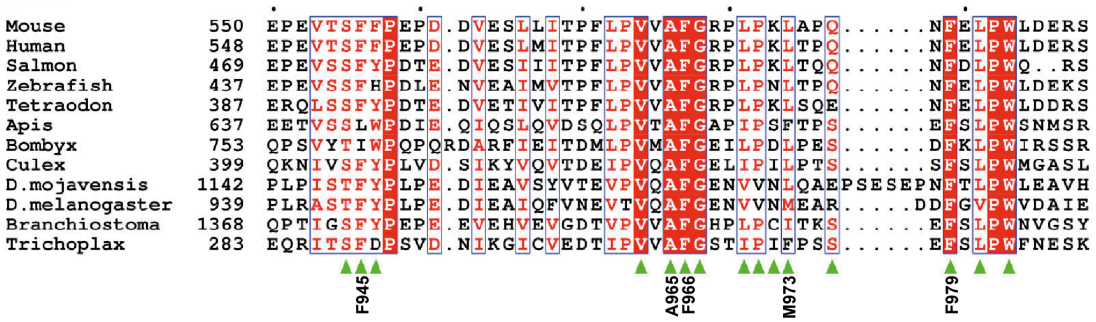
This and the MSL1-MSL3 structure (Figure 5B) clearly show that the MOF/MSL3 interacting region of MSL1 is not a genuine pre-folded domain and thus rather than PEHE domain we will refer to it as PEHE region.

Next, we analyzed the interaction interface between the MOF HAT domain and MSL1. The MSL1 fragment forms a loop (residues 494-501) followed by a 52Å long helix (residues 502-533). Both elements interact extensively with the N-terminal part of the MOF HAT domain with numerous, mainly charged contacts between the two molecules. In MSL1 the key interacting residues include Glu498, Asp502, Arg508, His509, Glu513 and Glu516 which form multiple hydrogen bonds and salt bridge interactions with MOF (Figure 7A). Additionally, Leu500, Phe505 and Leu512 are inserted in hydrophobic pockets in the center of the interface. All the MSL1 interacting residues are very well conserved among species, reflecting the importance of this interaction for the functional integrity of the MSL complex (Figure 7A).

Msl3 can be removed from the complex without any overall effect on other protein-protein interactions

To study the incorporation of MOF and MSL3 into the MSL complex *in vivo* we designed mutations in the full-length MSL1 based on predictions from the crystal structure. Since MSL proteins and their key interacting residues are evolutionary conserved and their role is better understood in *Drosophila*, we studied the effect of these mutations using Drosophila Msl1. The mutations are indicated in Figure 6B. Msl1 mutants were sub-cloned in pAc5.1 vector under the strong actin promoter and they all contain a C-terminal Flag epitope tag. Msl1 mutants were expressed in SL-2 cells by transient transfection and the corresponding MSL complexes were immunopreciptated using an anti-Flag resin. All tested Msl1 mutations in the C-terminal part of the PEHE region were unable to co-purify endogenous Msl3, while they had no effect on the remaining MSL components, Mof, Msl2 and Mle (Figure 6C).

A



B

Msl1 mut.1: wild type
Msl1 mut.2: F945E, A965E, F979E
Msl1 mut.3: F945E, A965E, F979E, M973E
Msl1 mut.4: F945E, F966E, ,F979E, M973E

C

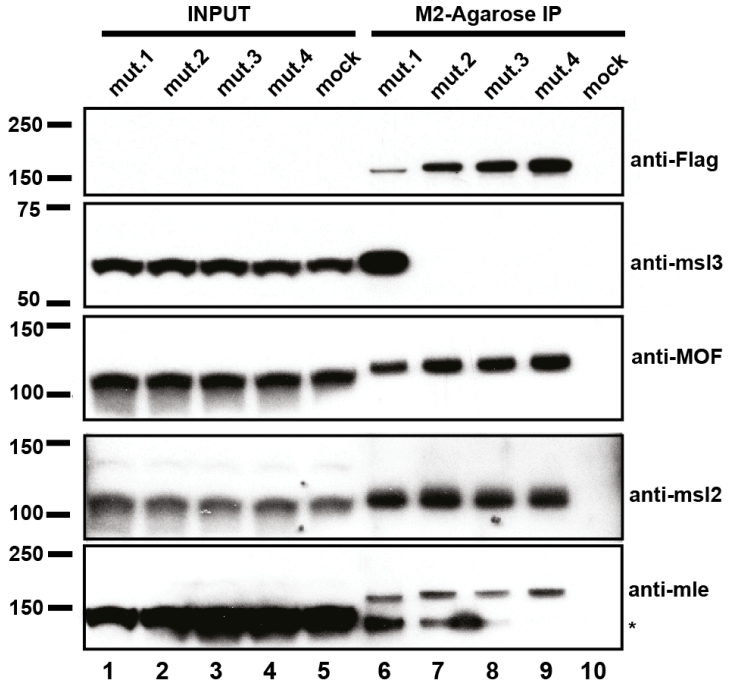


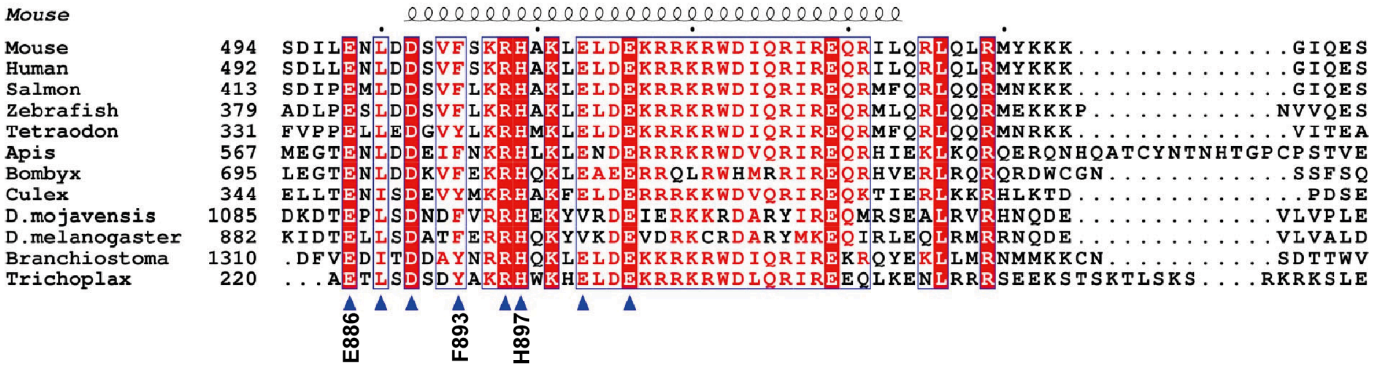
Figure 6. Msl1 mutagenesis for disruption of Msl3 interaction

(A) Sequence alignment of the MSL1 fragment that is involved in the interaction with MSL3. Identical residues are in red boxes. The interacting residues are indicated with green triangles. The five residues targeted for point mutations are shown for Drosophila melanogaster species. **(B)** Msl1 wild type with Flag epitope (mut.1) and three PEHE mutants predicted to disrupt Msl3 interactions. **(C)** Flag immunoprecipitation of Msl1 mutants in SL-2 cells. Wild type and indicated Msl1 mutants were transiently transfected in SL-2 cells. After 48 hours, the cells were harvested and immunoprecipitations using Flag-Agarose resin were carried out from whole cell extracts. Mock lane represents the empty vector. Asterisk in MLE blot is an unspecific cross-reacting band. The slight running difference between INPUT and IP lanes is due to different denaturing buffers. Anti-Flag antibody was used to detect exogenous Msl1 proteins. Transient transfections were always below the limit of detection for INPUT lanes.

Mof interaction with MSL complex can be disrupted by point mutations on Msl1

After the observation that Msl3 can be successfully dislocated from the complex, we tried to achieve the same effect for Mof. Therefore we designed point mutations on Msl1 based on the structure. The Msl1 mutants for Mof interaction are shown in Figure 7B. In this case a reduced binding of Mof was obtained for the single F893R mutation (Msl1 mut.6) in the N-terminal portion of the Msl1 PEHE region but not for single mutations E886R and H897R (Msl1 mut.5 and 7 respectively) (Figure 7C). Further reduction was observed for the

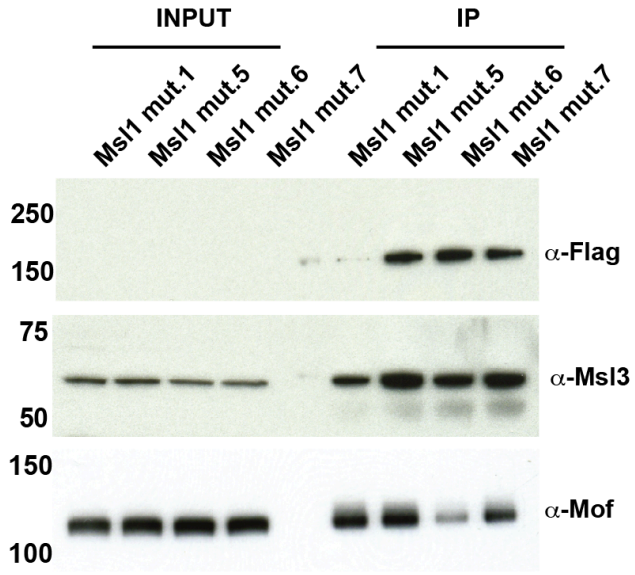
A



B

Msl1 mut.1: wild type
Msl1 mut.5: E886R
Msl1 mut.6: F893R
Msl1 mut.7: H897R
Msl1 mut.8: E886R, F893R
Msl1 mut.9: E886R, F893R,H897R

C



D

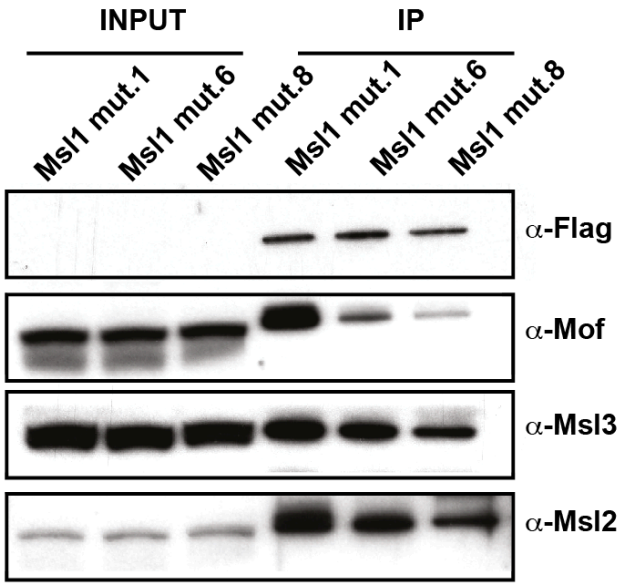


Figure 7. Msl1 mutagenesis for disruption of Mof interaction

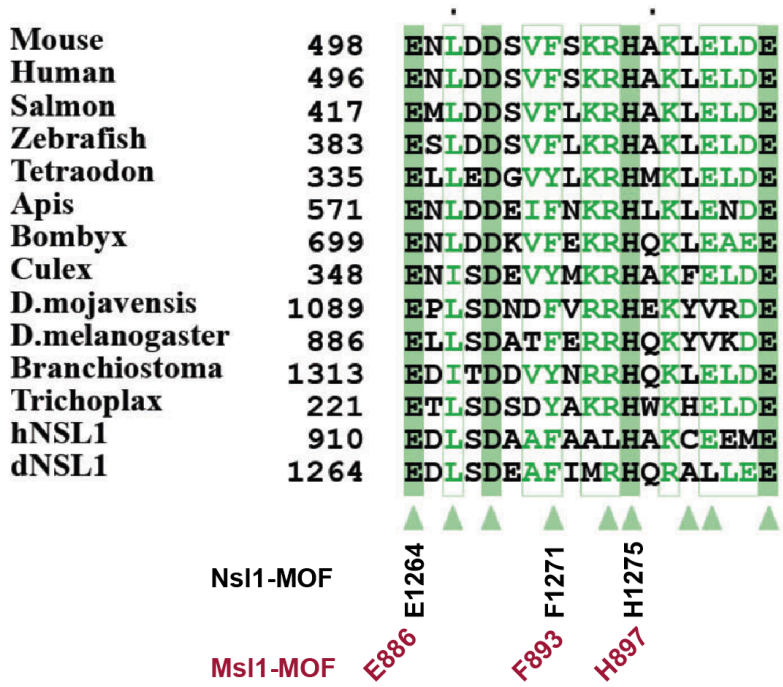
(A) Sequence alignment of the MSL1 fragment that is involved in the interaction with MOF. Identical residues are in red boxes. The interacting residues are indicated with blue triangles. The three residues targeted for point mutations are shown for Drosophila melanogaster species. **(B)** Msl1 wild type with Flag epitope (mut.1) and five PEHE mutants predicted to disrupt Mof interactions. **(C)** Flag immunoprecipitations from whole cell extracts of SL-2 cells transiently transfected with Msl1 mut.1,5,6 and 7. Western blots are shown for the indicated antibodies. Flag INPUT signals were below detection limit. **(D)** Same experiment as in C, including Msl1 mut.8.

E886R/F893R double mutant (Msl1 mut.8) (Figure 7D). A partial reduction was observed also for Msl3 and Msl2 incorporation, suggesting that the presence of Mof in the complexmight be important for Msl1 stability (Figure 7D). These results indicate that Msl3 and Mof are incorporated into the MSL complex via the Msl1 scaffold and show that at least Msl3 can be disassembled from the complex without an apparent effect on the molecular interactions of other members of MSL complex.

PEHE region of Nsl1 utilizes similar interaction network for MOF as in Msl1

MOF resides in two functionally distinct complexes in *Drosophila* as well as in mammals namely the MSL complex and the novel NSL complex[83]. There is a high degree of conservation between MSL1 and NSL1 in the N-terminal part of the PEHE region (Figure 8A). To investigate whether Nsl1 uses the same interaction surface for Mof contact *in vivo*, amino-acids predicted to interact with Mof were mutated in full length *Drosophila* Nsl1 and the mutant proteins were expressed in SL-2 cells. Remarkably, the Nsl1 mutant E1264R/F1271R showed a strong loss of Mof interaction (Figure 8B) whereas keeping MBDR-2 and Nsl3 still interacting. This indicates that Mof uses similar surfaces for the integration into either NSL or MSL complexes in *Drosophila* cells.

A


Mouse 498 ENLDDSVFSKRHAKLELDE
Human 496 ENLDDSVFSKRHAKLELDE
Salmon 417 EMLDDSVFLKRHAKLELDE
Zebrafish 383 ESLDDSVFLKRHAKLELDE
Tetraodon 335 ELLEDGVYLKRHMKLELDE
Apis 571 ENLDDEIFNKRHLKLENDE
Bombyx 699 ENLDDKVFEKRHQKLEAEE
Culex 348 ENISDEVYMKRHAKFELDE
D.mojavensis 1089 EPLSDNDFVRRHEKYVRDE
D.melanogaster 886 ELLSDATFERRHQKYVKDE
Branchiostoma 1313 EDITDDVYNRRHQKLELDE
Trichoplax 221 ETLSDSDYAKRHWKHELDE
hNSL1 910 EDLSDAAFAALHAKCEEME
dNSL1 1264 EDLSDEAFIMRHQRALLEE
Nsl1-MOF
E1264
F1271
H1275
Msl1-MOF
E886
F893
H897

B

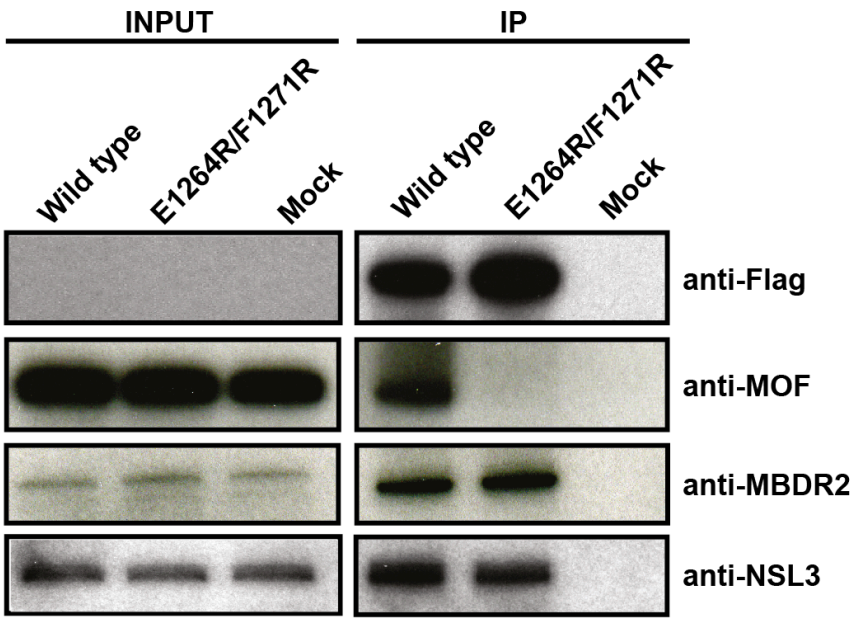

INPUT
IP
Wild type
E1264R/F1271R
Mock
Wild type
E1264R/F1271R
Mock
anti-Flag
anti-MOF
anti-MBDR2
anti-NSL3

Figure 8. Nsl1 PEHE region interacts with Mof similar to Msl1

(A) Alignment of PEHE regions of MSL1 proteins together with Drosophila and human NSL1. Identical residues are indicated with green boxes. Green triangles indicate contact points with MOF. Provisional point mutations are shown on Drosophila Nsl1 and corresponding residues of Msl1 are shown in red. **(B)** Flag immunoprecipitations from whole cell extracts of SL-2 cells transiently transfected with wild type Flag tagged Nsl1 and Flag tagged E1264R/F1271R double mutant. Mock represents empty vector transfection. Western blots are shown for the indicated antibodies. Flag INPUT signals were below detection limit. Anti-Flag signal corresponds to Nsl1.

Mof and Msl3 interaction is mostly mediated by Msl1

The observation that Msl3 and Mof can be removed from the complex without affecting each other considerably raises the question as to the significance of the interaction between Mof and Msl3, reported by Buscaino *et al*[27]. To further investigate this putative interaction, which was not confirmed by Morales *et al.* [15], we performed MSL complex reconstitution assays with the full-length proteins expressed in Sf21 insect cells. In the presence of Msl1, Mof clearly co-purifies with Msl3 (Figure 9), however in the absence of Msl1, this interaction could be seen only using western blot detection (Figures 9), suggesting that the Msl3-Mof interaction does occur marginally *in vitro*, albeit significantly weaker than in a trimeric complex.

Msl1 can be localized to chromatin and X chromosome without Mof or Msl3

In order to observe the consequences of taking Mof or Msl3 from the complex, and whether the transiently expressed proteins are incorporated/targeted to chromatin, we performed chromatin fractionation assays, where first nuclei were separated from cytoplasm and the nucleoplasm was extracted at physiological salt concentration (150 mM NaCl) with detergent perforation of the membrane. The remaining chromatin fraction was solubilized by nucleases and all pools were analyzed by immunodetection. In wild type cells, all MSL members have both nucleoplasmic and chromatin distributions with an enrichment in the chromatin bound pool (Figure 10B). Interestingly, Mof can also be observed in the cytoplasm. Upon increase of salt

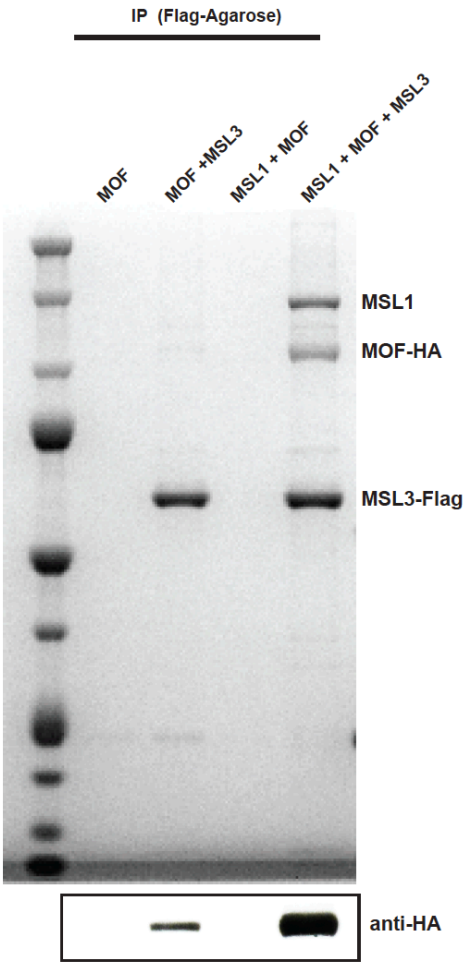


Figure 9. Mof interaction with Msl3 is largely mediated by Msl1

(A) Msl3-Flag, Mof-HA and non-tagged Msl1 proteins were expressed individually in Sf21 insect cells. Equal amounts of whole cells extracts were mixed and anti-Flag resin was used to pull down the Msl3. After washing, the proteins were eluted by Flag peptides and TCA precipitated. Elutions were analyzed on a gradient gel by Coomasie staining. Western blot analysis of HA antibody to detect Mof is shown below.

concentration, most members detach from chromatin pool and remain nucleoplasmic, however even in the stringent condition, a slight pool of MSL members remain on the chromatin. (Figure 10B). For chromatin localization, we chose one mutant for the loss of Msl3 (Msl1 mut.2) and one for Mof (Msl1 mut.8) and repeated the experiment under physiological conditions. When Msl1-Flag and the mutants were transiently expressed, they were mostly detected in the chromatin fractions indicating that our constructs were incorporated into chromatin similar to endogenous MSL complexes (Figure 10C).

Next, we wanted to see if X chromosomal targeting is affected by the disruption of Msl1 interaction with either Mof or Msl3. In order not to exceed physiological protein levels, we expressed the constructs under the cupper inducible MtnB promoter under

uninduced conditions and used an anti-Flag antibody to visualize the exogenous Msl1-Flag and mutant derivatives by immunofluorescence (IF) microscopy. MSL members are known to show a crescent shape in the nucleus in SL-2 IF cells, named as nuclear periphery[83]. All constructs were able to target to the X chromosome showing co-localization with endogenous MOF (Figure 11).

