Drosophila dosage compensation
Males are from Mars, females are from Venus

Plamen Georgiev, Sarantis Chlamydas and Asifa Akhtar*
Max Planck Institute of Immunobiology and Epigenetics; Freiburg, Germany

Key words: dosage compensation, MSL, roX, transcription, acetylation, methylation, X chromosome, nuclear pore complex

Dosage compensation of X-linked genes is a phenomenon of concerted, chromosome-wide regulation of gene expression underpinned by sustained and tightly regulated histone modifications and chromatin remodeling, coupled with constrains of nuclear architecture. This elaborate process allows the accomplishment of regulated expression of genes on the single male X chromosome to levels comparable to those expressed from the two X chromosomes in females. The ribonucleoprotein Male Specific Lethal (MSL) complex is enriched on the male X chromosome and is intricately involved in this process in Drosophila melanogaster. In this review we discuss the recent advances that highlight the complexity lying behind regulation of gene expression by just 2-fold.

The need to compensate for the hemizygosity of the X-linked genes in heterogametic males and to equalize their expression levels with that of the two X chromosomes in females has led to the development of dosage compensation mechanisms. These mechanisms have evolved through the combination of preexisting proteins acquiring new functions in the context of de novo complexes and gene products specifically dedicated to the new dosage compensation activity. The fascinating insights into the sex-specific dosage compensation phenomena are mainly derived from three groups of organisms: fruit flies (Drosophila melanogaster), round worms (Caenorhabditis elegans) and mammals (mice, Mus musculus and humans, Homo sapiens) (Fig. 1).

In these model organisms, males are the heteromorphic sex (XY in Drosophila and mammals and XO in C. elegans), whereas the homomorphic animals (XX) are either females (Drosophila and mammals) or hermaphrodites (C. elegans).

In flies, dosage compensation of the X-linked genes leads to a two-fold upregulation in males in comparison to females. This is mediated by the dosage compensation complex (DCC) also known as Male Specific Lethal complex (MSL), due to the male specific lethality phenotype upon loss-of-function of its major components. The Drosophila MSL is a ribonucleoprotein complex and is composed of at least five proteins, namely MSL-1 (male-specific lethal 1, scaffolding protein), MSL-2 (male-specific lethal 2, RING finger protein), MSL-3 (male-specific lethal 3, chromodomain protein), MOF (males absent on the first, histone acetyltransferase) and MLE (maleless, RNA helicase), and two functionally redundant long non-coding RNAs: roX1 (RNA on the X 1) and roX2 (RNA on the X 2) (Figs. 2 and 3). The MSL complex decorates the male X chromosome which is also hyperacetylated at histone H4 lysine 16 (reviewed in refs. 1 and 2).

In worms, dosage compensation is achieved by downregulation by half of the transcriptional output from both X chromosomes in hermaphrodites (Fig. 1). This process is mediated by a multicomponent complex, also known as the DCC, which consists of a number of condensin-like proteins. Similar to flies, the worm DCC is also enriched on the dosage compensated X chromosome (reviewed in ref. 3). However, in contrast to Drosophila or mammals (see below), non-coding RNAs contributing to its function have not been identified so far. The exact mechanism by which the C. elegans DCC is targeted to the X chromosomes is unknown, but discrete recruitment elements, namely rex45 and dax67 sites on the X chromosome have recently been identified.

In mammals, dosage compensation is achieved by inactivation of one of the two female X chromosomes. Interestingly, X chromosome inactivation requires the concerted action of a number of non-coding RNAs (reviewed in ref. 8). A recent new addition to the list of non-coding RNAs involved in X inactivation is Jpx. Work over the last two decades has shown that X-inactivating spe specific transcript (Xist), a non-coding RNA expressed from the X chromosome inactivation center (Xic), plays a major role in X inactivation. It acts in cis by coating the entire chromosome from which it is produced and triggers a cascade of chromatin modification events (Fig. 1). The search for protein components mediating X chromosome inactivation is currently ongoing. Nevertheless, recruitment of the Polycomb repressive complexes and histone H3 lysine 27 methylation has been implicated. Furthermore, the E3 ubiquitin ligase RNF12 has recently been shown to also play an important role in X inactivation.13,14

In this review, we focus on recent advances in our understanding of the dosage compensation process in Drosophila.