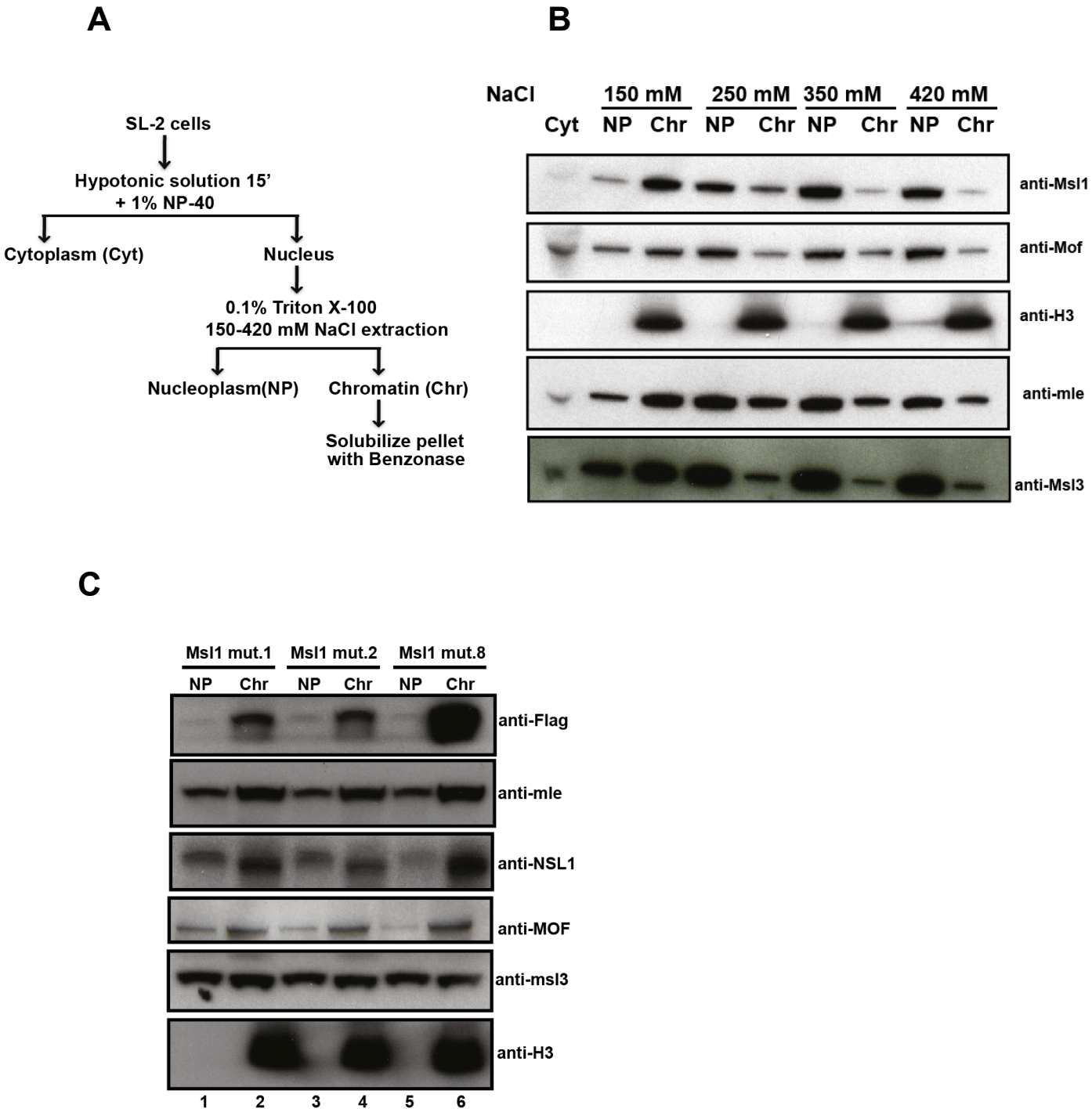


Figure 10. Fractionation assay in SL-2 cells

(A) Schematic representation of the flow of the experiment. **(B)** Fractionation experiment to monitor endogenous MSL proteins under increasing salt concentrations. Equal amounts of each fraction is resolved on SDS-PAGE and blotted with indicated antibodies. Histone 3 was used as a positive control for the chromatin pool. **(C)** Fractionation assay for Msl1 mut.1 (wt), mut.2 (Msl3 losing mutant) and mut.8 (Mof losing mutant) under 150 mM salt concentration. Flag antibody was used to detect the exogenously expressed Msl1 proteins. Only nucleoplasmic (NP) and chromatin pools (Chr) are shown. Endogenous proteins were detected by their respective antibodies.

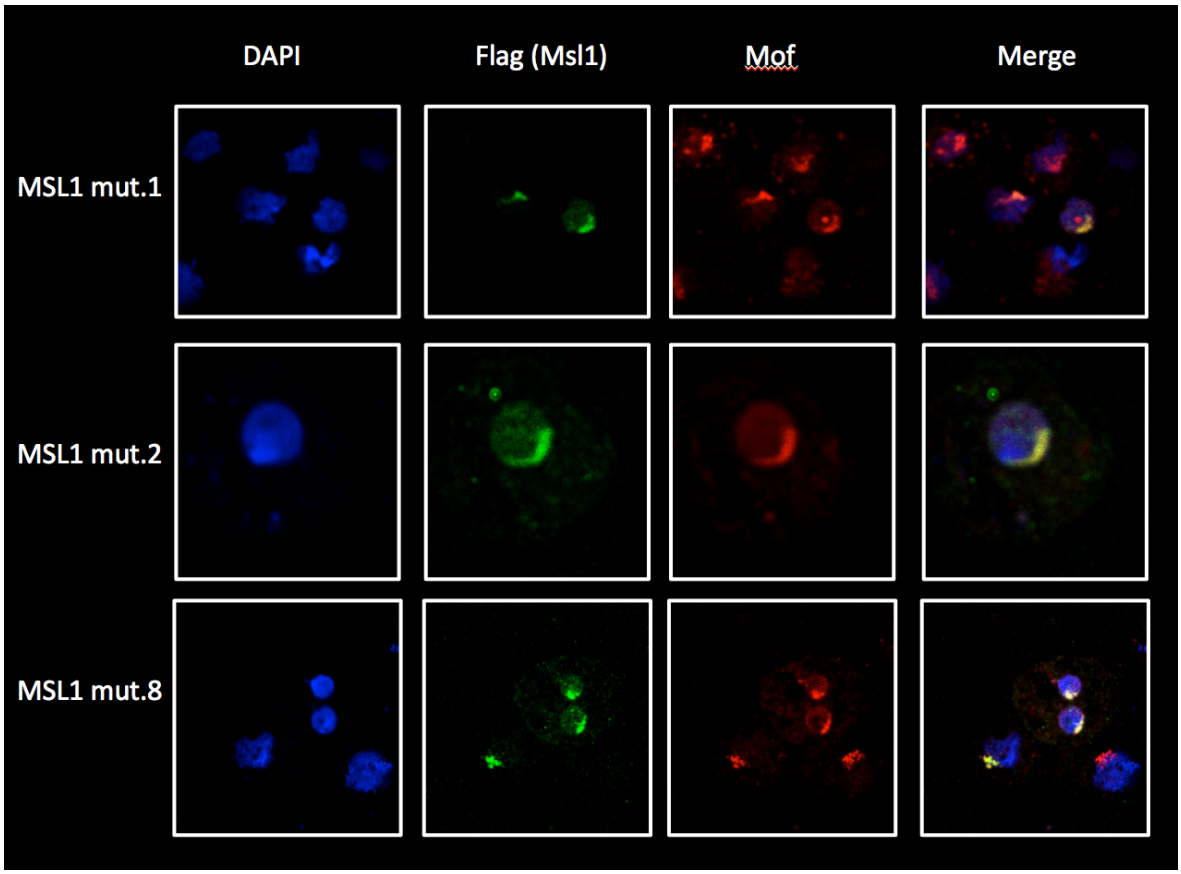


Figure 11. X chromosome localization can be seen for Msl1 mut.2 and mut.8

Immunofluorescence of SL-2 cells expressing Msl1-Flag and its derivatives under leaky MtnB promoter. Transiently transfected cells were spun in a cytospin machine to poly-Lysine coated slides. Msl1-Flag and endogenous MOF were detected by the indicated

antibodies and DNA was detected by DAPI staining. Due to transient transfection, only some of the cells express the Msl1 and its mutants.

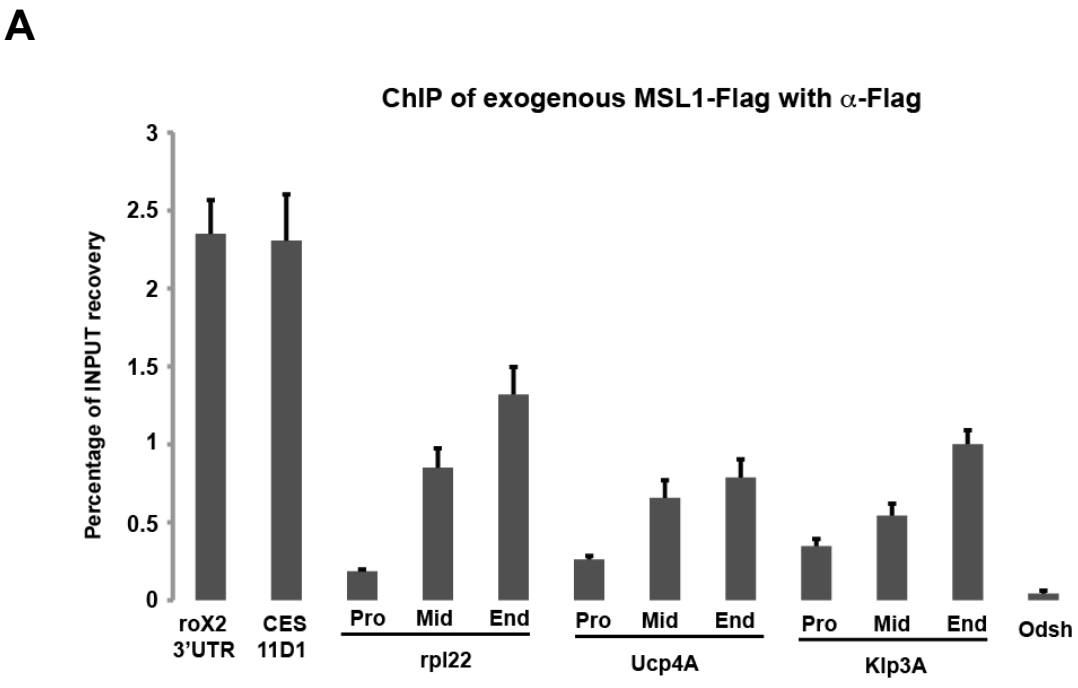
Consequence of loss of Msl3 or Mof on the localization of Msl1 on X chromosome

Since immunofluoresence microscopy does not provide sufficient resolution to observe targeting to individual loci, we decided to perform chromatin immunoprecipitation of the Msl1 derivatives on X-linked genes using an anti-Flag antibody to specifically pull down the exogenous proteins. Our first trials to perform ChIP with transiently transfected cells were not successful, possibly due to low efficiency of transient transfection. To generate a stable line, it is required to co-transfect the selection cassette together with the plasmid of interest and the ratio between these plasmids is crucial and requires optimization. In order to overcome this optimization, we generated a new vector that expresses Neomycin cassette under Actin promoter and SV40 terminator. Another multiple cloning site also exists under MtnA promoter, which enables us to induce expression with $CuSO_4$. We call this plasmid pIBU1 and all our stable cell lines were generated with this plasmid. We also replaced Flag epitope at C terminal of Msl1 constructs with 3x Flag-6His to enhance the recovery with flag epitope. 3xFlag 6 His epitopes would also enable us to perform tandem affinity purifications. We first confirmed that the 3xFlag epitope containing Msl1 and the mutant cognates showed identical co-immunoprecipitation behaviours as their single Flag carrying counterparts (data not shown). MSL proteins have been shown to enrich towards the 3'UTR of ORFs on X- linked genes (See introduction). After establishing ChIP protocol in Sl-2 cells with endogenous Msl1 (Figure 12A), we made a minor modification in the protocol including a bridge antibody amplification step for enhancing the Flag signal (See Methods). Upon these modifications, we were able to show that Msl1-3xFlag showed similar binding profiles as the endogenous MSL1 (Figure 12B), displaying the quality of the Flag ChIP. When this method applied to all mutants, we observed that in contrast to the wild-type Msl1, Msl1 E886R/F893R (mut.8) and MSL1 F945E/A965E/F979E (mut.2) showed significantly reduced binding on the body of X-linked genes (Figure 13A). We also observed a low but consistent Msl1 signal towards

promoters of the X linked genes. Interestingly, this signal remained largely unaffected in MSL1 derivatives (Figure 13A). We will examine the importance of this binding further.

High affinity sites are qualitatively different

We next asked whether compromised chromatin binding of the Msl1 mutants is restricted to low affinity sites or whether targeting to high affinity sites was also impaired. For this purpose, we chose thirteen different high affinity sites recently mapped by Kuroda and colleagues[77] and compared the binding profiles of wild-type Msl1 and its mutant derivatives. The roX2 gene was used as a control as it is a high affinity site for MSL complex assembly and it has been shown previously that MSL1 binding on this site is independent of Msl3 or Mof [87]. Since these high affinity sites are located at different loci on the X chromosome, we separated them into positional categories (promoter proximal, 5'UTR, exon, intron and 3'end) to investigate any site-specific differences. Interestingly, this analysis revealed that disruption of MOF or MSL3 interactions also affects optimal binding of Msl1 to these high affinity sites especially when they were located away from promoter regions (Figure 13B). However, for sites that are promoter proximal such as 2E1 and 2C4, Msl1 mutants remained bound at comparable levels to the wild-type Msl1 (Figure 13B).



B

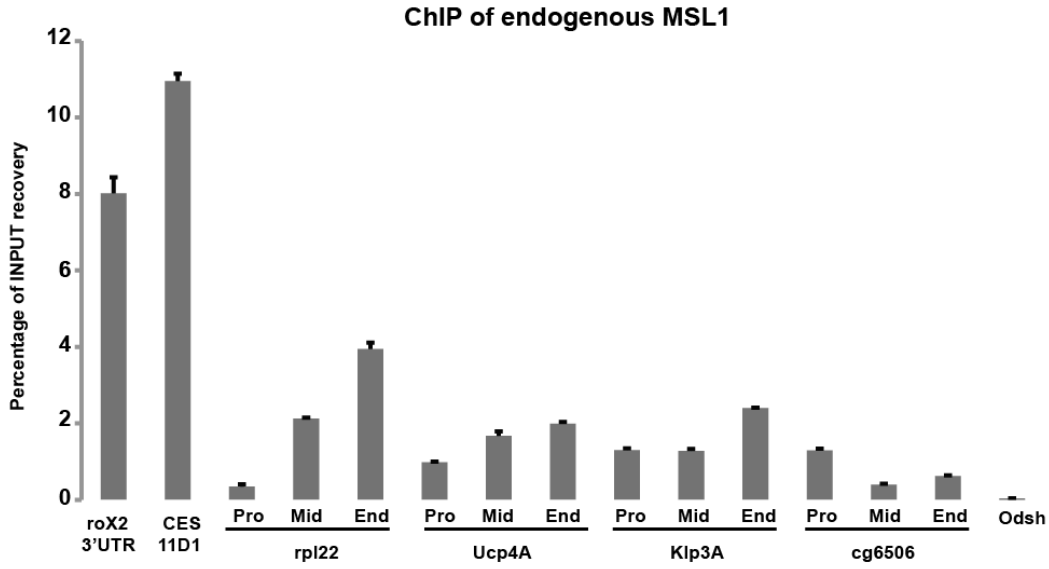


Figure 12. ChIP from wild type SL-2 cells.

(A) Anti-Msl1 antibody was used to show the binding profile of endogeneous Msl1. Rox2 and CES11D1 are High Affinity sites and represent high recovery of Msl complexes and serve as positive controls. Four X- linked genes were tested as dosage compensated genes and ODSH downstream intergenic region was used as a negative control. Each gene was probed with three different primer pairs targeted to the promoter (Pro), middle (Mid) and 3'UTR (End). Each bar represents the average of three independent IPs and the error bars are standard deviations. **(B)** ChIP from SL-2 cells stably expressing Msl1-Flag. The cells were induced with 0.5 μM $CuSO_4$ for 12 hours and immunoprecipitation was performed by using anti-Flag mouse monoclonal antibody. The immunocomplexes were collected with blocked Protein A-Sepharose beads. To enhance the recovery, a bridging anti-mouse antibody produced in rabbit had been coupled to the beads. Same primer pairs were used as in (A) to show the similar binding profile of endogenous Msl1 and Msl1-Flag.

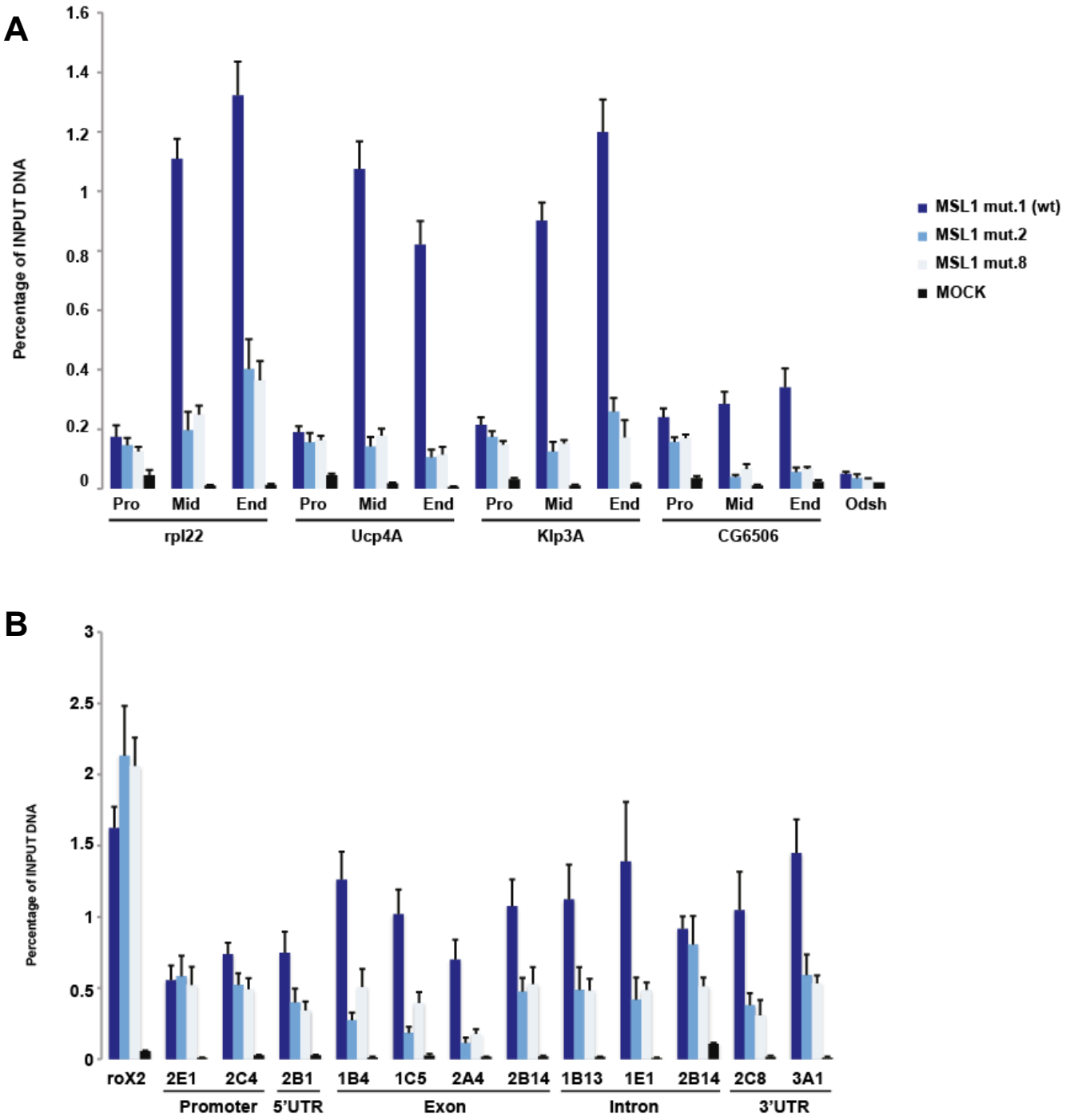


Figure 13. Effect of loss of Msl3 and Mof on MSL complex targeting on X chromosome

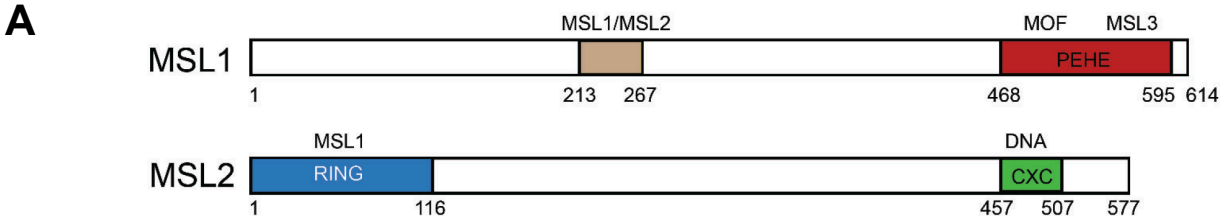
(A) ChIP of Msl1-Flag and derivatives. Stable cell lines expressing wild type Msl1-3xFlag or derivatives, under cupper inducible promoters, were grown to same density and induced with 0.5 μM $CuSO_4$. The ChIP was performed as in Figure 12B. Quantitative real time PCR was performed to the regions corresponding to promoters (P), middle (Mid) and 3'UTR of the genes (End). All tested genes are dosage compensated and located on

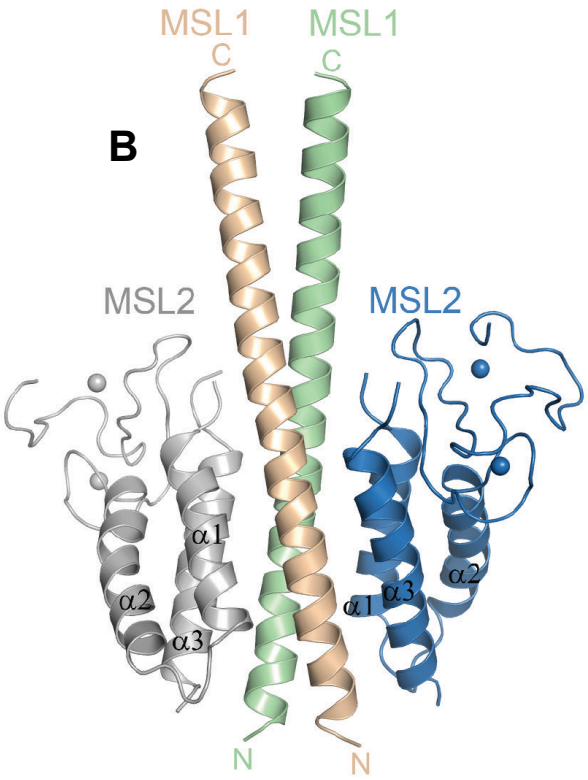
the X chromosome. ODSH upstream region was used as a negative control for the binding of the MSL complex. Each bar represents the average of four independent IPs and error bars indicate the standard deviation. Mock sample corresponds to wild-type cells. **(B)** ChIP of Msl1-3xFlag and derivatives on High Affinity Sites. The experiment was performed as explained in (A). 13 High Affinity Sites were chosen from[77]. The sites were grouped according to their genomic positions. Note that two 2B14 sites are different.

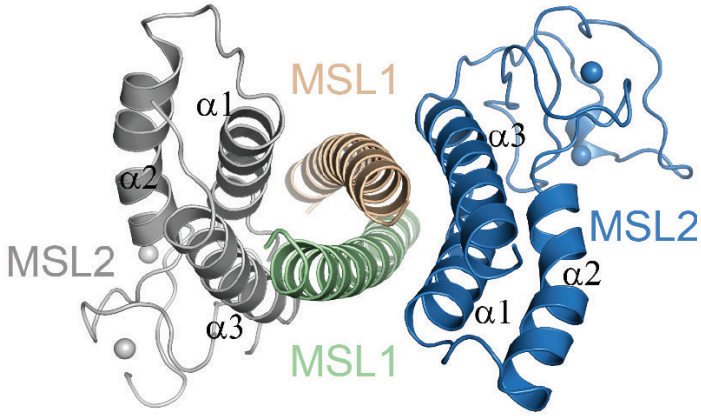
MSL1 and MSL2 form a heterotetrameric core of the MSL complex

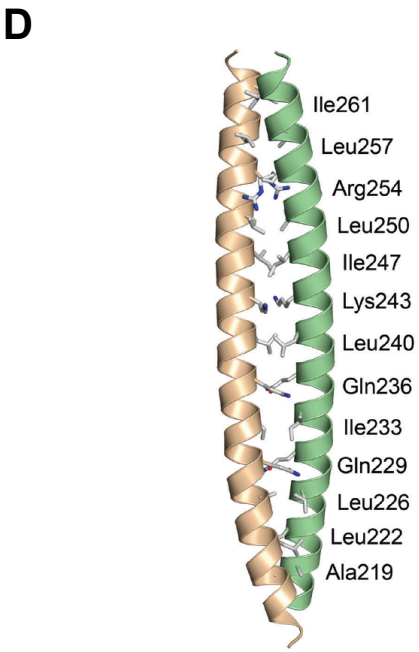
The complex between the predicted coiled-coil region of human MSL1 (residues 213-310) and the N-terminal portion of MSL2 (residues 1-116) was formed by co-expression in bacteria. Using trypsin limited proteolysis we identified a shorter MSL1 fragment spanning residues 213-267 that was sufficient for the MSL2 binding (Figure 14A). The structure of this complex was determined by X-ray crystallography at a resolution of 3.5Å (Figure 1). MSL1 and MSL2 proteins were originally suggested to dimerize via their putative coiled coil regions[13,14]. However unexpectedly, both of our structures show that instead, these two proteins form a heterotetrameric core of the MSL complex, where two MSL1 subunits form a dimeric coiled-coil which serves as a binding platform for two molecules of MSL2 (Figure 14B,C).