Solution for Keeping X Chromosomal Balance in Female Flies

An inevitable consequence of dosage compensation systems in which X-linked genes get transcriptionally upregulated is the risk of overexpression of the X-linked genes in females, where excess of the corresponding products could be as deleterious as their...
deficiency in males. Therefore, mechanisms have co-evolved with dosage compensation in males to prevent overexpression in females. In Drosophila, an exquisite interplay between two proteins MSL-2 (expressed in males) and Sex Lethal (SXL; expressed in females) ensures that the “holo” MSL complex is stably expressed only in males.

In females, expression of SXL is triggered from an “early” Sxl promoter.15 In the presence of early SXL, female-specific splicing of Sxl pre-mRNA is set up and maintained in females from a “late” promoter and the translation-terminating exon 3 is removed after splicing.16 Since little or no SXL is present in males, Sxl expression is present in males, Sxl pre-mRNA is spliced by default, the sequence that carries a translational stop codon is retained and this results in a truncated and inactive SXL protein. Following this stage, Sxl expression is shut down in males but maintained in females by a positive auto-regulatory loop determining sexual identity for the rest of the life cycle (reviewed in ref. 17).

Upon establishment of stable SXL expression in females, msl-2 expression is controlled at two levels. First, SXL binds multiple poly(U) sequences within msl-2 pre-mRNA and guides the alternative splicing of the first intron; out in males but retained in females. This leads to instability of the msl-2 mRNA and a substantial downregulation of the transcript in females.18,19 Second, MSL-2 is translationally silenced in Drosophila females by SXL, which binds to poly(U) stretches in the msl-2 mRNA 5’ and 3’ UTRs.18,20 An SXL 3’ UTR repressor complex blocks the 43S pre-initiation complex, independently of a separate 5’UTR bound SXL mechanism, which is able to inhibit ribosomal scanning.21 The resulting absence of MSL-2 in females leads to destabilization and partial degradation of MSL-1,22 and to a lower extent MSL-3 and roX RNAs. Consequently, this renders the dosage compensation complex inactive in females, which is a prerequisite for their viability. This is confirmed by the striking observation that ectopic MSL-2 expression in female flies induces 85% lethality which goes up to 100% upon simultaneous ectopic expression of MSL-1,23 whereas reduction of MSL-1 by half completely relieves the MSL-2 overexpression toxicity.22

Interestingly, a recent study has suggested that early during embryonic development, prior to the full Sxl activation, a transient assembly of MSL complexes and dosage compensation occurs. This is reliant on the maternal contribution and zygotic roX1 expression and also occurs in females. It is required for the sex determination signal accumulation, the upregulation of Sxl transcription and subsequent shut down of MSL-2 expression.24 This effect is surprising given that neither the MSL-2 protein alone22,25 nor the MSL complex have been previously detected in female embryos.26,27 However, maternal effect of msl-2 has been previously described.28 Nevertheless, to firmly establish the mechanism of this regulation and whether there is indeed MSL-2 protein expression and functional MSL assembly concomitant with targeting to the Sxl locus, in female embryos, further investigation is required.

**Recognizing and Targeting the Male X Chromosome**

Regardless of the sex and the need to upregulate or downregulate X-linked gene expression, the recognition and targeting to the X chromosome is a major problem faced by the dosage compensation complex.

The recent genome-wide techniques such as chromatin immunoprecipitation coupled with microarray technology (ChIP-on-chip) or deep sequencing (ChIP-seq) have allowed high-resolution mapping of MSL proteins.29-34 One major finding was that on a genome-wide scale, the MSL targets are enriched in the gene body with a peak at their 3’ends29-33 (Fig. 3). Furthermore, not all MSL proteins behave similarly as the MOF protein also binds to promoter proximal regions on the X chromosome as well as to a large number of autosomal promoters, independently of the other MSL complex members.32 Accordingly, MOF has recently been reported to be a component of a second complex named non-specific lethal (NSL) complex which binds to more than 4,000 genes at their promoter regions and acts as a major transcriptional regulator in Drosophila.35-37 Fascinatingly, the existence of NSL complexes and their interaction with MOF has also been reported in mammals. However, while the mammalian MSL-associated MOF acetylates nucleosomal histone H4 almost exclusively on lysine 16, NSL-associated MOF exhibits a relaxed specificity and also acetylates nucleosomal histone H4 on lysines 5 and 8.35,38 Future studies will be crucial in revealing the interplay between the MSL and the NSL complexes in transcription regulation in Drosophila and mammals.

The enrichment of the MSL complex towards the 3’ end of genes has led to the proposal that the MSL complex could facilitate elongation of transcription.39 Most target genes seem to be actively transcribed, reaffirming previous reports that MSL complex recruitment depends on passage of the transcription machinery, but not on the type of promoter or directionality of transcription.40 At the level of polytene chromosome stainings, it is not possible to detect MSL complex recruitment on autosomal translocations on the X chromosome.41,42 However, it was shown recently by ChIP analysis that active autosomal transgenes
inserted on the X chromosome are indeed capable of MSL recruitment. These results suggest that in an X chromosomal context transcription activity does contribute to MSL complex recruitment, such that also active autosomal genes benefit, if inserted on the X chromosome.

Since not all actively transcribed genes are bound by the MSL complex, transcription per se may not be the only signal for MSL recruitment and dosage compensation. This notion is further supported by the observation that MSL-1 binding profiles at different developmental stages are similar. Therefore, the decision of which genes will be subject to dosage compensation is very likely taken early during development and most of them are bound regardless of developmental changes. However, a recent report showing that many X chromosomal genes bind MSL complexes only too transiently for the interaction to be detected by conventional methods, make it still conceivable that transcription activation is a prerequisite but may not be sufficient for MSL recruitment.

Is it possible that the X chromosome possesses special sequences that help to recruit MSL complexes? Despite the numerous known MSL targets a universal targeting sequence motif has been difficult to identify. Early in situ hybridization studies on Drosophila polytene chromosomes showed that a \((dC - dA)_n(dG - dT)_n\) sequence is enriched on the X chromosome. More recently whole genome sequence analysis revealed that the X chromosome can be distinguished from the other chromosomes based on the CAn and G/Tn repeats as a sequence signature. A number of high affinity binding sites (HAS) on the X chromosome have been defined on the basis that partial complexes of MSL-1-MSL-2 could still bind these sites even in the absence of the rest of the MSL components, enforcing the idea that sequence recognition motifs for MSL targeting do exist. Interestingly, the two roX genes located on the X chromosome act as high affinity sites and play an important role in assembly and spreading of the MSL complex (Fig. 3) (reviewed in ref. 2). HAS were also previously called chromatin entry sites (CES) as according to the currently accepted model, MSL complexes bind to these discrete number of sites along the X chromosome and spread into the local chromatin. It has been a reasonable hypothesis that high affinity sites should have the highest probability for carrying the MSL targeting sequence motifs and recent efforts have been focused on analysis of these sequences. A “one-hybrid” assay approach suggested degenerate sequence motifs that represent weak DNA determinants whose cumulative contributions are required to form a HAS for MSL-2 targeting. A ChIP-on-chip and ChIP-seq approach identified roughly 150 putative HAS and a GA-rich or TC-rich MSL recognition element (MRE) that seems to be of functional relevance. A parallel study identified 131 HAS carrying GA- and CA-rich motifs. Most recently, oligonucleotide profiling, a newly developed method for DNA sequence analysis, observed lower complexity of the X chromosome in Drosophila in comparison with the autosomes. A repetitive sequence motif \([G(CG)N4]\)-like sequences within the low affinity sites. It is possible that a perfect sequence satisfying all the targeting criteria does not exist. At this stage, it seems likely that the bona fide targeting of MSL complexes to the X chromosome is a result of the combination of different factors including active transcription and sequence motifs, which are insufficient for complex recruitment when present alone. It is also becoming clear that the local chromatin context may also be a defining factor in DCC recruitment and its specificity.

Histone Modification Crosstalk on the Male X

It has become increasingly evident in the past several years that factors, in addition to the DNA sequence motifs, could be required for the faithful targeting of MSL to the X chromosome.