The crystallized fragment of MSL1 (residues 213-267) forms a 75Å long parallel dimeric coiled coil, where 10 hydrophobic and 4 polar residues (Gln229, Gln236, Lys243 and Arg254), that were originally thought to be involved in the interaction with MSL2 [14,14], pack in layers with a regular heptad (3-4) periodicity (Figure 14D,E). The coiled coil also contains other stabilizing interactions between Gln236 and Gln237, Lys243 and Glu244 or Glu253 and Arg254. The dimer's two Glu229 and Glu236 residues form respectively interhelical hydrogen bonds at its core (Figure 14E). Most of the residues involved in the MSL1 dimerization are highly conserved across species, reflecting the importance of this interaction for the functional integrity of the MSL complex (Figure 15A). Upon dimerization, the MSL1 coiled coil forms two composite, mostly hydrophobic binding sites for two molecules of MSL2.


A
MSL1
MSL1/MSL2
MOF
MSL3
PEHE
1
213
267
468
595
614
MSL2
MSL1
RING
DNA
CXC
1
116
457
507
577


B
MSL1
C
MSL1
C
MSL2
MSL2
α1
α2
α3
α1
α3
α2
N
N


C
MSL1
α1
α3
α2
MSL2
MSL2
α2
α1
α3
MSL1


D
Ile261
Leu257
Arg254
Leu250
Ile247
Lys243
Leu240
Gln236
Ile233
Gln229
Leu226
Leu222
Ala219

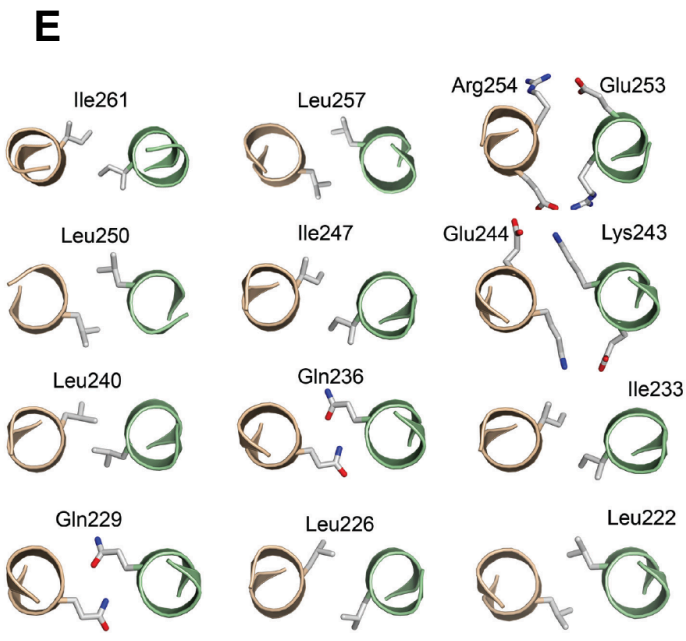

E
Ile261
Leu257
Arg254
Glu253
Leu250
Ile247
Glu244
Lys243
Leu240
Gln236
Ile233
Gln229
Leu226
Leu222

Figure 14. Crystal structures of the MSL1-MSL2 complex.

(A) Schematic representation of the domain structures of human MSL1 and MSL2. The binding partners are indicated above individual domains. **(B)** Ribbon diagram of the human $MSL1_{213-267}$/$MSL2_{1-116}$ complex. Two molecules of MSL1 form the central dimeric coiled coil (shown in brown and green). The N-terminal RING finger containing domains of MSL2 are shown in blue and gray. **(C)** The $MSL1_{213-252}$/$MSL2_{1-116}$ structure rotated by 90° along the horizontal axis relative to (B). **(D)** Ribbon diagram of the MSL1 dimeric coiled coil. Residues at the a and d heptad positions are labeled. (E) Details of the MSL1 dimerization contacts. This figure is provided by Jan Kadlec.

It is important to note that *Drosophila* Msl1 was originally proposed to self-associate via a so called glycine-rich region between residues 26-84, possibly mediating oligomerization of MSL complexes on male X-chromosome[14]. However, we show that MSL1 forms dimers (rather than higher oligomers) via its coiled coil region while the upstream glycine rich region is not required for the dimerization.

MSL2 was suggested to interact with MSL1 via its RING finger[16]. In contrast, our structure shows that its interaction with the MSL1 dimer is exclusively mediated by helices α1 and α3, while the RING finger has no contact with MSL1. Interestingly, the putative role of the *Drosophila* Msl2 RING finger in the interaction with Msl1 was established by identification of 13 mutations, which in light of the present structure would nearly all destabilize the RING finger and thus probably also the entire Msl2 [16]. Only 2 of these mutations (M14K and C107R) would probably directly affect the binding of helix α1 and α3 to Msl1. The helices of the two MSL2 molecules bind to MSL1 in an anti-parallel fashion forming an eight-helical bundle (Figure 14B,C) with multiple contacts within several hydrophobic and polar layers along the first three heptad repeats of MSL1. The key interacting residues of MSL1 form a short highly conserved cluster between Ser117 and Gln239 (Figure 15A).

MSL1 dimerization is independent of and prerequisite to MSL2 binding

Since the role of the MSL complex is better understood in *Drosophila* and the key residues in all interaction interfaces MSL1 makes with MSL1, MSL2, MSL3 and MOF are evolutionary conserved, we performed all our functional studies with *Drosophila* proteins in cell lines as well as transgenic flies. All the *Drosophila* Msl1 mutants used for dimerization studies and the corresponding mutations to human counterparts are summarized in Figure 15 and they all have a C-terminal 3xFlag epitope unless indicated otherwise.

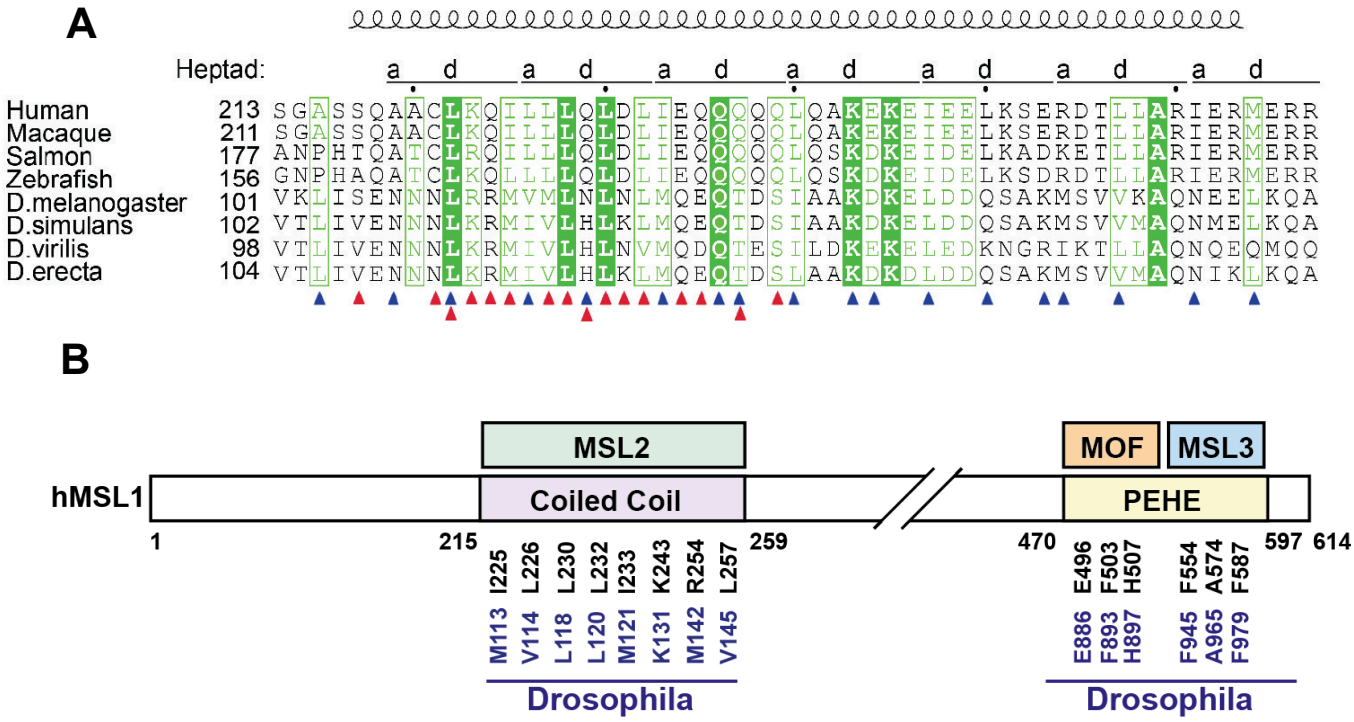


Figure 15. Generation of Drosophila Msl1 dimerization mutants

(A) Sequence alignment of MSL1 proteins comparing vertebrates and Drosophila species. Only the sequence of the coiled coil region is shown. Identical residues are in green boxes and conserved residues are shown in green. Blue triangles indicate residues involved in the MSL1 dimerization while red triangle show residues interacting with MSL2. **(B)** Mutated residues in Drosophila and their human homologoues are represented on the human MSL protein scheme. **(C)** Drosophila Msl1 mutants used in this study. All Msl1 mutants, including wild type, have a C-terminal 3xFlag tag. **(D)** Flag IP from whole cell extract of SL-2 cells transiently expressing Msl1 mut.1 and mut.10.

Previously we showed that the Msl3 and Mof interactions with Msl1 can be disrupted without any apparent influence on the other protein-protein interactions within the complex. To further support this finding and functionally separate the N-terminal interactions of Msl1 with Msl2 and the C-terminal interactions with Mof and Msl3 (through the PEHE region), we generated an Msl1 mutant (Msl1 mut.10) that binds neither Msl3 nor Mof (Figure 15D). Using co-immunoprecipitation we could show that the Msl1 interaction with Msl2 remains unaffected even when both Msl3 and Mof are eliminated from the complex.

To test the dimerization of the full-length Msl1 *in vivo*, we transiently co-expressed the wt Msl1-Flag and Msl1-myc proteins and immunoprecipitated Msl1-Flag bound proteins using a Flag antibody-coupled resin. Indeed, we could show that the Flag tagged Msl1 co-immunoprecipitated with Msl1-myc as well as Msl2, Msl3 and Mof (Figure 16A lane 1). Furthermore, we observed that Msl1 can dimerize even in the absence of Msl3 and Mof (Figure 16A lane 2). Next, we were interested in identifying Msl1 mutations that would disrupt its dimerization, without directly affecting the residues interacting with Msl2. Thus, we mutated either 4 or 5 residues at *a* or *d* heptad positions along the coiled coil to aspartates (Msl1 mut.11 and 12). Both mutants, although they were expressed more, failed to co-purify Msl1-myc and Msl2 while the

interaction with Msl3 and Mof was unaffected (Figure 16A lane 3 and 4). This experiment confirms that the interaction with Msl2 requires the entire composite Msl2 binding site formed by the Msl1 dimer (Figure 14B) while the monomeric Msl1 is not sufficient. It is important to note that, neither the Msl1 dimerization nor Msl2 binding is required for the interaction with Msl3 and Mof. These results emphasize the modular nature of Msl1 interactions with different members of the MSL complex.

Next we designed a mutant that would not interact with Msl2 but would preserve the integrity of the Msl1 dimer. Thus, we mutated three residues in the Msl1/Msl2 interface that do not lie at *a* or *d* heptad positions to arginines (M113R, L118R, L120R: Msl1 mut.13). The Msl1 mut.13 was still able to dimerize with Msl1-myc, bind Msl3 and Mof, while the interaction with Msl2 was lost (Figure 16A, lane 5) indicating that the presence of Msl2 is not required for the Msl1 dimerization. Finally, we showed that a single additional mutation in *a* heptad position (V114E) was sufficient to disrupt directly both Msl1 dimerization and Msl2 binding (Figure 16A, lane 6). Similar results were obtained when we repeated the co-IP experiments with an HA tagged wild type Msl1 (Figure 16B). To further support the hypothesis that Msl1 dimer can exist without Msl2, we performed the co-IP experiments in Kc cells, a cell culture model for *Drosophila* female cells, where Msl2 translation is inhibited (Figure 16C). In these cells, wt Msl1 forms a dimer (Figure 16C, lane 1) and loss of Msl3 and Mof does not affect the dimerization (Figure 16C, lane 2). Msl1 mut.11 and 12, which shows abolished dimer formation in SL-2 cells also show compromised

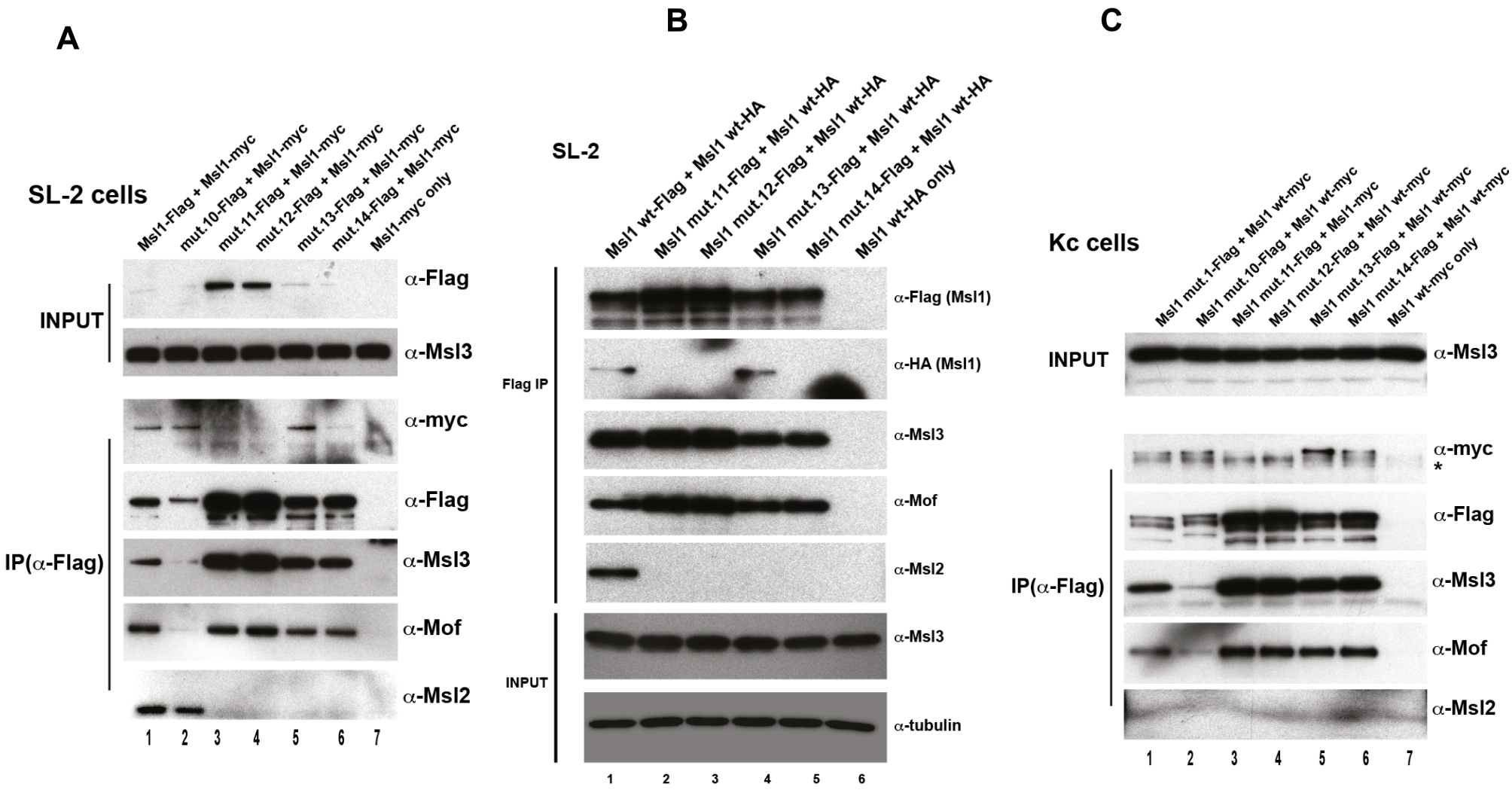

A
SL-2 cells
Msl1-Flag + Msl1-myc
mut.10-Flag + Msl1-myc
mut.11-Flag + Msl1-myc
mut.12-Flag + Msl1-myc
mut.13-Flag + Msl1-myc
mut.14-Flag + Msl1-myc
Msl1-myc only
INPUT
IP(α-Flag)
α-Flag
α-Msl3
α-myc
α-Flag
α-Msl3
α-Mof
α-Msl2
1 2 3 4 5 6 7
B
SL-2
Msl1 wt-Flag + Msl1 wt-HA
Msl1 mut.11-Flag + Msl1 wt-HA
Msl1 mut.12-Flag + Msl1 wt-HA
Msl1 mut.13-Flag + Msl1 wt-HA
Msl1 mut.14-Flag + Msl1 wt-HA
Msl1 wt-HA only
Flag IP
INPUT
α-Flag (Msl1)
α-HA (Msl1)
α-Msl3
α-Mof
α-Msl2
α-Msl3
α-tubulin
1 2 3 4 5 6
C
Kc cells
Msl1 mut.1-Flag + Msl1 wt-myc
Msl1 mut.10-Flag + Msl1 wt-myc
Msl1 mut.11-Flag + Msl1 wt-myc
Msl1 mut.12-Flag + Msl1 wt-myc
Msl1 mut.13-Flag + Msl1 wt-myc
Msl1 mut.14-Flag + Msl1 wt-myc
Msl1 wt-myc only
INPUT
IP(α-Flag)
α-Msl3
α-myc
*
α-Flag
α-Msl3
α-Mof
α-Msl2
1 2 3 4 5 6 7

Figure 16. Msl1 dimerization is independent of and prerequisite to Msl2 binding

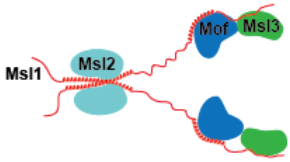
(A) Flag immunoprecipitation of Msl1 mutants in SL-2 cells. Msl1-Flag mutants are co-transfected with wild type myc-tagged Msl1 and Flag beads were used for IP. Western blots were performed with the indicated antibodies. Flag and myc tag indicates C-terminal 3xFlag and 3xMyc tag, respectively. **(B)** Experiment repeated with HA tagged wilt type Msl1 as in (A) **(C)** Same experiment in (A) performed in Kc cells. Asterisk in anti-myc blot indicates a contamination band. Msl2 absence is a marker for Kc cells.

---

dimer formation in Kc cells while mut.13 still dimerizes as predicted. Interestingly, Msl1 mutants that lose the Msl2 interaction were consistently observed to be more abundant than the wild type and mut.10, indicating a possible effect of Msl2 on Msl1 turnover. The schematic summary of the all the mutant Msl1 containing complexes is represented in Figure 17. Taken together, these results conclusively show that Msl1 dimeric coiled coil is a platform for Msl2 interaction *in vivo* and PEHE domain interactions are rather independent from MSL2 interaction, furthermore supporting the modular nature of the MSL complex.

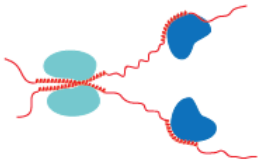
Full MSL coplex

Msl1 mut.1: wild type



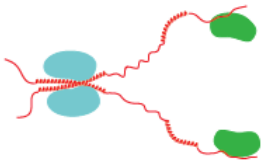
Msl3 losing mutants

Msl1 mut.2: F945E, A965E, F979E
Msl1 mut.3: F945E, A965E, F979E, M973E
Msl1 mut.4: F945E, F966E, F979E, M973E



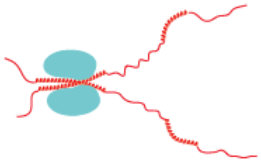
Mof losing mutants

Msl1 mut.5: E886R
Msl1 mut.6: F893R
Msl1 mut.7: H897R
Msl1 mut.8: E886R, F893R



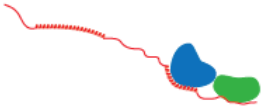
Mof and Msl3 losing mutant

Msl1 mut.10: E886R, F893R,H897R
F945E, A965E, F979E,



Dimerization mutants

Msl1 mut.11:V114D,M121D,M142D,V145D
Msl1 mut.12:V114D,M121D,M142D,V145D,K131D
Msl1 mut.14:M113R,L118R,L120R,V114E



Msl2 losing mutant

Msl1 mut.13:M113R,L118R,L120R

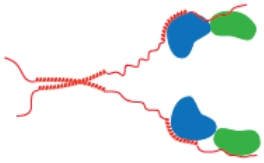


Figure 17. Summary of Msl1 mutants and the partial MSL complexes

Msl2 is an E3 ligase

Our structure revealed that RING finger of Msl2 is not involved in Msl1 interaction but rather solvent exposed, indicating that this fold can act as an enzyme. We thus tested the ubiquitination activity of *Drosophila* Msl2 using purified full-length protein expressed in Sf21 insect cells in an *in vitro* ubiquitination assay. We could show that Msl2 can auto-ubiquitinate itself, which is a hallmark of E3 ligase proteins (Figure 18). Msl1, which served as a negative control, did not exhibit any ubiquitination activity. Surprisingly, in the presence of Msl2, Msl1 also showed higher molecular species indicating that Msl1 is a substrate of Msl2 *in vitro* (Figure 18, lanes 9-10).

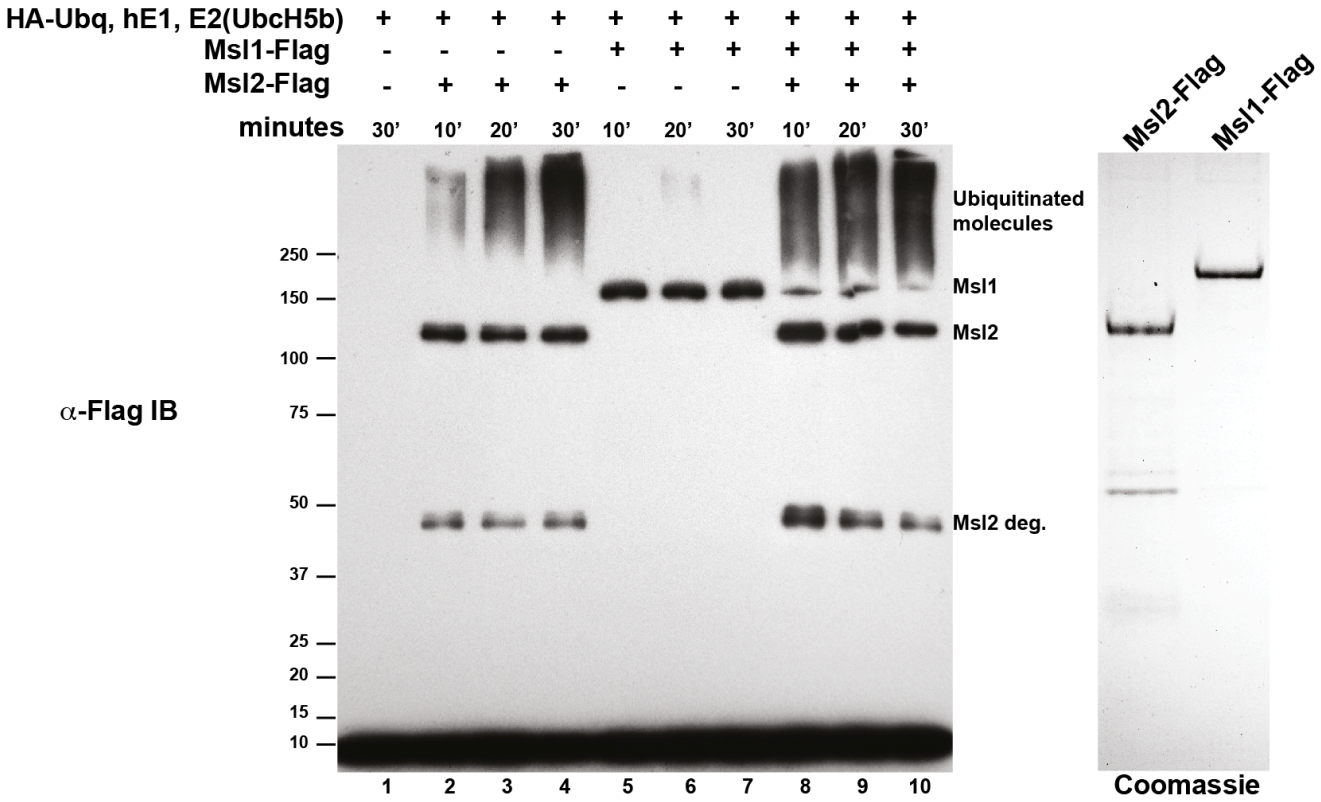
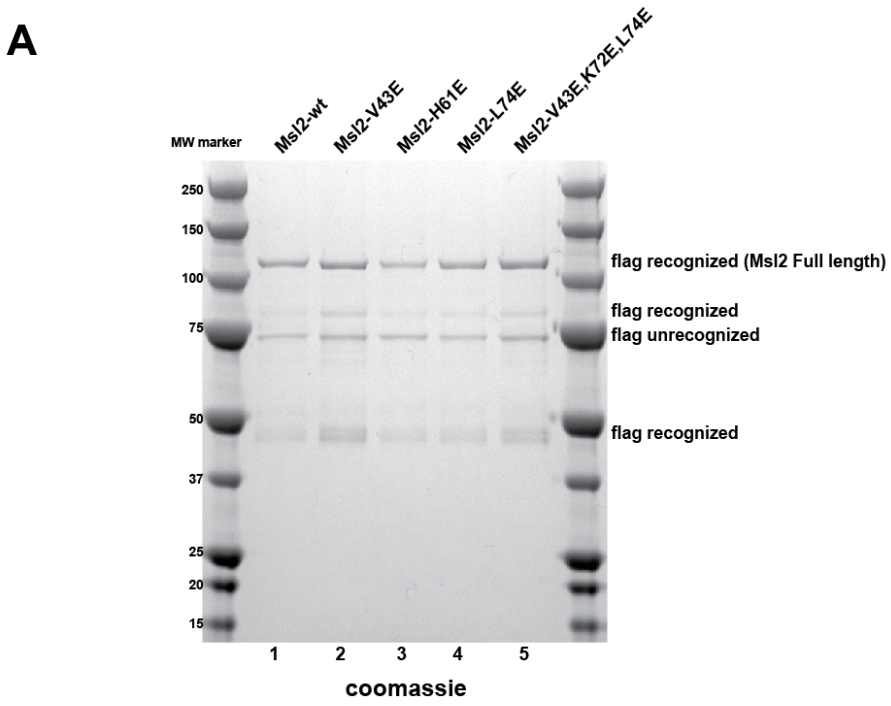


Figure 18. Msl2 is an E3 ligase

In vitro ubiquitination assay with recombinant Flag tagged Msl1 and Msl2. Equal amounts of proteins were assayed in 10 minutes time interval. Flag antibody was used to determine ubiquitinated pools of Msl1 and Msl2. Coomassie gel of purified Msl1 and Msl2 is shown in the right panel.

To understand the importance of the E3 ligase activity of Msl2, we reasoned that if we could disrupt the E2 binding without affecting Msl1 binding, we could measure the effects of loss of ubiquitination activity on the complex formation. In this regard, we prepared several mutations aimed to disrupt its interaction with E2 enzymes. As MSL2 RING does not possess a consensus E2 binding surface we mutated the *Drosophila* counterparts of Val46, Met75 and Met77 of the loop occluding the putative E2 binding surface (Val43, Lys72 and Met74). We first purified these mutants from Sf21 cells by baculovirus mediated expression (Figure 19A). We observed that the triple mutant V43E, K72E, L74E had a significantly reduced E3 activity whereas the single mutations did not (Figure 19B). These mutations, however, also affected the overall Msl2 structure, as this mutant no longer interacted with Msl1 upon transient expression in Kc cells, where no endogenous Msl2 exists (Figure 19C). So far, we were unable to identify a mutant that would uncouple Msl1 from E2 enzyme binding by Msl2, suggesting a tight interdependence between the two proteins.



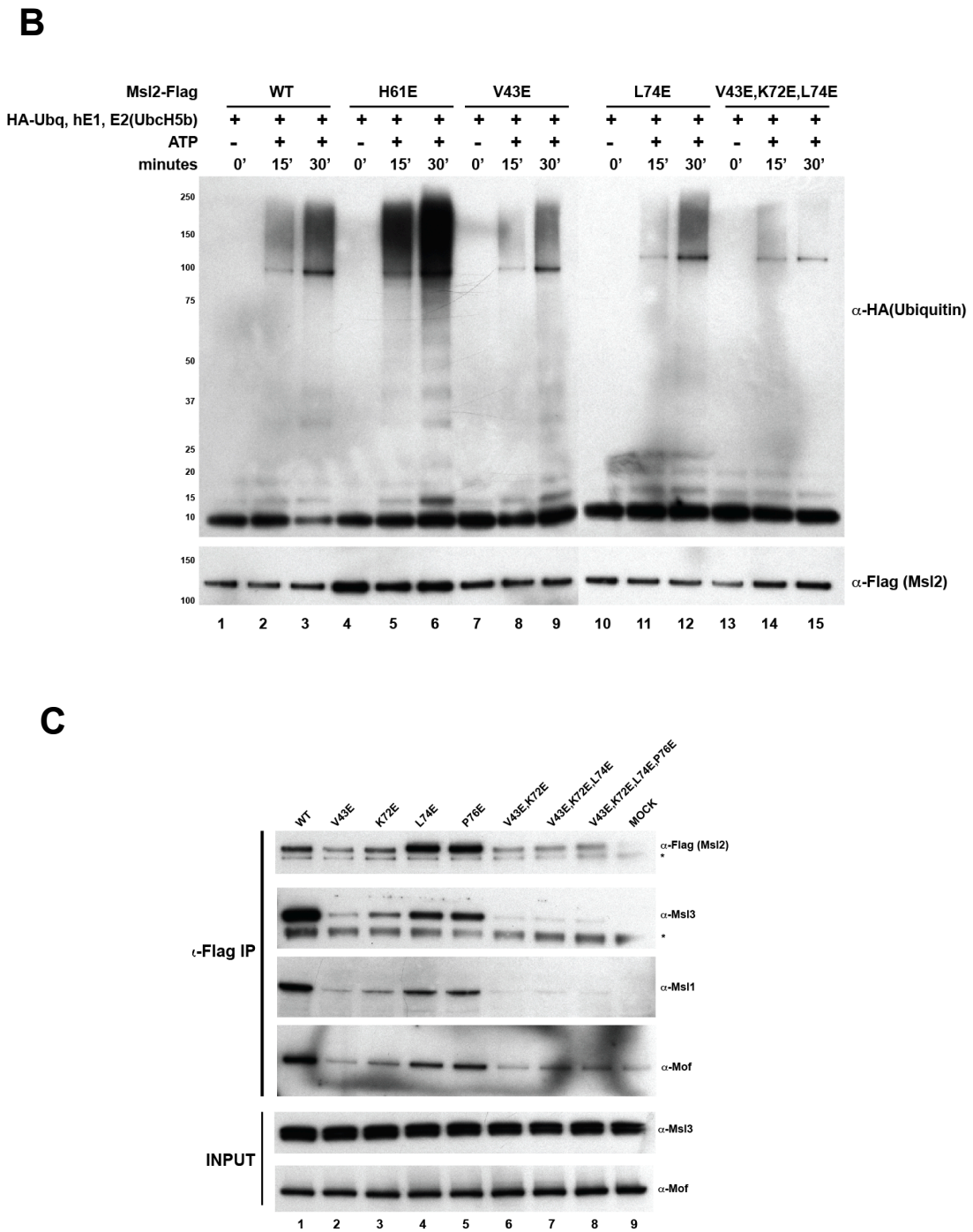


Figure 19. Msl2-E2 interaction surface disruption trials

(A) Purification of Msl2 mutants in Sf21 cells by Baculovirus mediated expression system. All mutants are purified by Flag agarose beads and eluted with Flag peptide. Coomasie gel is shown for Flag elutions. Major degradation bands are indicated according to their recognition by Flag antibody upon western blotting. **(B)** In vitro

ubiquitination assay of Msl2-Flag mutants. Assay is performed as in Figure 18. “0” time represents no-ATP control. HA antibody is used to assay HA-Ubiquitin. Flag blot shows the loading control for the amounts of purified Msl2 proteins. **(C)** Transient expression of Msl2 mutants in Kc cells. In order to avoid the effects of endogenous Msl2, Kc cells were transfected with indicated mutant constructs and immunoprecipitated with Flag agarose beads. Msl2 mutants differ in their levels of expression. Asterisks in the Flag and Msl3 blot indicate nonspecific bands.

---

roX2 RNA integration requires the full complex

The MSL complex is a ribonucleoprotein complex, containing two functionally redundant long non-coding RNAs, roX2 and/or roX1[49]. In the absence of both of the roX RNAs, the MSL complex binds to several sites along the X chromosome, some autosomal sites and chromocenter[50], indicating a role of roX RNAs in spreading of the complex from a relatively few sites along the X-chromosome. Although essentially all the proteins of the MSL complex have a potential of interacting with nucleic acids, the actual mode by which the complex binds RNA remains unknown[19,31,45,159]. We used the Msl1 mutants to study roX integration into the complex *in vivo* by RNA immunoprecipitation (RIP) method, where fixed complexes are pulled down and RNAs are quantitatively measured by quantitative PCR [158]. First, we optimized the RIP protocol in SL-2 cells, where roX2 but not roX1 is expressed, using the Mle subunit as a bait protein as its interaction with roX2 is well established [45]. Mle bound roX2 *in vivo*, as verified by two different primer pairs, and did not bind a non-specific nuclear RNA, 7SK (Figure 20A). RIP by Flag antibody gave only background levels of signal from wild type SL-2 cells (Figure 20A), ensuring the specificity of signals obtained from RIP of Msl1 mutants (Figure 20C). In order to capture partial complexes more efficiently, we generated stable SL-2 cell lines for all the Msl1 mutants except for Msl1 mut.11 because it behaves very similar to mut.12 (Figure 16). Since the level of expression can affect the recoveries of RIP, we optimized induction with Cu^{+2} concentrations to achieve similar levels of expression for each mutant (Figure 20B). Rox2 binding to the exogenous wt Flag-tagged Msl1 was recapitulated

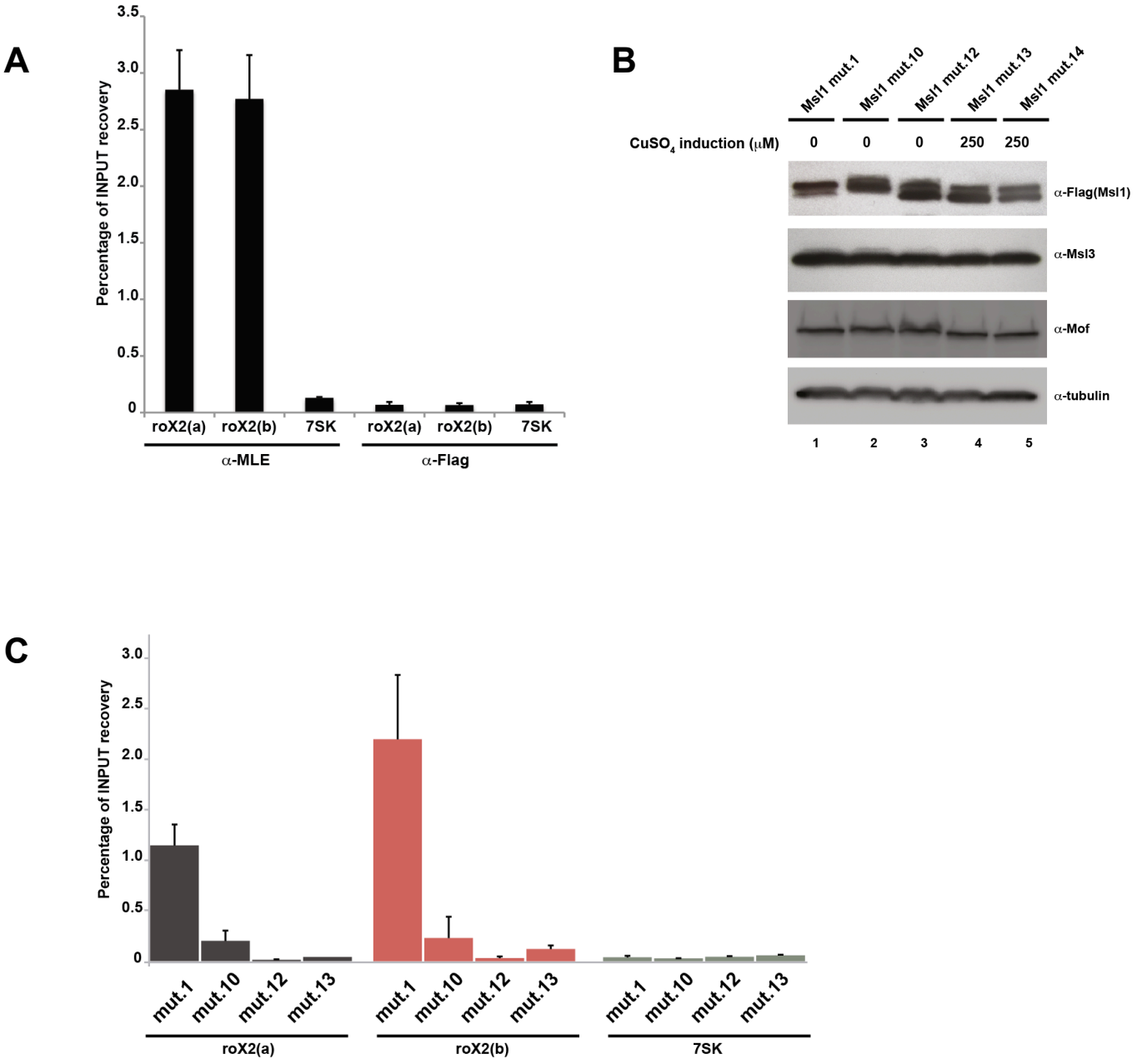


Figure 20. RoX2 interaction with partial MSL complexes

(A) RNA immunoprecipitation (RIP) in SL-2 cells with Mle and Flag antibody. RIP on Mle protein is used as a positive control for roX2 RNA binding. Two different roX2 sites are quantitatively amplified (roX2 a, roX2 b). 7SK is used as a nuclear RNA negative control. RIP with Flag antibody is repeated on same targets in wild type SL-2 cells to show background levels of RNA recovery. The error bars represent the standard deviation of 3 independent experiments. (B) Equal levels of Msl1 mutants used in ChIP and RIP experiments were achieved by differential induction of MtnA promoter with indicated

$CuSO_4$ amounts in SL-2 stable lines. MtnA promoter has a leaky expression without any induction. Each line contains Msl1 construct with 3xFlag tag in C terminus. Flag antibody was used to determine the levels of Msl1 proteins. (C) Flag RIP experiment in SL-2 stable cell lines that express Msl1 mutants. 2 rox2 RNA target sites and a negative control RNA target (7SK) are amplified.

---

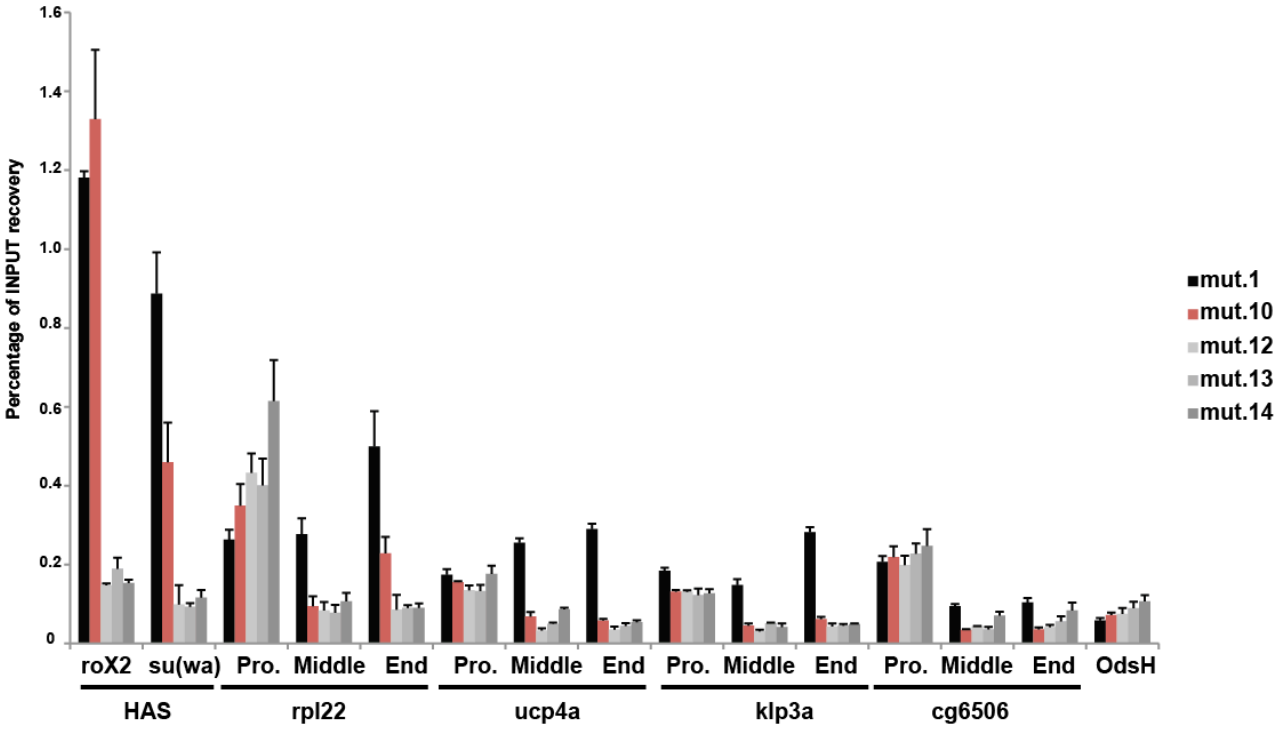
from the stable cell line. While Msl1 mut.10 showed a significant reduction of RNA recovery, mut.12 and mut.13 completely lost the binding despite their equivalent expression levels (Supplementary Figure 5A). The lack of roX2 signal in these assays is a direct indication of RNA loss of incorporation to the partial complexes, since roX2 RNAs are stabilized by the presence of endogenous proteins. These results indicate that Msl1-Msl3-Mof trimeric complex (Msl1 mut.12) and hexameric complex (Msl1 mut.13) cannot bind roX2 in the absence of Msl2. The Msl3 or Mof proteins are also required for complete incorporation of the RNA but their contribution is not detectable by this method when Msl2 is not present in the complex. Msl2 thus appears to be a key subunit for stable roX2 integration into the MSL complex.

Msl1 dimer platform and its association with Msl2 is required for X chromosome recognition

In order to understand the importance of the Msl1 dimerization in X chromosome targeting, we tested our mutants for their interactions with chromatin by ChIP analysis in stable SL-2 cell lines where each Msl1 mutant is under a Cu^{+2} inducible metallothionein A (MtnA) promoter. Expression levels were equilibrated as in figure 20B. We used the Flag epitope for IP to selectively pull down mutant derivatives, avoiding endogenous Msl1. First we analyzed Msl1 binding to two high-affinity site (HAS) targets (roX2 and su(wa)) and several low affinity sites within four X-linked genes (Figure 21A). Msl1 mut.10 ChIP shows that roX2 HAS binding is independent of both Msl3 and Mof and su(wa) showed a reduced binding of the partial complex whereas spreading across the body of the X-linked genes was completely lost. This result further supports our previous hypothesis that not all high affinity sites are identical and show differential affinities towards various surfaces of the MSL complex. Strikingly all the other mutants (Msl1 mut.12, 13, 14) did neither bind either to HAS nor

low affinity site gene bodies. Exceptions were observed for the promoter regions of the same genes where binding remains unaffected (see below). In order to ensure that X chromosome recognition is lost starting from the HAS, we tested 12 more HAS targets determined by Kuroda and colleagues [77] (Figure 21B). Remarkably, all the tested targets show reduced binding of Msl1 mut.10 and completely abolished binding of the Msl1 mut.12, 13 and 14. The abolished binding of Msl1 mut.13 importantly shows that Msl1 dimer *per se* cannot target the X chromosome but requires the composite action with Msl2. Taken together, these clearly indicate that Msl1 dimerization mediated Msl2 binding, is necessary for the recognition X chromosomal genes.

A



B

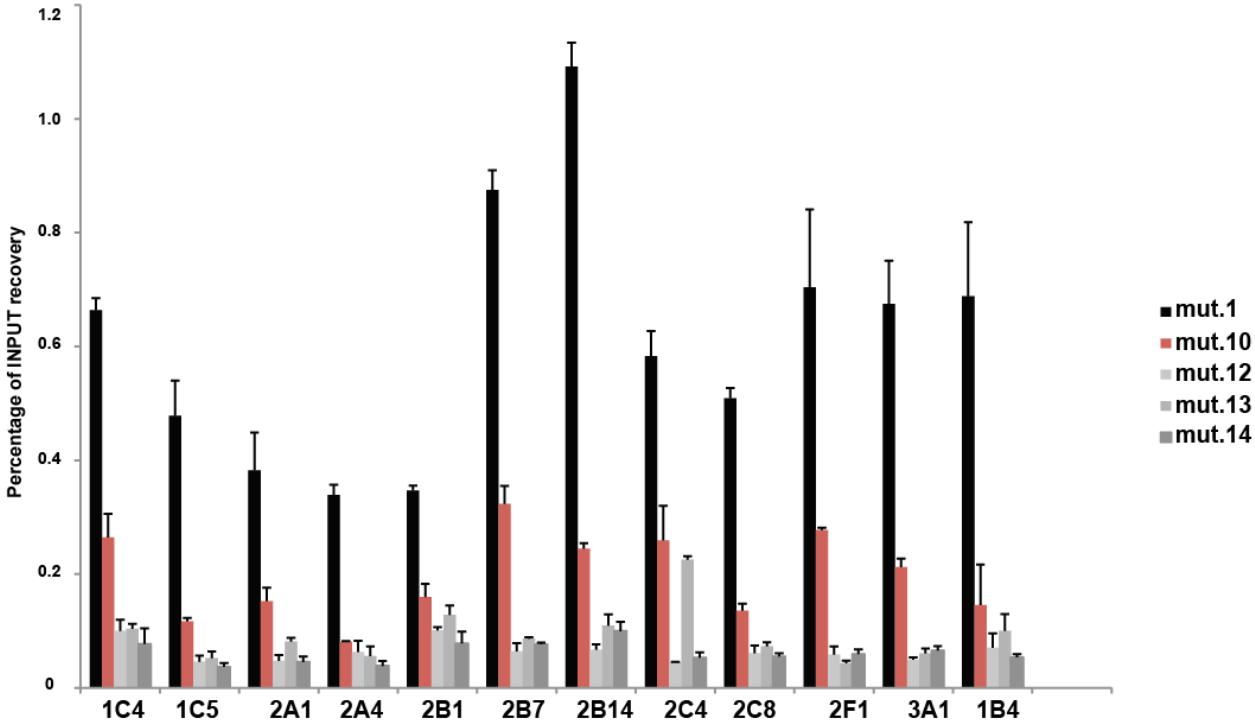


Figure 21.Chromatin interactions of partial MSL complexes

(A) ChIP of Msl1 mutants with Flag antibody in SL-2 stable cell lines. Transcription of Msl1 mutants is induced by different Cu^{+2} concentrations for similar level of expression (See Supplementary Figure 5A). 2 High Affinity sites and 4 X-linked genes were chosen as X-chromosomal targets. OdsH target is used as a negative control. P, M and E indicate promoter, middle and end of the genes respectively. The error bars represent the standard deviation of 3 independent experiments. **(B)** Same experiment as in (A) is performed on selected twelve High Affinity Sites.

Msl1 binds to promoters in male and female cells

Reproducible Msl1 binding to the promoters of X-chromosomal genes and its independent nature from Msl3, Mof, Msl2 and dimerization (Figure 13 and 21) prompted us to hypothesize that this binding could be independent from its role in

dosage compensation. In such a scenario, Msl1 might also be detectable at the promoters of autosomal genes, where dosage compensation does not occur. Indeed, by ChIP we detected strong enrichments at the promoters of eight random autosomal targets, while ORF binding was at the background level (Figure 22A, black bars). Given the tight interaction between Msl1 and Msl2 and its role in transcriptional regulation at the promoters in mammalian cells[18], we analyzed also the Msl2 binding to the promoters of autosomal and X-chromosomal genes. Interestingly, Msl2 follows similar binding patterns as Msl1, occupying promoters on the autosomal genes while on X-linked genes enrichment peaks towards the end of the genes in male cells (Figure 23A). We were next interested in identifying the distinguishing factor between autosomal and X-linked genes. Since MOF is also present in autosomal promoters [73], we checked Msl3 systematically on the same autosomal and X-linked genes in males and females. Surprisingly, Msl3 was absent on autosomal promoters in male cells (Figure 23B) and on female promoters (data not shown) compared to X chromosomal targets. Msl3 occupancy at X-linked promoters was either absent or very low relative to Msl1 and Msl2. Taken together; these results suggest that Msl1 binding at the promoters is independent from its role in dosage compensation. The exclusive presence of Msl3 provides a distinguishing feature for the X-chromosomal genes versus autosomal targets with concomitant spreading of the MSL complex.