Figure 2. MSL proteins decorate the male X chromosome. Polyten stainings from third instar larvae immunostained with specific antibodies raised against MSL-1 and MOF. The figure shows enrichment of these proteins on the male X chromosome. DNA is stained with Hoechst322 (blue), MSL-1 (red), MOF (green). MOF protein can also been seen localized on autosomes albeit with lower intensity compared to the X chromosome.
involvement of the MOF and histone H4K16Ac in a variety of processes during evolution ranging from embryonic development\(^6^3,6^4\) and DNA damage repair\(^6^5-6^8\) to general transcription regulation.\(^6^9\) Future studies should elaborate the role of MSL proteins beyond dosage compensation.

Interestingly, the peak of MSL binding at the 3’ end of the dosage compensated genes also coincides with trimethylation of lysine 36 on histone H3 (H3K36me3),\(^7^0,7^1\) a well-established mark for active genes.\(^7^2,7^3\) A crosstalk between H3K36me3 and H4K16ac was suggested based on the observation that down-regulation of Hypb, the enzyme required for the methylation of H3K36me3 in Drosophila, leads to reduced H4K16 acetylation on X-linked genes.\(^7^0\)

MSL-3 binds H3K36me3 nucleosomes in in vitro assays.\(^7^1\) However, recent structural work suggests co-recognition of DNA and a histone H4 tail monomethylated on lysine 20 (H4K20me) by the MSL-3 chromodomain.\(^7^4\) In a parallel study, MSL-3 affinity for H4K20me was confirmed.\(^6^7\) However, MSL-3/nucleic acid interaction was not reported.\(^6^7\) Interestingly, in both reports H3K36me3 peptide was shown not to be the preferred substrate for the MSL-3 chromodomain.\(^7^4,7^5\) How the presence of H3K36me3 on active X-linked genes integrates into the targeting model needs to be further addressed. At the same time, there is and dosage compensation. Local chromatin context may constitute one obvious factor. Posttranslational modifications of histones within the nucleosomes can have a huge impact on chromatin structure and gene transcription, and represent a wealth of epigenetic information.\(^5^2\) Examples of covalent modifications of histone tails include acetylation, phosphorylation, sumoylation, ubiquitination, and methylation.

A large body of evidence compiled in the recent years has demonstrated the impact of histone acetylation on transcriptional activity (reviewed in refs. 53–55). Gene activation for dosage compensation involves the MSL-associated MOF acetyltransferase activity on H4K16 (histone H4 lysine 16), which represents a hallmark of the male X chromosome.\(^3^2,5^6-5^8\) The mechanism by which H4K16ac contributes to dosage compensation remains an active area of study.\(^3^3,3^6,4^5,5^9\) One possible mechanism could be mediated by the structural changes imposed by the presence of this modification on nucleosomes, thus creating an open chromatin environment for transcription associated complexes.\(^6^0\) A recent genome-wide study showed increased DNA accessibility at active promoters and chromosomal regions that are hyperacetylated at H4K16.\(^5^9\) A similar function is attributed to H3S10ph\(^6^1\) and H3K4me2.\(^5^2\) Interestingly, data begins to accumulate elaborating on the involvement of the MOF and histone H4K16Ac in a variety of processes during evolution ranging from embryonic development\(^6^3,6^4\) and DNA damage repair\(^6^5-6^8\) to general transcription regulation.\(^6^9\) Future studies should elaborate the role of MSL proteins beyond dosage compensation.

Interestingly, the peak of MSL binding at the 3’ end of the dosage compensated genes also coincides with trimethylation of lysine 36 on histone H3 (H3K36me3),\(^7^0,7^1\) a well-established mark for active genes.\(^7^2,7^3\) A crosstalk between H3K36me3 and H4K16ac was suggested based on the observation that down-regulation of Hypb, the enzyme required for the methylation of H3K36me3 in Drosophila, leads to reduced H4K16 acetylation on X-linked genes.\(^7^0\)

MSL-3 binds H3K36me3 nucleosomes in in vitro assays.\(^7^1\) However, recent structural work suggests co-recognition of DNA and a histone H4 tail monomethylated on lysine 20 (H4K20me) by the MSL-3 chromodomain.\(^7^4\) In a parallel study, MSL-3 affinity for H4K20me was confirmed.\(^6^7\) However, MSL-3/nucleic acid interaction was not reported.\(^6^7\) Interestingly, in both reports H3K36me3 peptide was shown not to be the preferred substrate for the MSL-3 chromodomain.\(^7^4,7^5\) How the presence of H3K36me3 on active X-linked genes integrates into the targeting model needs to be further addressed. At the same time, there is
Phosphorylation of histone H3 on serine 10 (H3S10ph) mediated by JIL-1 kinase has also been associated with the MSL complex. JIL-1 and H3S10ph have been shown to be enriched on the male X chromosome. However, JIL-1 null mutants lead to lethality in both sexes suggesting an additional, broader role for this enzymatic activity. The mechanism by which JIL-1-mediated H3S10ph phosphorylation leads to transcriptional defects in interphase cells remains a matter of some controversy. On the one hand, H3S10ph has been shown to facilitate RNA polymerase II release from promoter-proximal pausing in Drosophila. The histone crosstalk between H3S10ph and H4K16ac has also been proposed to generate a histone code that mediates transcription elongation in mammals. The latter is reaffirmed by a recent study showing that the 14-3-3 protein interacts with the Elp3 acetyltransferase (acetylates H3 on lysine 9, H3K9ac), which functions during transcription elongation. However, other studies support a different model in which Pol II-mediated transcription does not require H3S10ph and suggest that the transcriptional defects observed in the absence of JIL-1/H3S10ph are a result of structural alterations of chromatin. Interestingly, JIL-1-mediated ectopic H3S10 phosphorylation is sufficient to induce a change in higher-order chromatin structure from a “heterochromatin-like” state to a more open “euchromatic” state. Furthermore, in the absence of JIL-1, levels of H3K9ac are significantly reduced leading to ectopic spreading of the major heterochromatin markers H3K9me2, HP1α and Su(var)3-7.

Conventional core histones can be replaced by histone variants which have different structural, mobility and stability characteristics, as well as altered susceptibility to modifications, and hence can affect the overall dynamics of chromatin (reviewed in ref. 89). Interestingly, histone H3.3 has shown to be frequently replaced on active genes and is enriched on the dosage compensated male X chromosome.

The primary role of the MSL complex appears to be X chromosomal regulation, but it seems that its function is facilitated by a number of more general factors. These include the DNA supercoiling factor SCF, ISWI, NURF301, HP1, UNR and Su(var)3-7, which display complex genetic interactions with the MSL complex. However, the mechanistic details of these interactions and how they mediate the formation of proper chromatin architecture along the male X chromosome to facilitate dosage compensation are largely unknown. Nevertheless, one emerging theme is the requirement for a permissive chromatin structure to be established and maintained to allow and facilitate the hypertranscription underlying the dosage compensation process. Furthermore, these histone modifications have a great impact on chromatin folding and flexibility, which in turns is essential for the genome architecture and function (reviewed in ref. 97). Further studies are required to elucidate the complex interplay between the various histone modifications and their role for chromatin organization and how they impact dosage compensation.

A recently developed genetic system in Drosophila provides an exciting tool for studying the “histone code” and histone-dependent chromatin assembly in vivo.

Chromosomal Context and Dosage Compensation

In recent years it has become increasingly evident that nuclear three-dimensional structure and genome organization during interphase are of functional importance. It has been shown that gene positioning within the nucleus in relation to other genes and subnuclear compartments contributes to transcriptional control, modulating activity and ensuring maximal expression in some cases or repression in others (reviewed in ref. 100). Since dosage compensation and transcription activation are intrinsically linked, it is not surprising that evidence has started to accumulate indicating a role of nuclear architecture in dosage compensation.

In higher eukaryotes individual chromosomes occupy discrete chromosome territories in the nucleus forming subchromosomal domains of various size. Interestingly, gene-poor or inactive domains reside more internally while active ones tend to localize more peripherally and even loop out of their chromosomal territory and dynamically relocate to specialized subnuclear compartments in association with their activation status. The conventional perception of the nuclear periphery as a site of gene repression has been challenged by findings that physical interactions between actively transcribed genes and the nuclear pore complex (NPC) members can exist (reviewed in ref. 107). It has been proposed that the NPCs could represent platforms for the preassembled transcriptional machineries such as transcription factories. This would help create increased local concentration of transcription factors and enzymatic activities for the coordinated regulation of gene expression. NPCs have been implicated in transcriptional regulation through post-translational modifications of transcriptional factors such as phosphorylation or sumoylation. Facilitation of mRNA processing and export is obviously one of the most classical functions attributed to NPCs (reviewed in ref. 114). NPCs might also have a role in creating boundaries preventing the actively transcribed domains from the invasion of repressive signals from the surrounding peripheral lamina-associated heterochromatin. In addition, NPC components could also establish promoter-end gene loop formation, facilitating RNA Pol II recycling during transcription but also serving as a transcriptional memory allowing rapid reinduction of transcription. At the same time, chromatin structure can also affect nuclear envelope integrity and NPC formation.