Msl1 dimerization is essential for male viability

In order to assess the functional relevance of these *msl1* mutations in *Drosophila in vivo,* we generated transgenic flies expressing the mutant variants of *msl1* (wt Flag-tagged, mut.10, 12, 13, 14) in a spatiotemporally regulated manner using the UAS/Gal4 binary system. All transgenes were inserted in the same genomic location (65B2) by phiC31 integrase-mediated transformation to avoid the influence of position effects on gene expression and facilitate direct comparison upon phenotypic analysis[160]. Furthermore, using the fly system also enabled us to directly compare sex-specific effects of different mutations.

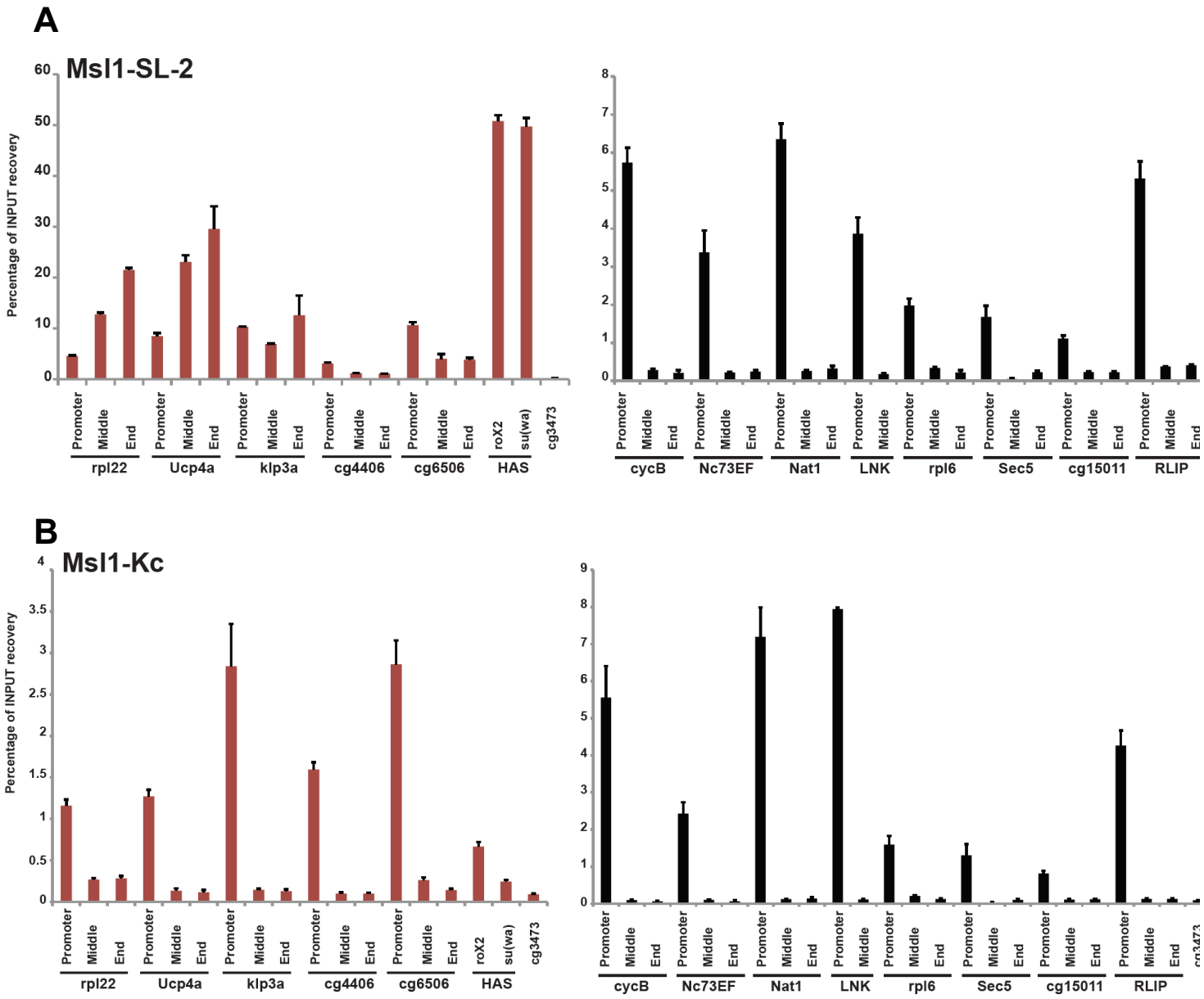


Figure 22. Msl1 binds to promoters of X and autosomes in a sex-independent manner.

(A) ChIP of endogenous Msl1 in SL-2 cells. 5 X-linked genes and 2 HAS are chosen for canonical X chromosome enrichment (Red bars). Cg3473 is a negative control target site. 8 autosomal target sites are shown with black bars. The error bars represent the standard deviation of 3 independent experiments. **(B)** Same experiment is performed for endogenous Msl1 in Kc cells

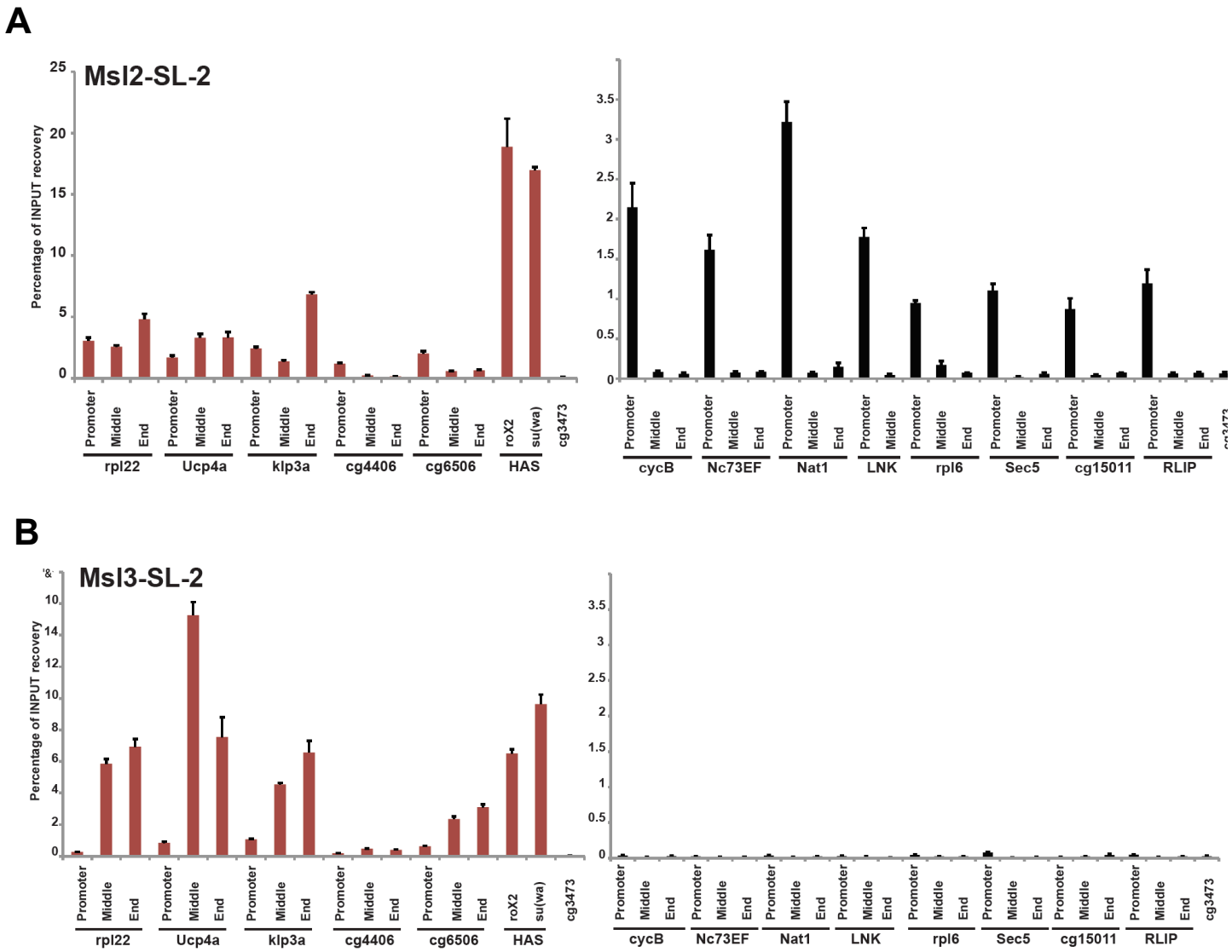


Figure 23. Msl2 binds to the promoters but not Msl3

Standard ChIP with Msl2 **(A)** and Msl3 **(B)** antibodies as performed in Figure 22 with same target sites.

We first induced ectopic expression of these mutants in a wild type background ubiquitously with a strong *tubulin-Gal4* driver at 25°C [161]. Strikingly, expression of Msl1 mut. 12, 13, 14 caused both male and female lethality, whereas Msl1 mut.10 caused only male-specific lethality and wild type Msl1 expression did not have any observable effects on viability (Figure 24B). Western blot analysis showed that Msl1 mut. 12,13 and

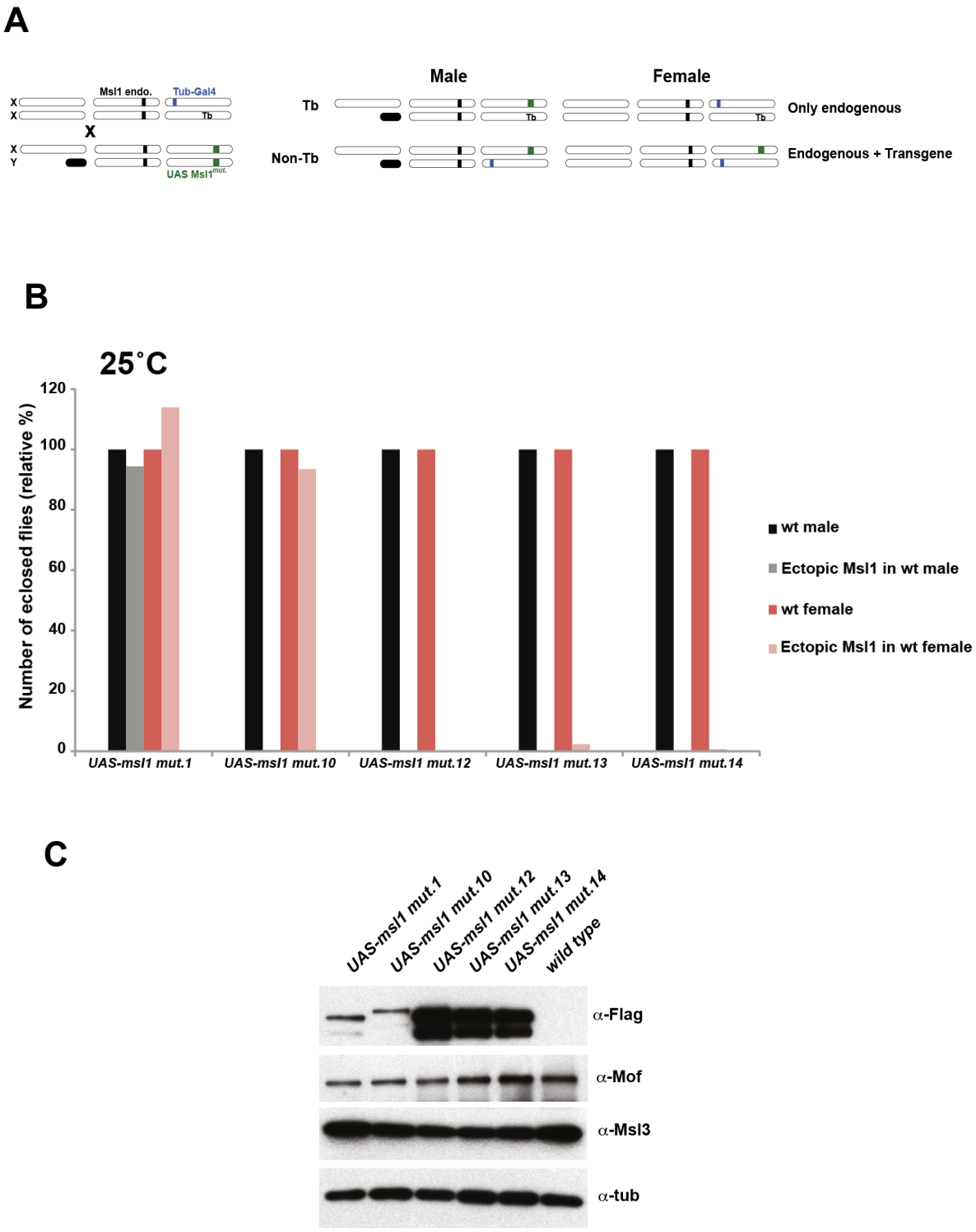


Figure 24. Overexpression of Msl1 mutants in wild type background at 25°C

(A) Schematic representation of the cross used in this study. Details can be found in Methods. **(B)** Ectopic expression of wild type Msl1 and Msl1 mut.10, 12, 13 and 14 in a wild type background at 25˚C. The non-expressing *TM6Tb/UAS-msl1** internal controls are designated in black (males) and red (females). The *tubGal4/UAS-msl1** males with

transgene expression are shown in grey and females in pink. Viability of adult flies for each genotype is represented as % from the total number of offspring for each *UAS-msl1** mutant set of crosses shown in methods. Total number of flies counted for this assay was n= 3489. **(C)** Western blots from protein extracts prepared from second instar larvae carrying different *UAS-msl1** transgenes, all C-terminal 3xFlag tagged. Flag antibody was used to probe exogenous Msl1; Mof and Msl3 protein levels are shown for comparison. Tubulin levels were used as a loading control.

14, which lose Msl2 interaction, are more abundant than wt Msl1 and mut.10 (Figure 24C), suggesting a down-regulating effect of Msl2 on Msl1, which is consistent with our cell culture observations and *in vitro* ubiquitination assays that Msl1 is an Msl2 ubiquitination substrate.

In order to ensure that lethality is not due to indirect effects of over-expression of the mutant proteins, especially for the coiled-coil Msl1 mut. 12, 13 and 14, we repeated the experiment at 18°C, where *tubulin-Gal4* induced transgene expression can be significantly decreased relative to 25°C[162] and (Figure 25A). Under these conditions, we observed that female viability is restored for Msl1 mut.12, 13 and partially for mut.14, whereas male-specific lethality was still observed for all mutants (albeit escapers for Msl1 mut.10 and mut. 13). Ectopic expression of wild type Msl1 in these conditions had no effects on viability. These results show that dominant negative effects of all mutations can be observed exclusively in males at both temperatures, whereas females become sensitive to the levels of Msl1 mut. 12, 13 and 14 at 25˚C.

To assess the direct effect of the mutations, we expressed the Msl1 mutant variants in $msl1^{L60}/msl1^{\gamma 269}$ null mutant flies to reconstitute Msl1 function. As expected, in the absence of Msl1, females are viable whereas males die as third instar larvae or at early pupal stages (Figure 26B). At 25°C, *tubulin-Gal4*-induced ectopic expression of wild type Msl1 rescued completely the *msl1* loss-of-function male-specific lethality (Figure 26B). Noticeably, none of the *msl1* mutants rescued male lethality (Figure 26B). Female viability dropped significantly in Msl1 mut.12, 13 and 14, similar to the dominant effect observed upon over-expression in a wild type background (Figure 26B). At 18°C, *tubulin-Gal4*-induced ectopic expression of wild type Msl1 rescued the *msl1* loss-of-

function male-specific lethality only partially and other mutants failed to do so (data not shown). These results clearly show that all of the residues that are determined from the crystal structure are absolutely essential for male viability.

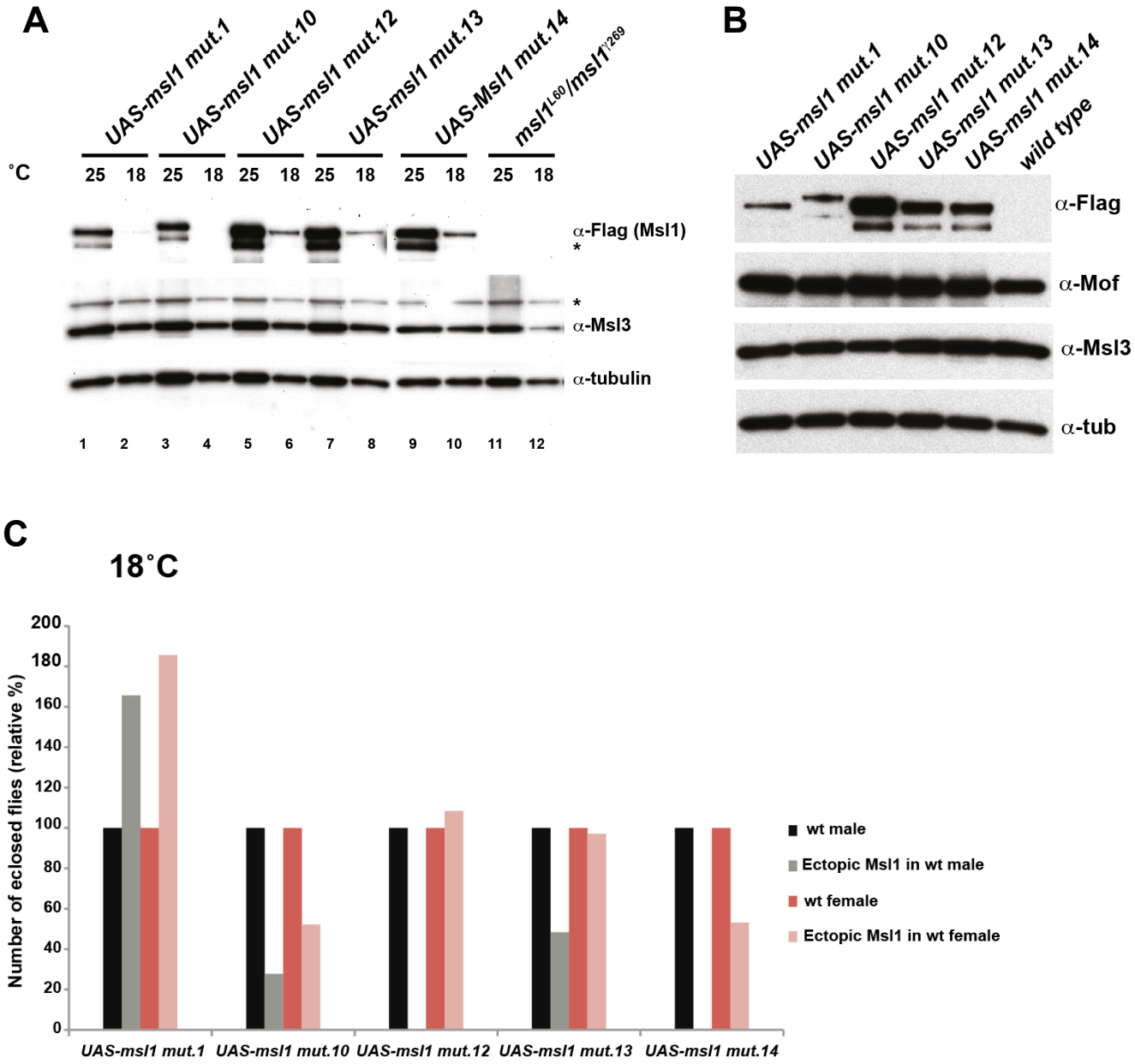


Figure 25. Overexpression of Msl1 mutants in wild type background at 18°C

(A) Temperature dependence of tub-Gal4 driven expression. Protein samples were prepared from larvae grown at 25˚C and 18˚C expressing UAS-msl1 genes with tub-Gal4 driver in msl1L60/ msl1γ269 null background. Tubulin expression was assayed as a loading control and endogenous Msl3 protein was shown for temperature independence of endogenous proteins. **(B)** Western blots were done as in (Figure 24C).

(C) Recovery of female but not male viability upon weaker/lower ectopic expression of Msl1 mut.12, 13 and 14 in a wild type background at 18˚C. Crosses were performed as in (Figure 24A**)** but the flies had been kept at 18˚C. Total number of flies counted for this assay was n= 2376.

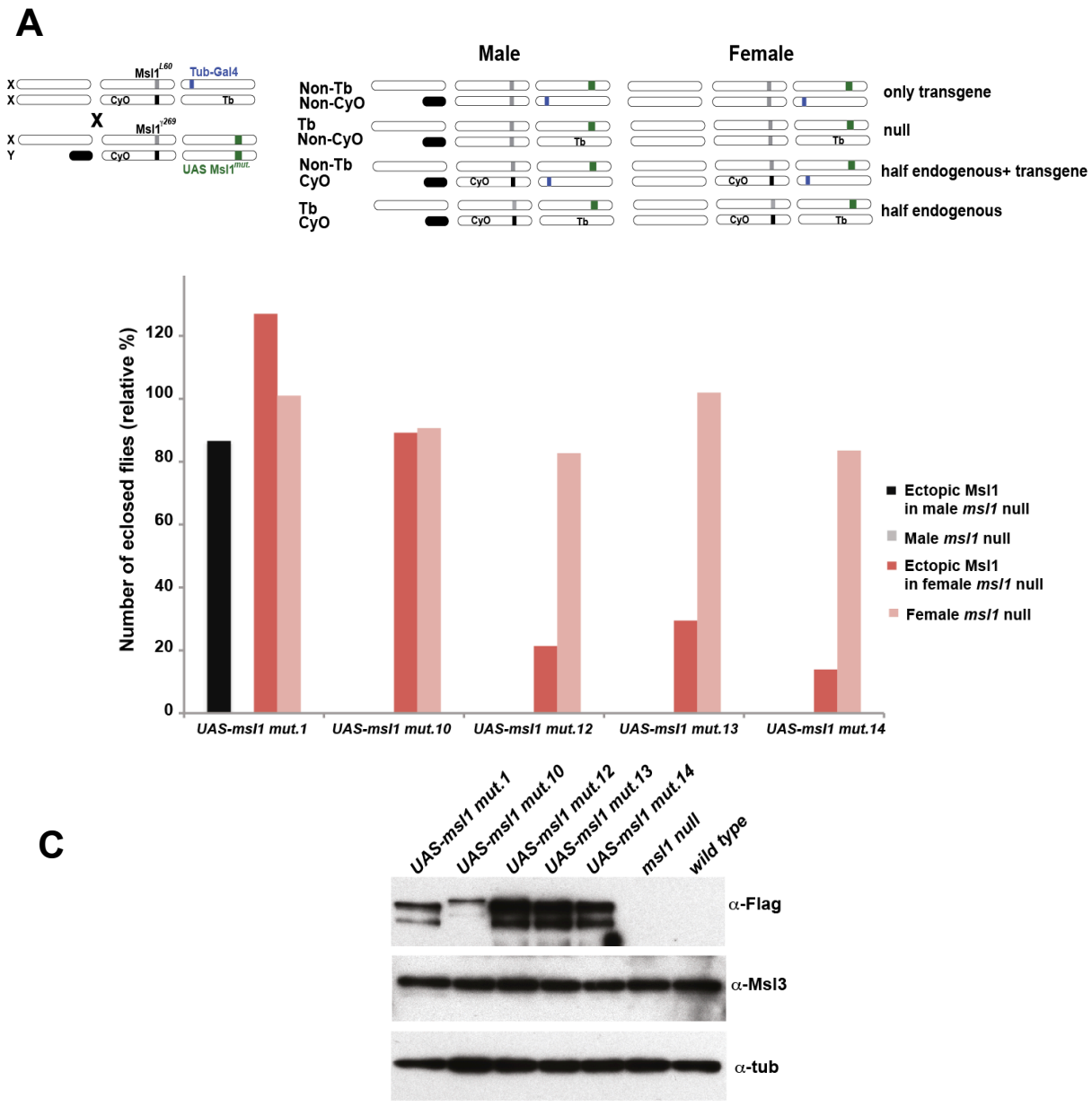


Figure 26. Rescue of the *msl1* loss-of-function male-specific lethality phenotype by ectopic expression of Msl1 mutants in *msl1*^{L60}/*msl1*^{,269} null mutant background at 25˚C

(A) The cross was performed as indicated in Methods and schematically respresented. **(B)** Transgene expressing *msl1^{L60}/msl1^{ν269}; tubGal4/UAS-msl1** males are represented with black and corresponding females with red bars. The non-expressing *msl1^{L60}/msl1^{ν269}; TM6Tb/UAS-msl1** internal controls are shown in grey for males and pink for females. Total number of flies counted was n=2428. **(C)** Tubulin antibody was used for loading control and Msl3 and Flag antibodies were probed to show the levels of expression of transgenic Msl1 and endogenous Msl3.

DISCUSSION

DISCUSSION

Dosage compensation is an excellent example of how cells can fine-tune a whole chromosome in an epigenetic manner. At the heart of the dosage problem lies the imbalance resulting from the structure of sex chromosomes, where one of the homolog bears the sex determining region and structurally inert, as in Y chromosome in males. Although recently the "universality" of dose problem have been questioned[163], it is clear that, at least in model organisms ranging from *C. elegans* to mammals, there are micro-molecular machines that structure a whole chromosome. In Drosophila melanogaster, this machine is the MSL complex or DCC complex (Dosage compensation complex) which tries to transcriptionally up-regulate the single X chromosome two fold in a surprisingly accurate manner. It is difficult to comprehend how more than 2200 genes, all having unique promoter-distal elements evolution, can be elevated with a single machine. Therefore it is not unexpected to witness the complexity and the diversity of the elements MSL complex brings together. It has at least three enzymes, Msl2 and Mof having the potential to post-translationally modify histones, Mle being an RNA helicase/chaperone. It contains two functionally redundant long non-coding RNAs, which on the sequence level have minor similarity. It harbors a histone marker reader (Msl3) that can recognize methylation of histones [1]. Furthermore, the X chromosome in males seems to be positionally restrained to nuclear periphery[138]. Nevertheless, the complexity could be broken down to three different categories;

a) Assembly problem: How do the male cells promote the complex formation whereas females avoid it? Where in the cell does the complex form and how it chooses the integrate either of the non-coding RNAs? Is the chromatin needed to structurally support the complex formation? When in the developmental stage the complex starts to function and how is this coordinated with developmental clues?

b) Targeting problem: How does the complex recognize the X chromosome? What are the genetic and epigenetic marks that confine the complex to one single chromosome? What are the features of autosomal MSL binding sites?

c) Function problem: How does the MSL complex regulate the transcriptional up-regulation? How is actually two fold calculated?

Although these problems seem to ask different questions, their answers are usually entangled and give clues to one another. In this thesis, we tried to obtain the structure of the MSL complex with the hope that we could contribute to the above-mentioned problems. As expected, we gained considerable insight from the structural data.

Advantages of having point mutations

To understand the contribution of a single compartment of a system, in this case, a protein inside a complex, it is a general practice to knock down or knock out the gene of interest. However growing number of evidence show that proteins are not usually part of single complexes with single functions, but rather have multiple functions that may reside in more than one complex. A relevant example is MOF acetyltransferase. Mof is part of at least two complexes, NSL and MSL[83]. NSL is a major regulator of Drosophila genome in both sexes[84]. Furthermore mammalian MOF is indicated in DNA damage pathway acetylating p53[164], acetylation of TIP5, a part of the NORC complex responsible for regulation of rRNA genes[165]. Therefore, RNAi against Mof will not yield a dosage complex phenotype but rather a complex one. Our point mutation analysis on Msl1 scaffold yields surgical removal of proteins, creating genuine partial complexes and enabling us to monitor the absence of individual proteins in the context of dosage compensation. By using this method we gained important insights into targeting and spreading of the complex.

Dimerization of Msl1 enables spreading of the complex along gene bodies

The fact that Msl1 dimerizes through such an extended interface and the dimer formation is required for Msl2 binding dramatically changes our view on the dosage compensation complex structure and assembly. As it is possible to co-purify the recombinant human MSL1/MSL2/MSL3/MOF complex from insect cells using Flag-tagged MSL2 that can presumably only bind dimeric MSL1, it is very likely that the MSL complex contains all the subunits in pairs, including also MSL3 and MOF [18] (Figure 26). We therefore propose that MSL complex binding to the open reading frames of the X-linked genes in *Drosophila* happens through a dimer dependent nucleosome engagement. The presence of two copies of each of the chromatin modifying or modification binding domains of the complex would increase the number of possible, probably transient contacts with nucleosome(s), containing histones also in pairs. The

Drosophila Msl1 scaffold is a large, mostly disordered protein (1039 residues) that provides the MSL complex with high degree of flexibility. Msl2 and Mof/Msl3 binding regions of Msl1 are separated by 720 poorly conserved probably unstructured residues. It is thus possible that while some subunits are

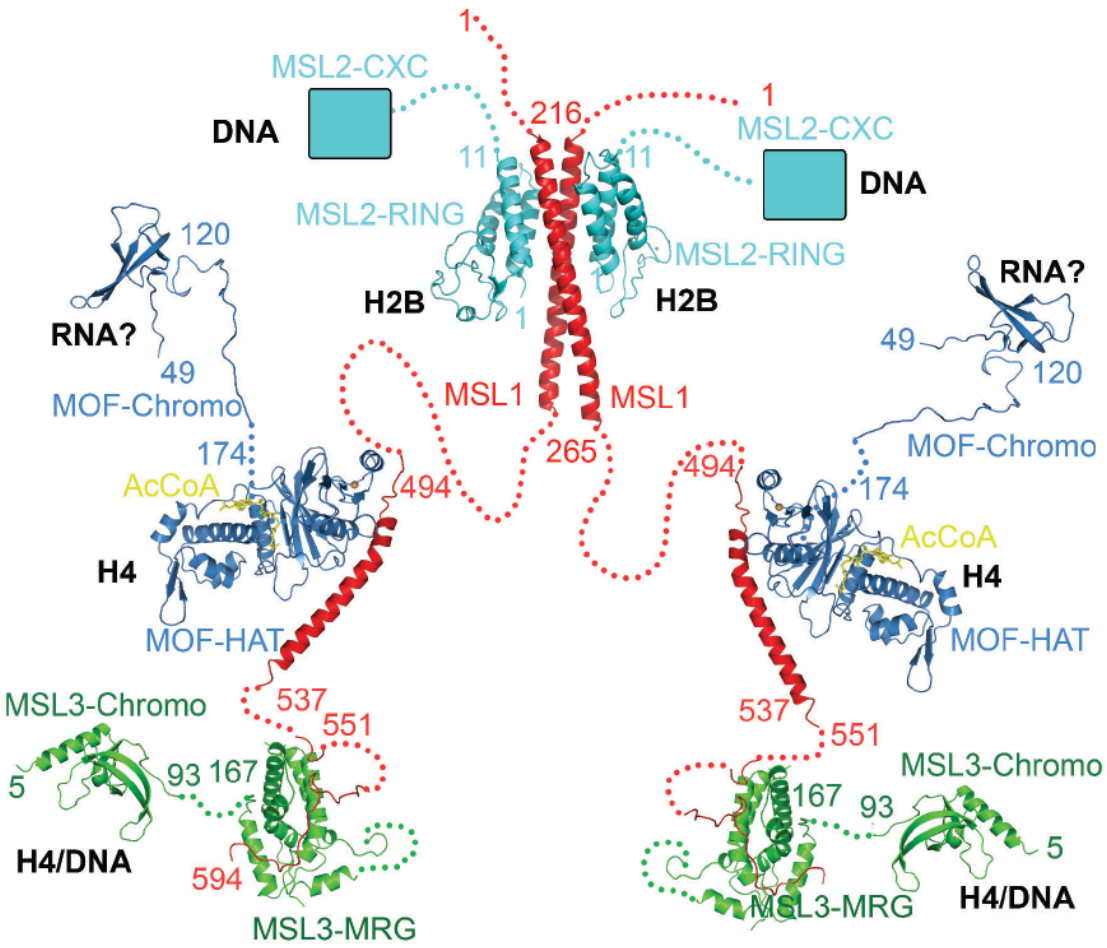


Figure 26. Summary of the structural information on the human MSL complex.

A schematic model showing all existing structures of the MSL sub-complexes and domains including the MSL1/MSL2 tetramer structure, the structures of MSL1 in complex with the MRG domain of MSL3 and the HAT domain of MOF (PDB codes 2Y0N, 2Y0M) the chromo barrel domain of MSL3 (PDB code 3OB9, [42]) and chromo barrel domain of MOF (PDB code 1WGS). These structures, together with sequence alignments and secondary structure and disorder predictions suggest that the MSL complex is characterized by ordered functional domains separated by extensive natively disordered and flexible regions. In particular MSL1 contains short helical (or unstructured) interaction peptides that interact with MSL2, MSL3 and MOF separated

by a very long, disordered region (shown as red dotted line). Similarly the chromo barrel domains are separated from the MRG and HAT domains by rather long, probably unstructured linkers (green and blue dotted line). The only defined domain of unknown structure remains the CXC DNA binding domain of MSL2, which is again separated by long linker from the N-terminal RING finger (dotted line in cyan). As it is currently unknown whether any additional interactions exist among the individual proteins and their domains, this model should be regarded only as schematic and is not intended to imply any particular three-dimensional arrangement of the MSL complex. Putative nucleosome and RNA binding sites are labeled.

---

attached to chromatin, others, connected by the flexible Msl1 linker can browse the surrounding nucleosomes for new attracting histone marks. It is well established that MSL complex can spread in *cis* from a HAS translocated to autosomes depending on the transcription status of the nearby genes [35]. Since transcription machinery will generate a linear gradient of active marks along the genes, which will be also linearly erased, MSL complex 'walking' could be unidirectional along the gene. The dimer dependent spreading can also be deduced from the ChIP analysis of Msl3/Mof deficient Msl1 mutant in the endogenous Msl1 background (Figure 21). Msl1 mut.10 can still dimerize with the endogenous intact Msl1, albeit low levels observed from our IP analysis (Figure 16), however it cannot spread to the open reading frames, which indicates that both copies of Msl3 and Mof are required spreading.

Msl1 and X chromosome recognition

MSL complex has to differentiate X chromosome from autosomal ones. One of the key concepts in this manner is the High Affinity sites (HAS) already suggested in 1994[85].High affinity sites were suggested to be initial recruitment sites to the X chromosome but only years later their sequence identity is discovered through high throughput sequencing[77]. Nevertheless, it is not clear what exactly these 'landing sites' are and if there are other clues than the mere DNA sequence. In our study we were able to distinguish HAS qualitatively. Our work strongly suggests that Msl1 *per se* cannot recognize other X-chromosomal features but promoters. However, binding of Msl2 to the Msl1 dimer has two important consequences: rudimentary recognition of X

chromosome and roX2 RNA integration into the complex. Msl1 mut.10 ChIP experiment shows that roX2 HAS only requires Msl1-Msl2 while all other tested HAS show reduced level of the complex (Figure 13 and 21). We propose that chromatin regions like roX2 HAS are the "**elementary high affinity sites**" where initial enrichment of Msl1 on X chromosome was mediated by a composite surface of Msl1-Msl2. We currently do not know how many such elementary sites exist on the X chromosome. High throughput analysis with different Msl1 mutants is required to understand the qualitative differences between these sites and other high affinity sites. Furthermore Msl1 mut.2 and mut.8 CHIP showed that **non-elemenatary HAS** requires Mof and Msl3 for their optimal binding (Figure 13). Grouping these sites according to their location in a transcription unit revealed that non-elementary HAS behave very similar to typical X-linked genes, meaning that their ChIP recoveries drop significantly in the absence of Msl3 and Mof however unaffected if they are at the promoters.

This can easily be explained with the hypothesis that ORF spreading is merely mediated by nucleosome engagement, most probably Mof/Msl3 head of the machine but the promoter bindings are independent of it (Figure 27).

MSL complex and roX2 RNA

Our RIP results clearly show that Msl1-Mof-Msl3 trimer or hexamer cannot bind roX RNA (corresponding to Msl1 mut. 12 and 13), which indicates an active role of Msl2 in this binding. Interestingly Msl1 mut.1 also shows significant loss of roX RNA interaction implying that full integration happens only in the context of the whole complex. A recent technology, named CHIRP, has shown that roX2 RNA exclusively follows the binding profile of Msl3 [166]. It is tempting to speculate that roX2 RNA is present only in ORF linked MSL complex and enables crosstalk between the two distant N-terminal Msl1-Msl2 and C terminal Msl1-Msl3-Mof catalytic centers of the MSL complex.

Although Mle is a major player in roX RNA shaping and we could not monitor its status with respect to our mutants. However Mle targeting to chromatin is RNAse sensitive in contrast to Msl1 and Msl2 [44]. Mle can not be immunoprecipitated with Msl1, Msl2 or Msl3 in SL-2 nuclear extracts [16]. Richter et.al also observed that Mle is not associated with Msl2 although all other members of the MSL complex were co-immunoprecipited. In fact Mle can only be IP'ed with MSL components under low ionic strength and RNAse

friendly conditions and upon 0.4 M NaCl titration, the interaction is diminished [26]. In our experiments we also failed to see Mle being co-IP'ed with Msl1 under 250 mM salt concentrations. All these experiments strongly suggest that Mle bears no protein contact with the rest of the complex but only is bridged to it by RNA, presumably roX RNAs. Therefore RIP experiments with Msl1 mutants are still relevant because we measure roX2 RNA directly.

Sex independent binding of Msl1 at promoters

Our study reveals that Msl1 binds Msl2 at promoters in male cells. In mammals, MSL2 ubiquitinates H2BK34 triggering H3K4 methylation and ultimately expression of the tested genes[18]. Since this crosstalk mainly happens at the promoters, we believe that Msl2 in *Drosophila* might function similarly. Indeed, *in vitro*, *Drosophila* Msl2 is capable of ubiquitinating Hela nucleosomes[18]. Interestingly, Kuroda and colleagues also observed an increase in RNAPII occupancy at autosomal promoters after Msl2 RNAi, arguing for a role of Msl2 on autosomes as well [167]. Furthermore, Msl2 can also ubiquitinate other members of the complex, which we have shown at least for Msl1 *in vitro*, and may have a regulatory role on the assembly of the complex. The stabilization of Msl1 mutants that cannot bind Msl2 both in cell culture and *in vivo* models indicates a role of Msl2 in controlling the amount of Msl1.

The occurrence of Msl1 at the promoters in both sexes and its independence from other members of the complex for this binding suggests the possibility of an evolutionary conserved function in higher eukaryotes. All complex members, except for Msl2 and Msl1, have origins traceable to yeast [17]. The emergence of "Msl1 like genes", namely Msl1 and Nsl1 in *Drosophila*, both having a PEHE region to bind Mof through the same surface (Figure 8), seemed to focus this ubiquitous acetyl-transferase to promoter regions of a large portion of *Drosophila* genome. Indeed Mof binds to promoters in both sexes and is responsible for the promoter chromatin H4K16 acetylation[73]. It was also observed that RNAi of Nsl1 or Msl1 does not completely diminish Mof occupancy at the promoter, probably because both proteins have complementary roles [84,84]. It will be interesting to delineate possible functional interplay of Msl1 and Nsl1 at promotes as well as the distribution of Mof between these two proteins. It is important to note that Msl1 is not essential for female viability, possibly due to this redundancy between Msl1 and Nsl1 in terms of Mof recruitment to the promoters. Female viability decreases only

when Msl1 mutants that have intact PEHE region are expressed, probably causing mistargeting of Mof or diluting it away from promoters. In addition, no affect is observed upon expression of wild type Msl1 or mut.10, strengthening the hypothesis that observed female phenotypes are due to Mof rather than Msl1.

The distinguishing factor between the promoter complex and the dosage compensation complex could be Msl3, whose binding on the autosomal promoters was undetectable and X-linked promoters was very low. It will be crucial to understand how Msl3 is excluded from the promoters and which part of Msl1 is responsible for promoter binding (Figure 27).

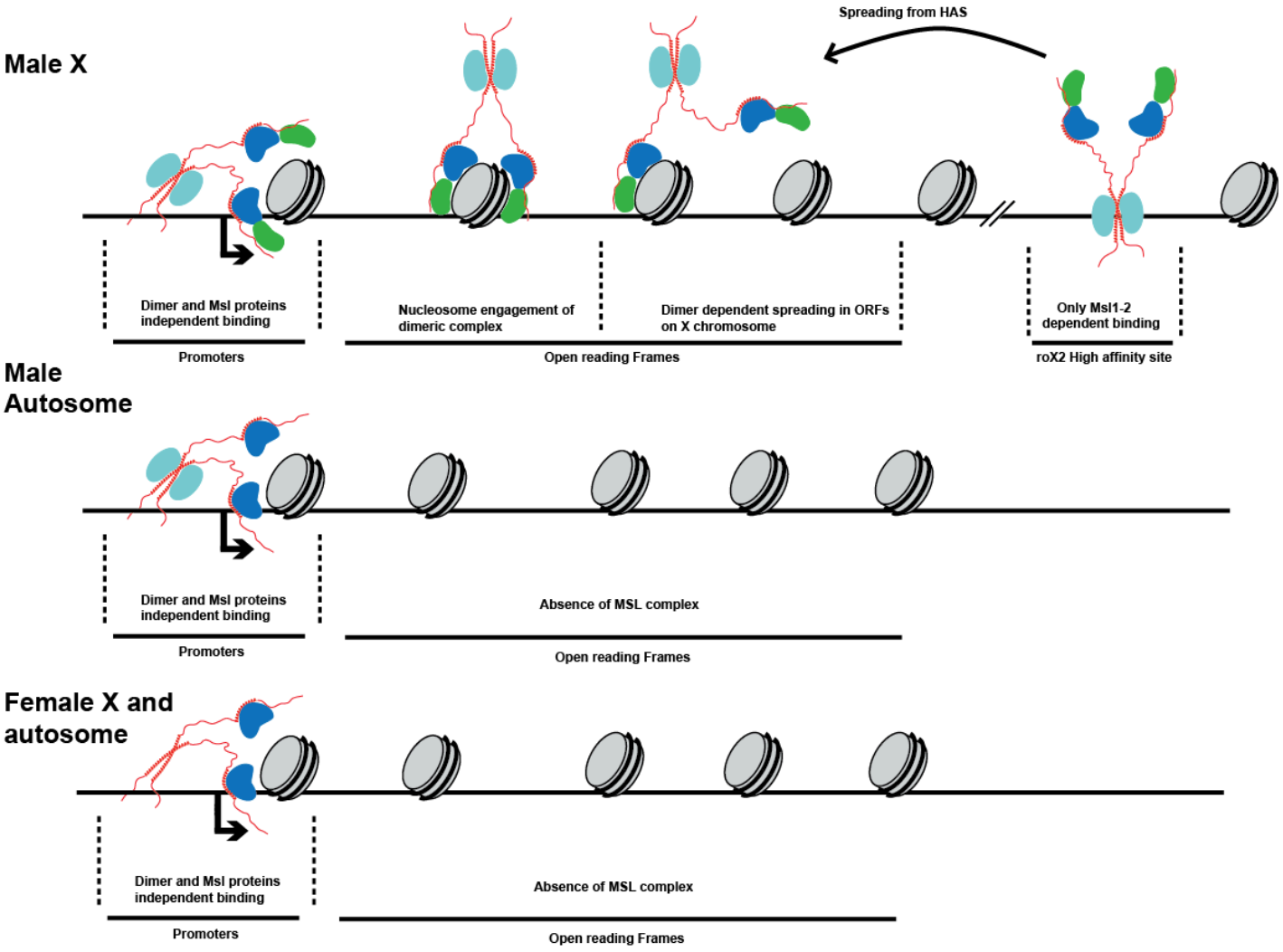


Figure 27. Summary of the MSL complex chromatin interactions derived from this work.

The MSL complex binds to roX2 high affinity site on male X chromosome in an Msl3-Mof independent manner. RoX2 HAS may represent one of the ''elementary high affinity sites''. The MSL complex must be in the dimeric state for binding to open reading frames on X chromosome. Long, flexible Msl1 linkers between the coiled coil and PEHE regions might enable the complex to interact with chromatin in various configurations. Msl1 binds to male and female promoters in a dimerization/Msl proteins-independent manner. In males and females, autosomal promoters are depleted of Msl3, however there is a slight enrichment on X-linked promoters in males. We cannot rule out the existence of another protein that binds Msl1 in a similar manner to Msl2 on promoters in females. Color coding of proteins is similar to Figure 17 and 26.

In summary, our study enhances our perspective on the architecture of MSL complex and how this configuration could help spreading of MSL complex on X chromosome. We show that Msl1 plays a key structural role in the assembly and function of the MSL complex in *Drosophila*, where it acts as scaffold protein. Msl1 dimerization is required for Msl2 binding and roX2 integration. Loss of either of these interactions has a severe consequence on X chromosome binding, which causes male lethality. We believe that a dimeric perspective of the MSL complex has a better explanatory power over the monomeric model. We showed the differences between high affinity sites and how they may form a gradient of affinities to help MSL spread more efficiently on the X chromosome. Our data also demonstrate for the first time dosage compensation complex independent function of Msl1 on male and female promoters.

MATERIALS AND METHODS

MATERIALS AND METHODS

Cloning of Msl1 mut.1-8 and Nsl1 mutants for expression in *Drosophila* SL-2 cells

Drosophila Msl1-Flag cDNA was initially cloned into a pFASTBac 6His vector (a vector generated from pFASTBac™1 -Invitrogen- by introduction of a 6His tag for the expression of N-terminally tagged proteins). The various mutants of Msl1 were generated in this vector by mutagenesis. For the expression in SL-2 cells, sub-cloning was carried out through NotI-XbaI restriction sites into the pAc5.1/V5-His A. For stable cell line generation, we sub-cloned all our constructs under Cu^{+2} inducible MtnA promoter in pIBU1. Nsl1 was initially cloned into SL-2 expression vector pBSactTAP, which is designed to express N-terminally TAP tagged proteins. An oligo containing a Flag/HA tag was inserted downstream of N-TAP. Mutagenesis was carried out in this vector.

Generation of 3xFlag tagged piBU constructs

pIBU vectors contain .Neomycin Resistance cassette expressed by Act5 promoter and terminated by SV40 teminator. PIBU1 series contain MtnA promoter for Cupper inducible gene expression and pIBU2 series have Act5c promoters for high levels of expression and generally used for transient transfections. 3xFLAG and 6 His tag (DYKDHDDYKDHDDYKDDDDKHHHHHH) sequence was inserted into piBU2 as a C terminal tag. Subcloning between pIBU1 and 2 was done by PacI-AscI rare cutters, which can swap the whole ORF in the multiple cloning site.

Transient transfections

Msl1 constructs for transient transfections are under short Actin 5c promoter and have 3xFlag-6his C terminal tags. For dimerization experiments, 2 μg of each vector were co-transfected to 25 million SL-2 and Kc cells using Effectene (Qiagen) reagent according to manufacturer's suggestions. After 48 hours cells were harvested. For Msl2 expression in Kc cells, Msl2 constructs were cloned under MtnA promoter, together with C terminal 3xFlag-6his epitope. 15 million cells were transfected in 10 cm dish and after 24 hours, induced with 250 μM $CuSO_4$. After 24 hours of induction, cells were harvested.

Generation of Stable cells lines

SL-2 cells were seeded on 6 well plates for 80% confluency, corresponding to 1-2 million cells. The cells were transfected with 0.5 μg of DNA by Qiagen Effectene Transfection Reagent. After two days, the cells were diluted to 1:5 and 1:10. Twelve hours later, the medium was changed with the one that contains 1 mg/ml Geneticin. The cells were selected at least two weeks, until the WT cells were completely eliminated. For the subsequent amplification of the cells, 0.5 mg/ml Geneticin concentration was used. All stable lines were controlled by MtnB promoter to prevent the over-expression of MSL1 and its mutants during the selection period and for the induction, 0.5 μM $CuSO_4$ was used for 12 hours.