Gene-to-pore association in yeast is mediated, at least partly, by DNA zip codes named gene recruitment sequences (GRSs). Whether similar sequences exist and are required for gene targeting to the nuclear periphery in other organisms remains unclear. Adaptor or bridging proteins involved in the interaction between chromatin and the NPC have also been reported including EUN2 and Xmas-2, the Htz1 histone variant, mRNA-export receptors and the CCCTC-binding factor CTCF. A number of nucleoporins have also been reported to be involved in gene-NPC interactions. The fact that many NPC components such as Nup153 and Megator (Mtor) shuttle between...
the nucleoplasm and the nuclear pores and form dynamic filamentous structures protruding into the nucleoplasm, has led to the suggestion that two pools of NPC proteins exist that have different functions depending on their subnuclear localization. Several different Nups have been recently reported to associate with actively transcribed genes including a large number of nucleoplasm-residing genes. However, whether these interactions directly activate transcription or bind as a consequence of activation is still under investigation. It is possible that they may facilitate mRNA export providing a physical route to the periphery of the nucleus. A role for Nups in mediating dynamic looping outside the chromosomal territory and transient associations with the NPCs, has also been suggested.

Intriguingly, the Drosophila MSL complex member MOF has been shown to co-purify with components of the NPC including Nup153 and Mtor. RNAi-mediated knockdown of these components led to reduction of the typical MSL binding pattern on the X, suggesting a role in the dosage compensation process. In support of this idea it was shown that Nup153 and Mtor bind genome-wide in large domains, the so-called nucleoporin associated regions (NARs), which are heavily enriched for active transcription marks. Interestingly, up to 75% of the male X chromosome is enriched in these domains. Taking into consideration the recently demonstrated recruitment of actively expressed genes by a common transcription factor (Klf-1) to “specialized” transcription factories in erythroid cells, it is tempting to speculate that nucleoporins might act in a similar manner. Nucleoporins may mediate the coordinated expression of a number of genes genome-wide, including the extreme case of chromosome-wide upregulation of X-linked genes underlying dosage compensation in Drosophila males. Hyperactivation of the X chromosome seems to correlate with a particular chromosomal conformation where HAS reside in proximity to each other, forming a dosage-compensated chromosomal domain in an MSL-1-MSL-2-dependent manner. A model has been proposed where MSL assembly from the sites of roX transcription in the interior of the chromosomal territory establishes a focal point of enrichment promoting radial spreading and a gradient of decreasing concentration of functional MSL complexes towards the periphery.

While many NARs seem to localize at the nuclear periphery, a subset of them are clearly nucleoplasmic showing that both soluble and peripheral bound pools contribute towards gene expression control. It is possible that transcription control of a subset of genes occurs within nucleoporin-associated domains to allow the coupling of transcription and post-transcriptional events thus facilitating transcriptional output. Future studies will be crucial in understanding how nuclear architecture influences gene expression and how this additional layer of control influences the X chromosome.

**Concluding Remarks**

Classical genetic studies have been instrumental in identifying and characterizing the dosage compensation complex members. Now, recent advances in the fields of biochemical and genomic analyses continue to reveal novel insights into the mechanisms by which dosage compensation is achieved. It is evident that MSL complex members not only take advantage of general factors to help regulate the X chromosome but are also involved in additional functions beyond X chromosomal regulation. Future studies promise to unravel how the MSL complex members balance their various roles and fine-tune gene expression.

**Acknowledgements**

We are grateful to Vera Titschen for helping with the preparation of figures. We also thank Erine Hallaci, Thomas Conrad, Matthew Turley and Juan M. Vaquerizas for critical reading of the manuscript and helpful suggestions.

**References**