Flag immunoprecipitations

Harvested cells were washed with cold PBS two times and resuspended in 1 ml HEMGT 150 buffer (25 mM Hepes/KOH 7.6, 0.1 mM EDTA, 12.5 mM $MgCl_2$, 10% Glycerol, 0.2% Tween-20, 150 mM KCl). After 3 times freeze thaw cycle in liq. N_2 and 37°C waterbath, extract was centrifuged for 30 minutes in 20000g. 30 μl bed volume of M2-Flag agarose beads (Sigma) were incubated for 3 hours at 4°C. The beads were washed for 5 times in HEMGT 250 and boiled with 40 μl of 4X Laemmli Buffer.

Fractionation of SL-2 cells

Wild-type or transiently transfected cells were harvested and washed once with PBS. After determination of the volume of the cell pellet, the cells were re-suspended in 5 PCV (pellet cell volume) of hypotonic buffer (10 mM HEPES/KOH 7.9, 1,5 mM $MgCl_2$, 10 mM KCl, 1 mM PMSF and 1X Protease Inhibitor cocktail solution) for 15 minutes at 4°C. After incubation, the swollen cells were vortexed for 30 seconds in 1% NP-40. The supernatant was kept as cytoplasmic extract. The nuclei were pelleted with 2000g for 5 minutes and washed with an isotonic solution (10 mM HEPES/KOH 7.6, 2 mM $MgCl_2$, 3 mM $CaCl_2$, 300 mM Sucrose, 1 mM PMSF and 1X Protease Inhibitor cocktail solution). The washed nuclei were re-suspended in 5 PCV of extraction buffer with 150 mM NaCl (20 mM HEPES/KOH 7.4, 10% glycerol, 150-420 mM NaCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 1 mM DTT, 1X Protease Inhibitor cocktail solution) and rocked for 2 hours in the cold room. The extracted nuclei were centrifuged at 22000g for 15 minutes. The supernatant was kept as the nucleoplasmic extract. The pellet was re-suspended in 5

PCV of 150 mM salt extraction buffer and solubilized by Benzonase. Equal amounts of fractions were analyzed by western blot.

Immunofluorescence for SL-2 cells

1 million cells were harvested and washed with PBS. The cells were swollen in 500 µl 0.5% sodium citrate for 7 minutes and loaded through a single chamber cytospin tunnel. The cells were spun for 10 minutes at 900 rpm. The slides were put in fixative solution (4% formaldehyde in PBS, 0.1% Triton X-100) for 8 minutes in a Coplin jar. After washing three times with PBS-T, the cells were blocked with 1% milk containing PBS-T. Anti-Flag mouse and anti-MOF rabbit were used in 1:100 dilutions in 1% milk containing PBS-T overnight at 4°C in a wet chamber. Secondary antibodies mouse-488 and rabbit TritC were used in 1:200 dilutions at 37°C. After DAPI staining and mounting, the cells were mounted and watched at 63X objective.

ChIP protocol from SL-2 cells

SL-2 cells were harvested and washed twice with PBS. The cells were resuspended in Fixation buffer (50 mM Hepes/NaOH pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl) and fixed with 1.8% Formaldehyde (final concentration). The reaction was quenched with 2.5 M Glycine and washed with the following buffers for 3 times; Paro Rinse 1 (10mM Tris pH 8.0, 10mM EDTA, 0.5mM EGTA, 0.25% TritonX-100), Paro Rinse 2 (10mM Tris pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA), and RIPA buffer (140mM NaCl, 25mM HEPES pH 7.5, 1mM EDTA, 1% TritonX-100, 0.1% SDS, 0.1% DOC, supplemented with Protease Inhibitors). The cells were sonicated in RIPA buffer with Branson sonicator (Power output 3%, Pulse duration 40, 30 cycles in total and each cycle 20 seconds pulse and 50 seconds off). The fragmentation of the DNA was checked with agarose gels. The sonicated sample was centrifuged at 14 krpm for 30 minutes and the chromatin supernatant was cleared with blocked protein A-Sepharose beads. The beads had been blocked with 800 µl RIPA, 100 µl salmone sperm DNA (10mg/ml) and 100µl (100X NEB BSA) for 1hr at 4°C. For each immunoprecipitation, 30 µg DNA containing chromatin was incubated overnight with 3 µl antibody and collected with 20 µl (bed volume) of blocked Protein A- Sepharose Beads. For the Flag tag containing samples, the beads were bridged with rabbit anti-mouse polyclonal antibody (from Active Motif). The beads were washed four times with RIPA buffer, one time with LiCl

buffer (250mM LiCl, 10mM Tris-HCl pH 8.0, 1mM EDTA, 0.5% NP-40, 0.5% DOC) and one time with TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA). After resuspension in 100 ul of TE buffer, the precipitated chromatin was decrosslinked overnight at 65°C. The sample was treated with 1μl of 10mg/ml RNAse A at 37°C for 30 minutes and 1.3μl of 10mg/ml Proteinase K (plus 5μl of 10% SDS) for 2 hours. The sample was purified with MiniElute columns of Qiagen according to the manufacturer's recommendations.

Chromatin Immunoprecipitaion in SL-2 cells with Chelex Method

ChIP was carried out with a few modifications. After overnight IP, 10% Inputs were taken, mixed with 3 volumes of 100% ethanol, 15 μg GlycoBlue (Ambion) and was incubated at -80°C during immunocollection by ProtA Sepharose beads (GE Healthcare). The beads were washed as before and finally after TE wash, the beads were resuspended in 100 μl 10 % Chelex solution. Input samples were centrifuged 20000g for 30 minutes, washed with 70% ethanol, vacuum dried at 30°C and resuspended in 100 μl 10% Chelex solution. All samples were decrosslinked at 95°C for 10 minutes, cooled and incubated with 20 μg of Proteinase K for 1 hour at 55°C. The samples were incubated at 95C again for 10 minutes and centrifuged at 20000g. Without taking any Chelex beads, supernatants were transferred to a new tube. The chelex modified protocol was adapted from [157]. Quantitative PCR was carried out with FastStart Universal SYBR Green Master (Roche).

RNA immunoprecipitation

RIP was carried out essentially the same as [158] with 25 million SL-2 stable lines that had been induced with $CuSO_4$ for 12 hours. The cells were fixed with 1% formaldehyde at room temperature for 15 minutes. Formaldehyde is quenched with 200mM Glycine for 5 minutes at room temperature. The cells are pelleted by centrifugation at 1000g for 5 minutes and washed twice with ice-cold PBS. Then the cells are re-suspended in 500μL of FA lysis buffer (50mM HEPES/KOH 7.6, 140 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 40u/mL RNasin, 1xProtease inhibitors) and spun at 3000g for 30 seconds. Supernatant is discarded and the pellet is re-suspended in 600μL of FA lysis buffer and sonicated using a Branson sonifier (Power output: 3, Duty cycle: 40, 10 cycles. Each cycle: 20 seconds on, 50 seconds off). Sonicated extract is transferred into a new tube and centrifuged at maximum speed for 10 minutes at 4°C.

Supernatant is transferred to a new tube and the centrifugation is repeated once. The extract is then pre-cleared over 50µL of Protein A sepharose beads for 1 hr at 4˚C. Beads are removed by centrifuging at 1000g for 2 minutes at 4˚C. Supernatant is transferred into a new tube and made to 25mM $MgCl_2$ and 5mM $CaCl_2$. 3µL RNasin and 6µL RQ1 RNase-free DNase (Promega) is added and the extract is incubated at 37˚C for 30 minutes. The reaction is stopped by adding EDTA to 20mM. Insoluble material is disposed of by centrifuging at maximum speed for 10 min at 4˚C. Supernatant is either used immediately or snap-frozen in liquid nitrogen and stored at -80˚C until use.

Extract coming from about a million cells is diluted to 500µL with FA lysis buffer for each IP. 50µL of this solution is saved as INPUT. 3µL of anti-FLAG (M2) or 3µL of anti-Mle antibody is added and the tubes are incubated in the cold room overnight with end-over-end rotation. The RIPs are centrifuged at maximum speed for 10 minutes at 4˚C and the supernatants are transferred into a new tube that contains 40µL of Protein A or G slurry. Incubation was for 90 minutes in the cold room. Then, the beads are pelleted by centrifugation at 1000g for 2 minutes and washed once with FA lysis buffer, once with FA500 buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) sodium deoxycholate, 1xProtease Inhibitors, 40u/mL RNasin), once with LiCl buffer (10 mM Tris·Cl, pH 8, 250 mM LiCl, 0.5% (v/v) Igepal CA 630, 0.1% (w/v) sodium deoxycholate 1 mM EDTA, 1xProtease Inhibitors, 40u/mL RNasin) and once with TE buffer (10 mM Tris·Cl, pH 8, 1 mM EDTA, 100 mM NaCl) for 5 minutes each in the cold room with end-over-end rotation. TE buffer is removed as much as possible and the immune complexes are eluted with 75µL of RIP elution buffer (100 mM Tris·Cl, pH 8, 10 mM EDTA, 1% (w/v) SDS, 40U/mL RNasin) by incubation at 37˚C for 10 minutes. The beads are pelleted by centrifugation at 1000g for 2 minutes at room temperature, the supernatant is transferred into a new tube and the elution is repeated with another 75µL of RIP elution buffer.

Eluates are pooled, made to 200mM NaCl and 0.13µg/µL Proteinase K. Protease digestion was carried out by incubating the eluates at 42˚C for 1 hour, followed by reversal of formaldehyde cross-links by incubating at 65˚C for 1 hour. Input samples saved before are processed in parallel. Onto each RIP, 100µL nuclease-free water and 250µL acid-phenol: chloroform is added. Phase separation is carried out in MaxTrak tubes (Qiagen) by centrifugation at 10000g for 3 minutes at room temperature. The

aqueous phase is transferred into a new tube and the RNA is precipitated by adding 25µL of 3M sodium acetate, 20µg glycogen and 625µL ice-cold absolute ethanol and incubating the mixture at -20˚C overnight. RIPs are centrifuged at maximum speed for 30 minutes are 4˚C. Supernatant is discarded and the pellet is washed with ice-cold 70% ethanol. After a final centrifugation at maximum speed for 5 minutes at 4˚C, supernatant is discarded and the pellet is air-dried for 5-10 minutes. 90µL nuclease-free water is used for re-solubulizing the RNA pellet. 10µL TURBO DNase buffer and 1µL TURBO DNase (Ambion) is added and the mixture is incubated at 37˚C for 30 minutes. 10µL DNase inactivation reagent is pipetted to each tube and the tubes are incubated at room temperature for 2 minutes with occasional shaking. The tubes are centrifuged at 10,000g for 2 minutes and the supernatants are transferred into fresh tubes. These RNA samples are either used immediately for reverse transcription-qPCR or stored at -80˚C until use. Reverse transcription and PCR were carried out in the same tube by using reverse transcriptase and RNasin in the SYBR mix, and adding a step of 50˚C for 30 minutes before the start of PCR amplification.

Baculovirus mediated expression in Sf21 cells

50 million exponentially grown Sf21 cells were infected with 1 ml high titer virus stocks. After two days infection, the cells were harvested, washed once with PBS and re-suspended in 10 ml of HEMGT 150 buffer. Whole cell extraction was carried out by three cycles of freeze-thaw. For each Flag pull down, 50 µl Anti-Flag M2 Agarose beads were incubated with the extracts for 2 hours at 4°C. The beads were washed 5 times with HEMGT 500 and once with HEMGT 200 buffer. The proteins were eluted with 250 ng/µl 3XFlag peptide overnight at 4°C.

***In vitro* reconstitution assay**

50 million exponentially grown Sf21 cells were infected with 5 ml high titer virus stock. After two days infection, the cells were harvested, washed once with PBS and re-suspended in 10 ml of HEMGT 150 buffer. Whole cell extraction was carried out by three cycles of freeze-thaw. For the dimeric or trimeric reconstitutions, equal amounts of extracts were mixed and incubated for 2 hours at 4°C. For each Flag pull down, 100 µl Anti-Flag M2 Agarose beads were incubated with the extracts for 2 hours at 4°C. The beads were washed twice with HEMGT 500 and once with HEMGT 200 buffer. The

proteins were eluted with 400 ng/μl Flag peptide overnight at 4°C. Eluted proteins were TCA precipitated and loaded on NuPAGE® Bis-Tris 4-12% gradient gels.

In vitro Ubiquitination reactions

hE1 (100 nM), hUbcH5b(250nM), 5 uM HA-Ubiquitin and 5mM ATP were used in the reactions. Proteins at 100 ug/ul were incubated up to 30 minutes at 30°C at slow shaking at a termoshaker. Negative controls were carried out without ATP.

Fly culture and rearing conditions

Unless otherwise specified, flies (*Drosophila melanogaster*) were reared on standard fly medium at 25°C and 70% relative humidity and 12 hrs dark/12 hrs light cycle. All transgenic lines carrying mutant versions of *msl1* were generated through phiC31 integrase-mediated germ-line transformation using *y^1 M{vas-int.Dm}ZH-2A w^*; PBac{y^+-attP-3B}VK00033* (Bloomington stock #24871). The following stocks were obtained from the Bloomington stock centre or were kindly donated: *y^1 w^*; P{tubP-GAL4}LL7/TM3, Sb^1* (Bloomington stock #5138), *$msl1^{L60}$/CyO* (Mitzi Kuroda), *$msl1^{\gamma 269}$ cn^1 bw^1/CyO* (John Lucchesi), *w^*; In(2LR)$noc^{4L}$$Sco^{rv9R}$, b^1/CyO, P{ActGFP}JMR1* (Bloomington stock #4533) *Eip75B$^{j^{n}51}$/TM6B, P{Ubi-GFP.S65T}PAD2, Tb^1* (Bloomington stock #23652). All lines used in this study were generated by standard genetic crosses from the above listed stocks. In order to obtain flies ectopically expressing mutant *msl1* in a wild type background *y^1 w^*; P{tubP-GAL4}LL7/ TM6B, P{Ubi-GFP.S65T}PAD2, Tb^1* virgin females were crossed with males homozygous for the appropriate *UAS-msl1** transgenic insertion. For analysis in *msl1* null mutant background *y^1 w^*; $msl1^{L60}$/CyO, P{ActGFP}JMR1; P{tubP-GAL4}LL7/ TM6B, Tb^1* virgin females were crossed with *y^1 w^*; $msl1^{\gamma 269}$ cn^1 bw^1/CyO, P{ActGFP}JMR1; P{UAS-msl1*}65B2* males.

Preparation of larval protein extracts for Western blot analysis

To extract proteins for immunoblotting, 50 second-instar larvae of the appropriate genotype were collected in 1.5ml Eppendorf tubes on ice and washed with 1ml ice-cold 1 x PBS buffer. The larvae were homogenized in 100 μl 2x ROTH lysis buffer with freshly added 1:1000 antifoam A (Sigma, A5633) and protease inhibitor cocktail (Roche) at 4^0C. Homogenates were sonicated for 10 sec in a Branson 250 sonicator at 40 pulse intensity 1, boiled for 10 min at 95°C and sonicated again. After 5 min

centrifugation at 16000g in a table-top centrifuge at 4°C supernatants were transferred to fresh Eppendorf tubes and either immediately used for SDS-PAGE or stored at -20°C.

Primers

CHIP PRIMERS	FORWARD	REVERSE
roX2 3'	ACG GTG CTG GCT TAG AGA GA	GGC GGA AAT GTA TTT GCA GT
Su(wa)	TGTGTTGTCAGGGATCCAAA	TTCGAGTTGATGCGAACAAG
1B4	ATCTCGCTGTTATCGGTTGC	TCAAGTTGATCGTGGAGCAG
1C4	AAAGAGCACCAACCGATCAC	AAAGACATTTGGTTGTGGGC
1C5	ATATTTTGATCGAGGTGGCG	TCTGGCACATCTTTGAGCAC
2A1	CTGATCGTCGTAAAAACGCA	TAAAAGGCTGCCACAGCTTC
2A4	CGTAGAGCAGTCTGCCATTG	CAATCGCGGTTTTATGACCT
2B1	GAAGCACGTGATATTGGGCT	GCACTACATCAGCACAGCGT
2B7	GACAGGGAGATTAATGGGCA	CTTGTCCGGTCGATTTTTGT
2B14	GCGACTATATAGGACCGCCA	TCCAGAAGTCCGTGTTCCAT
2C4	TCGCCATCTCTGTCTGTCTG	TACTTCCGTTTCCGCTCACT
2C8	TGAAACGTGTCATCTCGCTC	TGACTTCGTTAGGGAATGGC
2F1	GCTAAGTTTGGATTGCCCTG	CTGCATATAGGGTTTGCCGT
3A1	TGGTTGTCCACCAAAGTGAA	CAGGCAACCCATCTCTCTCT
rpl22 promoter	CAA TCC AAT GCG CAG TTA TG	AAG GCC TTG TTC GCA TAT TG
rpl22 middle	TAG CGG TAA GCT GGG CTA AA	GTC GCT CTG ATG GCA GTG TA
rpl22 end	GGC TAG CCC GAA GTT TTC TT	AGC TGA TCC CTT CAG TGG AA
Ucp4a promoter	CAAGTTGTCGCGAGTTGAAA	CAATTGCTTCGCTCTAGCTG
Ucp4a middle	CGCAAGGAGTTCACACAGAA	CTCCATTTGGATTTGCACCT
Ucp4a end	TTCATGTTACCCCGCCTTTA	CTCCTGACATTTGGGCATTC
klp3a promoter	GAGGTGCCGGTGTAGAAAGA	CACAAATCGTCCAACCACAT

klp3a middle	CATTCCCATTCGGAGGAGTA	GCAGCTCCTGTTTGAGATCC
klp3a end	TTGATGTTGGCTGTCGAAAA	TTGAATTCATTCCTGGGTCA
CG4406 promoter	CTGCTCGATAGCACGCAGT	TATCGACGGTCACACTGCTC
CG4406 middle	CCTGGAACTTGAGGAATCCA	GGCAGCAATGTGCTCATCTA
CG4406 end	AGCTCGGAAGGAAACTGTGA	GTGACCAAAAAGCCCTTCAA
CG6506 promoter	GCCGATGTTTACCGACAATC	CATGGTTGGTTATCGGGACT
CG6506 middle	ATCCGTGCCTAATGATACCG	ACGGTTGGTGTGAACCAAAT
CG6506 end	ACAGTCAGCTCCCAGCAGAT	AAAGTGGCGTGAAAGTTGCT
CG3473 upstream	ACTCGGTTCAGATCCTGTGG	GGCCAGTGGGCTTGTAATAA
ODSH	AAGATCCGCTAAGCGATGAA	GCCAGGAGTTGAAGTTGGTC
sec5 promoter	GCT GCT CAG CAA GGA GAC TT	CGG ACG AGC ATA AAA AGA GC
sec5 middle	CCA GGA AAA GGC AAA GAA CA	TCG CAG AAG TTA ACC CGA TT
sec5 end	ATC AAC GGC TTC ATC TTT CG	GCG TTT TCT TCC ATT TTC CA
cycb promoter	CAGTGGCGCTCGAGAATAAT	TATCGCACGTATCGCATCTC
cycb middle	CATCTTGTGGGACACCTCCT	AGCCCATTCACAAGGATCAC
cycb end	TCAGAGACATTTTTGGGAAACA	TCGCACATTCATACAAAAACAA
Nc73EF promoter	CGATATCTGGCACAAAACGA	GCGTTAGTGGTGGGACTGTT
Nc73EF middle	CAATCGGGTCTGGTAATGCT	GGAAGTAGTCGGGATCGTCA
Nc73EF end	TGTTTCAAACCAAGCAAAGC	TGCGCTCCATAAAGGGTATC
NAT1 promoter	CCACACCATGGATTGTGAAA	TGCAGCAATCGATAACCTGT
NAT1 middle	AATCATACGCACTCCGTTCC	CGATCATTGTTGCGCATATC
NAT1 end	CTGTGCCGTTGGTTCAAGTA	GTCGGAAATCTCCTCCTTCC
CG11815 promoter	GGCATACATTTCGGATTGCT	TTGCACGATAATCAGCCGTA
CG11815 middle	GCTTTCCAGCTTCTTCATGG	TCGATTGGACGGAGGACTAC
CG11815 end	TTTTGTTGTCATCGCCTTCA	CATGAGCGAGGATGAACTGA

LNK promoter	TTCCCACCATTTTGCTTACC	CACAGAAAACCCGATTACCG
LNK middle	TTCGGAGAAGGGTCAGTGTC	TTCCAGTGGTATCGGGTTGT
LNK end	CATTTGGGCTTTGGGTTTTA	TTCGTTACCTTATTTCAAACGAA
rpl6 promoter	CGACTTGGCCACCTTTTTAG	TGGTCACACCGCTCTAGACA
rpl6 middle	TTTGTGCTCGCTGAAGTTTG	GAAGAGCAAGGCCTCCTACC
rpl6 end	CGCGCTTGTACAGTGTTTTT	GAGGGGATGGATTTGGTTCT
rlip promoter	CATGACACAGTAGCGCCATC	CCGCTGCTATCCAACACTG
rlip middle	TGCAGGATCACCTGAAGTGT	TGTACAACCGCTTGAAGTGC
rlip end	CGGATTCCTCATGCCTAAAG	ATTTCGGTGGGTCAAGTCTG
RIP PRIMERS	**FORWARD**	**REVERSE**
roX2 (a)	TCGCAATGCAAACTGAAGTC	AGGCGCGTAAAACGTTACC
Rox2 (b)	GAAACGTTCTCCGAAGCAAA	GCGGAAATCGTTACTCTTGC
7SK	GATAACCCGTCGTCATCCAG	AGTAATTCTGCCTGGCGTTG

REFERENCES

REFERENCES

1. Conrad, T. & Akhtar, A. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13**, 123–134 (2012).

2. Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* **12**, 429–442 (2011).

3. Deng, X. *et al.* Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat Genet* **43**, 1179–1185 (2011).

4. Kharchenko, P. V., Xi, R. & Park, P. J. Evidence for dosage compensation between the X chromosome and autosomes in mammals. *Nat Genet* **43**, 1167–1169 (2011).

5. Gelbart, M. E. & Kuroda, M. I. Drosophila dosage compensation: a complex voyage to the X chromosome. *Development* **136**, 1399–1410 (2009).

6. Kelley, R. L. *et al.* Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. *Cell* **81**, 867–877 (1995).

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

8. Gorman, M., Franke, A. & Baker, B. S. Molecular characterization of the male-specific lethal-3 gene and investigations of the regulation of dosage compensation in Drosophila. *Development* **121**, 463–475 (1995).

9. Palmer, M. J. *et al.* The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a novel protein that associates with the X chromosome in males. *Genetics* **134**, 545–557 (1993).

10. Kuroda, M. I., Kernan, M. J., Kreber, R., Ganetzky, B. & Baker, B. S. The maleless

protein associates with the X chromosome to regulate dosage compensation in Drosophila. *Cell* **66**, 935–947 (1991).

11. Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A. & Lucchesi, J. C. mof, a putative acetyl 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 J* **16**, 2054–2060 (1997).

12. Beckmann, K., Grskovic, M., Gebauer, F. & Hentze, M. W. A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. *Cell* **122**, 529–540 (2005).

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

14. Li, F., Parry, D. A. D. & Scott, M. J. The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. *Molecular and Cellular Biology* **25**, 8913–8924 (2005).

15. Morales, V. *et al.* Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J* **23**, 2258–2268 (2004).

16. Copps, K. *et al.* Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. *EMBO J* **17**, 5409–5417 (1998).

17. Marín, I. Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes. *Journal of Molecular Evolution* **56**, 527–539 (2003).

18. Wu, L., Zee, B. M., Wang, Y., Garcia, B. A. & Dou, Y. The RING Finger Protein MSL2 in the MOF Complex Is an E3 Ubiquitin Ligase for H2B K34 and Is Involved in Crosstalk with H3 K4 and K79 Methylation. *Mol Cell* **43**, 132–144 (2011).

19. Morales, V., Regnard, C., Izzo, A., Vetter, I. & Becker, P. B. The MRG domain mediates the functional integration of MSL3 into the dosage compensation

complex. *Molecular and Cellular Biology* **25**, 5947–5954 (2005).

20. Bowman, B. R. *et al.* Multipurpose MRG domain involved in cell senescence and proliferation exhibits structural homology to a DNA-interacting domain. *Structure* **14**, 151–158 (2006).

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

22. Bone, J. R. *et al.* Acetylated histone H4 on the male X chromosome is associated with dosage compensation in Drosophila. *Genes Dev* **8**, 96–104 (1994).

23. Gu, W., Szauter, P. & Lucchesi, J. C. Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster. *Dev Genet* **22**, 56–64 (1998).

24. Akhtar, A. & Becker, P. B. The histone H4 acetyltransferase MOF uses a C2HC zinc finger for substrate recognition. *EMBO Rep* **2**, 113–118 (2001).

25. Akhtar, A. & Becker, P. B. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. *Mol Cell* **5**, 367–375 (2000).

26. Smith, E. R. *et al.* The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Molecular and Cellular Biology* **20**, 312–318 (2000).

27. Buscaino, A. *et al.* MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. *Mol Cell* **11**, 1265–1277 (2003).

28. Nielsen, P. R. *et al.* Structure of the chromo barrel domain from the MOF acetyltransferase. *J Biol Chem* **280**, 32326–32331 (2005).

29. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).

30. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).

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

32. Bouazoune, K. *et al.* The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. *EMBO J* **21**, 2430–2440 (2002).

33. Shimojo, H. *et al.* Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. *J Mol Biol* **378**, 987–1001 (2008).

34. Sural, T. H. *et al.* The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome. *Nat Struct Mol Biol* **15**, 1318–1325 (2008).

35. Larschan, E. *et al.* MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol Cell* **28**, 121–133 (2007).

36. Buscaino, A., Legube, G. & Akhtar, A. X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO Rep* 8 (2006).doi:10.1038/sj.embor.7400658

37. Zhang, P. *et al.* Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res* **34**, 6621–6628 (2006).

38. Sun, B. *et al.* Molecular basis of the interaction of Saccharomyces cerevisiae Eaf3 chromo domain with methylated H3K36. *J Biol Chem* **283**, 36504–36512 (2008).

39. Xu, C., Cui, G., Botuyan, M. V. & Mer, G. Structural basis for the recognition of methylated histone H3K36 by the Eaf3 subunit of histone deacetylase complex Rpd3S. *Structure* **16**, 1740–1750 (2008).

40. Li, B. *et al.* Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science* **316**, 1050–1054 (2007).

41. Buscaino, A., Legube, G. & Akhtar, A. X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO Rep* **7**, 531–

538 (2006).

42. Kim, D. *et al.* Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. *Nat Struct Mol Biol* **17**, 1027–1029 (2010).

43. Moore, S. A., Ferhatoglu, Y., Jia, Y., Al-Jiab, R. A. & Scott, M. J. Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono or dimethyl lysine 20 on histone H4. *J Biol Chem* (2010).doi:10.1074/jbc.M110.134312

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

45. Lee, C. G., Chang, K. A., Kuroda, M. I. & Hurwitz, J. The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. *EMBO J* **16**, 2671–2681 (1997).

46. Izzo, A., Regnard, C., Morales, V., Kremmer, E. & Becker, P. B. Structure-function analysis of the RNA helicase maleless. *Nucleic Acids Res* **36**, 950–962 (2008).

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

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

49. Franke, A. & Baker, B. S. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. *Mol Cell* **4**, 117–122 (1999).

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

51. Kageyama, Y. *et al.* Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site. *EMBO J* **20**,

2236–2245 (2001).

52. Stuckenholz, C., Meller, V. H. & Kuroda, M. I. Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster. *Genetics* **164**, 1003–1014 (2003).

53. Park, S.-W., Kuroda, M. I. & Park, Y. Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in roX noncoding RNAs. *Molecular and Cellular Biology* **28**, 4952–4962 (2008).

54. Meller, V. H. *et al.* Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. *Curr Biol* **10**, 136–143 (2000).

55. Meller, V. H. Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. *Mechanisms of Development* **120**, 759–767 (2003).

56. Schütt, C. & Nöthiger, R. Structure, function and evolution of sex-determining systems in Dipteran insects. *Development* **127**, 667–677 (2000).

57. Bashaw, G. J. & Baker, B. S. 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 (1995).

58. Merendino, L., Guth, S., Bilbao, D., Martínez, C. & Valcárcel, J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**, 838–841 (1999).

59. Förch, P., Merendino, L., Martínez, C. & Valcárcel, J. Modulation of msl-2 5' splice site recognition by Sex-lethal. *RNA* **7**, 1185–1191 (2001).

60. Grskovic, M., Hentze, M. W. & Gebauer, F. A co-repressor assembly nucleated by Sex-lethal in the 3'UTR mediates translational control of Drosophila msl-2 mRNA. *EMBO J* **22**, 5571–5581 (2003).

61. Gebauer, F., Grskovic, M. & Hentze, M. W. Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. *Mol Cell* **11**,

1397–1404 (2003).

62. Kelley, R. L., Wang, J., Bell, L. & Kuroda, M. I. Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. *Nature* **387**, 195–199 (1997).

63. Duncan, K. *et al.* Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3' UTR: translational repression for dosage compensation. *Genes Dev* **20**, 368–379 (2006).

64. Abaza, I., Coll, O., Patalano, S. & Gebauer, F. Drosophila UNR is required for translational repression of male-specific lethal 2 mRNA during regulation of X-chromosome dosage compensation. *Genes Dev* **20**, 380–389 (2006).

65. Patalano, S., Mihailovich, M., Belacortu, Y., Paricio, N. & Gebauer, F. Dual sex-specific functions of Drosophila Upstream of N-ras in the control of X chromosome dosage compensation. *Development* **136**, 689–698 (2009).

66. Rastelli, L. & Kuroda, M. I. An analysis of maleless and histone H4 acetylation in Drosophila melanogaster spermatogenesis. *Mechanisms of Development* **71**, 107–117 (1998).

67. Rastelli, L., Richman, R. & Kuroda, M. I. The dosage compensation regulators MLE, MSL-1 and MSL-2 are interdependent since early embryogenesis in Drosophila. *Mechanisms of Development* **53**, 223–233 (1995).

68. Gupta, V. *et al.* Global analysis of X-chromosome dosage compensation. *J Biol* **5**, 3 (2006).

69. Baker, B. S., Gorman, M. & Marín, I. Dosage compensation in Drosophila. *Annu. Rev. Genet.* **28**, 491–521 (1994).

70. Zhang, Y. & Oliver, B. Dosage compensation goes global. *Curr Opin Genet Dev* **17**, 113–120 (2007).

71. Franke, A., Dernburg, A., Bashaw, G. J. & Baker, B. S. Evidence that MSL-mediated dosage compensation in Drosophila begins at blastoderm. *Development* **122**, 2751–2760 (1996).

72. McDowell, K. A., Hilfiker, A. & Lucchesi, J. C. Dosage compensation in Drosophila: the X chromosome binding of MSL-1 and MSL-2 in female embryos is prevented by the early expression of the Sxl gene. *Mechanisms of Development* **57**, 113–119 (1996).

73. Kind, J. *et al.* Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. *Cell* **133**, 813–828 (2008).

74. Gilfillan, G. D. *et al.* Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex. *Genes Dev* **20**, 858–870 (2006).

75. Legube, G. X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in Drosophila. *Genes Dev* **20**, 871–883 (2006).

76. Alekseyenko, A. A. High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. *Genes Dev* **20**, 848–857 (2006).

77. Alekseyenko, A. A. *et al.* A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* **134**, 599–609 (2008).

78. Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A. & Becker, P. B. The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. *PLoS Genet* **4**, e1000302 (2008).

79. Kotlikova, I. V. *et al.* The Drosophila dosage compensation complex binds to polytene chromosomes independently of developmental changes in transcription. *Genetics* **172**, 963–974 (2006).

80. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).

81. Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).

82. Bell, O. *et al.* Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster. *Molecular and Cellular Biology* **28**, 3401–3409

(2008).

83. Mendjan, S. *et al.* Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Mol Cell* **21**, 811–823 (2006).

84. Raja, S. J. *et al.* The nonspecific lethal complex is a transcriptional regulator in Drosophila. *Mol Cell* **38**, 827–841 (2010).

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

86. Lyman, L. M., Copps, K., Rastelli, L., Kelley, R. L. & Kuroda, M. I. Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* **147**, 1743–1753 (1997).

87. Kelley, R. L. *et al.* Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. *Cell* **98**, 513–522 (1999).

88. Park, Y. *et al.* Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex. *Mol Cell* **11**, 977–986 (2003).

89. Fagegaltier, D. & Baker, B. S. X Chromosome Sites Autonomously Recruit the Dosage Compensation Complex in Drosophila Males. *PLoS Biol* **2**, e341 (2004).

90. Oh, H., Bone, J. R. & Kuroda, M. I. Multiple classes of MSL binding sites target dosage compensation to the X chromosome of Drosophila. *Curr Biol* **14**, 481–487 (2004).

91. Dahlsveen, I. K., Gilfillan, G. D., Shelest, V. I., Lamm, R. & Becker, P. B. Targeting Determinants of Dosage Compensation in Drosophila. *PLoS Genet* **2**, e5 (2006).

92. Ilik, I. & Akhtar, A. roX RNAs: Non-coding regulators of the male X chromosome in flies. *RNA Biol* **6**, (2009).

93. Li, F., Schiemann, A. H. & Scott, M. J. Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. *Molecular and Cellular Biology* **28**, 1252–1264 (2008).

94. Oh, H. Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. *Genes Dev* **17**, 1334–1339 (2003).

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

96. Deng, X., Rattner, B. P., Souter, S. & Meller, V. H. The severity of roX1 mutations is predicted by MSL localization on the X chromosome. *Mechanisms of Development* **122**, 1094–1105 (2005).

97. Deng, X. & Meller, V. H. Molecularly severe roX1 mutations contribute to dosage compensation in Drosophila. *Genesis* **47**, 49–54 (2009).

98. Kelley, R., Lee, O. & Shim, Y. Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. *Mechanisms of Development* **125**, 1009–1019 (2008).

99. Bai, X., Alekseyenko, A. A. & Kuroda, M. I. Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes. *EMBO J* **23**, 2853–2861 (2004).

100. Rattner, B. P. & Meller, V. H. Drosophila male-specific lethal 2 protein controls sex-specific expression of the roX genes. *Genetics* **166**, 1825–1832 (2004).

101. Lee, C.-G., Reichman, T. W., Baik, T. & Mathews, M. B. MLE functions as a transcriptional regulator of the roX2 gene. *J Biol Chem* **279**, 47740–47745 (2004).

102. Gu, W., Wei, X., Pannuti, A. & Lucchesi, J. C. 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 J* **19**, 5202–5211 (2000).

103. Morra, R., Smith, E. R., Yokoyama, R. & Lucchesi, J. C. The MLE subunit of the Drosophila MSL complex uses its ATPase activity for dosage compensation and its helicase activity for targeting. *Molecular and Cellular Biology* **28**, 958–966 (2008).

104. Kind, J. & Akhtar, A. Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes. *Genes Dev* **21**, 2030–2040 (2007).

105. Sass, G. L., Pannuti, A. & Lucchesi, J. C. Male-specific lethal complex of Drosophila targets activated regions of the X chromosome for chromatin remodeling. *Proc Natl Acad Sci USA* **100**, 8287–8291 (2003).

106. Carvalho, A. B. Origin and evolution of the Drosophila Y chromosome. *Curr Opin Genet Dev* **12**, 664–668 (2002).

107. Steinemann, S. & Steinemann, M. Y chromosomes: born to be destroyed. *Bioessays* **27**, 1076–1083 (2005).

108. Marín, I., Franke, A., Bashaw, G. J. & Baker, B. S. The dosage compensation system of Drosophila is co-opted by newly evolved X chromosomes. *Nature* **383**, 160–163 (1996).

109. Gallach, M., Arnau, V. & Marín, I. Global patterns of sequence evolution in Drosophila. *BMC Genomics* **8**, 408 (2007).

110. Bone, J. R. & Kuroda, M. I. Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila. *Genetics* **144**, 705–713 (1996).

111. Charlesworth, B., Charlesworth, D., Hnilicka, J., Yu, A. & Guttman, D. S. Lack of degeneration of loci on the neo-Y chromosome of Drosophila americana americana. *Genetics* **145**, 989–1002 (1997).

112. Macknight, R. H. The Sex-Determining Mechanism of Drosophila Miranda. *Genetics* **24**, 180–201 (1939).

113. Norman, R. A. & Doane, W. W. Dosage compensation and dietary glucose repression of larval amylase activity in Drosophila miranda. *Biochem Genet* **28**, 601–613 (1990).

114. Bachtrog, D. Expression profile of a degenerating neo-y chromosome in Drosophila. *Curr Biol* **16**, 1694–1699 (2006).

115. Steinemann, S. & Steinemann, M. Evolution of sex chromosomes: dosage compensation of the Lcp1-4 gene cluster on the evolving neo-X chromosome in

Drosophila miranda. *Insect Mol Biol* **16**, 167–174 (2007).

116. Yi, S., Bachtrog, D. & Charlesworth, B. A survey of chromosomal and nucleotide sequence variation in Drosophila miranda. *Genetics* **164**, 1369–1381 (2003).

117. Rodriguez, M. A., Vermaak, D., Bayes, J. J. & Malik, H. S. Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila. *Proc Natl Acad Sci USA* **104**, 15412–15417 (2007).

118. Kelley, R. L. Path to equality strewn with roX. *Dev Biol* **269**, 18–25 (2004).

119. Veneti, Z., Bentley, J. K., Koana, T., Braig, H. R. & Hurst, G. D. D. A functional dosage compensation complex required for male killing in Drosophila. *Science* **307**, 1461–1463 (2005).

120. Matyunina, L. V., Bowen, N. J. & Mcdonald, J. F. LTR retrotransposons and the evolution of dosage compensation in Drosophila. *BMC Mol Biol* **9**, 55 (2008).

121. Eisen, A. The Yeast NuA4 and Drosophila MSL Complexes Contain Homologous Subunits Important for Transcription Regulation. *Journal of Biological Chemistry* **276**, 3484–3491 (2001).

122. Smith, E. R. *et al.* A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Molecular and Cellular Biology* **25**, 9175–9188 (2005).

123. Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Molecular and Cellular Biology* **25**, 6798–6810 (2005).

124. Kruse, J.-P. & Gu, W. MSL2 promotes Mdm2-independent cytoplasmic localization of p53. *J Biol Chem* **284**, 3250–3263 (2009).

125. Nguyen, D. K. & Disteche, C. M. Dosage compensation of the active X chromosome in mammals. *Nat Genet* **38**, 47–53 (2006).

126. Yildirim, E., Sadreyev, R. I., Pinter, S. F. & Lee, J. T. X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nat Struct Mol Biol* **19**, 56–61 (2011).

127. Bhadra, M. P., Bhadra, U., Kundu, J. & Birchler, J. A. Gene expression analysis of the function of the male-specific lethal complex in Drosophila. *Genetics* **169**, 2061–2074 (2005).

128. Bhadra, U., Pal-Bhadra, M. & Birchler, J. A. Histone acetylation and gene expression analysis of sex lethal mutants in Drosophila. *Genetics* **155**, 753–763 (2000).

129. Bhadra, U., Pal-Bhadra, M. & Birchler, J. A. Role of the male specific lethal (msl) genes in modifying the effects of sex chromosomal dosage in Drosophila. *Genetics* **152**, 249–268 (1999).

130. Birchler, J. A., Pal-Bhadra, M. & Bhadra, U. Abstract - SpringerLink. *Genetica* **117**, 179–190 (2003).

131. Straub, T., Dahlsveen, I. K. & Becker, P. B. Dosage compensation in flies: mechanism, models, mystery. *FEBS Lett* **579**, 3258–3263 (2005).

132. Hamada, F. N., Park, P. J., Gordadze, P. R. & Kuroda, M. I. Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster. *Genes Dev* **19**, 2289–2294 (2005).

133. Henry, R. A., Tews, B., Li, X. & Scott, M. J. 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. *J Biol Chem* **276**, 31953–31958 (2001).

134. Kelley, R. L. & Kuroda, M. I. The Drosophila roX1 RNA gene can overcome silent chromatin by recruiting the male-specific lethal dosage compensation complex. *Genetics* **164**, 565–574 (2003).

135. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844–847 (2006).

136. Deniaud, E. & Bickmore, W. A. Transcription and the nuclear periphery: edge of darkness? *Curr Opin Genet Dev* **19**, 187–191 (2009).

137. Branco, M. R. & Pombo, A. Chromosome organization: new facts, new models.

Trends in Cell Biology **17**, 127–134 (2007).

138. Akhtar, A. & Gasser, S. M. The nuclear envelope and transcriptional control. *Nat Rev Genet* **8**, 507–517 (2007).

139. Köhler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* **8**, 761–773 (2007).

140. Brown, C. & Silver, P. Transcriptional regulation at the nuclear pore complex. *Curr Opin Genet Dev* **17**, 100–106 (2007).

141. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).

142. Jin, Y., Wang, Y., Johansen, J. & Johansen, K. M. JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *The Journal of Cell Biology* **149**, 1005–1010 (2000).

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

144. Zhang, W. *et al.* The JIL-1 histone H3S10 kinase regulates dimethyl H3K9 modifications and heterochromatic spreading in Drosophila. *Development* **133**, 229–235 (2006).

145. Ivaldi, M. S., Karam, C. S. & Corces, V. G. Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in Drosophila. *Genes Dev* **21**, 2818–2831 (2007).

146. Saunders, A., Core, L. J. & Lis, J. T. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7**, 557–567 (2006).

147. Cai, W. *et al.* RNA polymerase II-mediated transcription at active loci does not require histone H3S10 phosphorylation in Drosophila. *Development* **135**, 2917–2925 (2008).

148. Spierer, A., Seum, C., Delattre, M. & Spierer, P. Loss of the modifiers of

variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. *Journal of Cell Science* **118**, 5047–5057 (2005).

149. Deuring, R. *et al.* The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* **5**, 355–365 (2000).

150. Bouazoune, K. & Brehm, A. ATP-dependent chromatin remodeling complexes in Drosophila. *Chromosome Res* **14**, 433–449 (2006).

151. Corona, D. F. V., Clapier, C. R., Becker, P. B. & Tamkun, J. W. Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep* **3**, 242–247 (2002).

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

153. Bai, X., Larschan, E., Kwon, S. Y., Badenhorst, P. & Kuroda, M. I. Regional control of chromatin organization by noncoding roX RNAs and the NURF remodeling complex in Drosophila melanogaster. *Genetics* **176**, 1491–1499 (2007).

154. Spierer, A., Begeot, F., Spierer, P. & Delattre, M. SU(VAR)3-7 links heterochromatin and dosage compensation in Drosophila. *PLoS Genet* **4**, e1000066 (2008).

155. Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).

156. Furuhashi, H., Nakajima, M. & Hirose, S. DNA supercoiling factor contributes to dosage compensation in Drosophila. *Development* **133**, 4475–4483 (2006).

157. Nelson, J., Denisenko, O. & Bomsztyk, K. The fast chromatin immunoprecipitation method. *Methods Mol Biol* **567**, 45–57 (2009).

158. Selth, L. A., Gilbert, C. & Svejstrup, J. Q. RNA Immunoprecipitation to Determine RNA-Protein Associations In Vivo. *Cold Spring Harbor Protocols* **2009**, pdb.prot5234–pdb.prot5234 (2009).

159. Fauth, T., Müller-Planitz, F., König, C., Straub, T. & Becker, P. B. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. *Nucleic Acids Res* **38**, 3209–3221 (2010).

160. Groth, A. C. Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage C31. *Genetics* **166**, 1775–1782 (2004).

161. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).

162. Mondal, K. *et al.* Design and Isolation of Temperature-sensitive Mutants of Gal4 in Yeast and Drosophila. *J Mol Biol* **370**, 939–950 (2007).

163. Mank, J. E., Hosken, D. J. & Wedell, N. SOME INCONVENIENT TRUTHS ABOUT SEX CHROMOSOME DOSAGE COMPENSATION AND THE POTENTIAL ROLE OF SEXUAL CONFLICT. *Evolution* **65**, 2133–2144 (2011).

164. Bhadra, M. P. *et al.* The role of MOF in the ionizing radiation response is conserved in Drosophila melanogaster. *Chromosoma* **121**, 79–90 (2011).

165. Zhou, Y. *et al.* Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding RNA-dependent silencing. *Nat Cell Biol* **11**, 1010–1016 (2009).

166. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol Cell* (2011).doi:10.1016/j.molcel.2011.08.027

167. Larschan, E. *et al.* X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. *Nature* **471**, 115–118 (2011).

PUBLISHED WORK

1. Chromosome Res. 2009;17(5):603-19.

X chromosomal regulation in flies: when less is more.

Hallacli E, Akhtar A.

In Drosophila, dosage compensation of the single male X chromosome involves upregulation of expression of X linked genes. Dosage compensation complex or the male specific lethal (MSL) complex is intimately involved in this regulation. The MSL complex members decorate the male X chromosome by binding on hundreds of sites along the X chromosome. Recent genome wide analysis has brought new light into X chromosomal regulation. It is becoming increasingly clear that although the X chromosome achieves male specific regulation via the MSL complex members, a number of general factors also impinge on this regulation. Future studies integrating these aspects promise to shed more light into this epigenetic phenomenon.

2. Nat Struct Mol Biol. 2011 Feb;18(2):142-9. Epub 2011 Jan 9.

Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1.

Kadlec J*, **Hallacli E**,* Lipp M, Holz H, Sanchez-Weatherby J, Cusack S, Akhtar A.

The male-specific lethal (MSL) complex is required for dosage compensation in Drosophila melanogaster, and analogous complexes exist in mammals. We report structures of binary complexes of mammalian MSL3 and the histone acetyltransferase (HAT) MOF with consecutive segments of MSL1. MSL1 interacts with MSL3 as an extended chain forming an extensive hydrophobic interface, whereas the MSL1-MOF interface involves electrostatic interactions between the HAT domain and a long helix of MSL1. This structure provides insights into the catalytic mechanism of MOF and enables us to show analogous interactions of MOF with NSL1. In Drosophila, selective disruption of Msl1 interactions with Msl3 or Mof severely affects Msl1 targeting to the body of

dosage-compensated genes and several high-affinity sites, without affecting promoter binding. We propose that Msl1 acts as a scaffold for MSL complex assembly to achieve specific targeting to the X chromosome.
* Joint first authors

This paper contains my work related to the PEHE domain of Msl1 and qualitative differences in the High affinity sites.

3 Dev Cell. 2012 Mar 13;22(3):610-24.

The MOF Chromobarrel Domain Controls Genome-wide H4K16 Acetylation and Spreading of the MSL Complex.

Conrad T, Cavalli FM, Holz H, **Hallacli E,** Kind J, Ilik I, Vaquerizas JM, Luscombe NM, Akhtar A.

The histone H4 lysine 16 (H4K16)-specific acetyltransferase MOF is part of two distinct complexes involved in X chromosome dosage compensation and autosomal transcription regulation. Here we show that the MOF chromobarrel domain is essential for H4K16 acetylation throughout the Drosophila genome and is required for spreading of the male-specific lethal (MSL) complex on the X chromosome. The MOF chromobarrel domain directly interacts with nucleic acids and potentiates MOF's enzymatic activity after chromatin binding, making it a unique example of a chromo-like domain directly controlling acetylation activity in vivo. We also show that the Drosophila-specific N terminus of MOF has evolved to perform sex-specific functions. It modulates nucleosome binding and HAT activity and controls MSL complex assembly, thus regulating MOF function in dosage compensation. We propose that MOF has been especially tailored to achieve tight regulation of its enzymatic activity and enable its dual role on X and autosomes.

For this paper, I made stable SL-2 cell lines expressing different Mof mutants and checked their interactions with the rest of the complex by immunoprecipitation. I also investigated interaction of Mof mutants with roX2 RNA.

WORK IN SUBMISSION

MSL1 mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila

Erinc Hallacli, Michael Lipp, Plamen Georgiev, Clare Spielman, Stephen Cusack Asifa Akhtar and Jan Kadlec

The male specific lethal (MSL) complex regulates dosage compensation of the male X chromosome in *Drosophila*. Here, we report the crystal structure of its MSL1-MSL2 core, where two MSL2 subunits bind to a dimer formed by two molecules of MSL1. Analysis of structure-based mutants revealed that MSL2 can only interact with the MSL1 dimer, but MSL1 dimerization occurs independent of MSL2. We show that MSL1 is a substrate for MSL2 E3 ubiquitin ligase activity. ChIP experiments revealed that MSL1 dimerization is essential for targeting and spreading of the MSL complex on X-linked genes, however MSL1 binding to promoters of male and female cells is independent of the dimer status and other MSL proteins. Finally, we show that loss of MSL1 dimerization leads to male specific lethality in transgenic flies. We propose that MSL1 mediated dimerization of the entire MSL complex is required for MSL2 binding, X chromosome recognition and spreading along the X chromosome.

This body of work includes the data related to interaction of Msl1 and Msl2 and dimerization of Msl1.