**Drosophila** dosage compensation: a complex voyage to the X chromosome

Marnie E. Gelbart and Mitzi I. Kuroda*

Dosage compensation is the crucial process that equalizes gene expression from the X chromosome between males (XY) and females (XX). In **Drosophila**, the male-specific lethal (MSL) ribonucleoprotein complex mediates dosage compensation by upregulating transcription from the single male X chromosome approximately twofold. A key challenge is to understand how the MSL complex distinguishes the X chromosome from autosomes. Recent studies suggest that this occurs through a multi-step targeting mechanism that involves DNA sequence elements and epigenetic marks associated with transcription. This review will discuss the relative contributions of sequence elements and transcriptional marks to the complete pattern of MSL complex binding.

**Introduction**

The genetic control of sex determination is often associated with dimorphic sex chromosomes (see Glossary, Box 1). In the XY system, females are homogametic (XX), whereas males are heterogametic (XY) (see Glossary, Box 1). Although originally homologous to the X chromosome, the Y chromosome has degenerated over time, creating an imbalance in X-linked gene products between the two sexes. Dosage compensation mechanisms have evolved to equalize X-linked gene expression between males and females, thereby ensuring the appropriate balance of X-chromosomal and autosomal gene products in each sex (Charlesworth, 1996).

Dosage compensation has been best studied in worms, flies and mammals, revealing three distinct strategies for equalizing X chromosome expression (Lucchesi et al., 2005) (Fig. 1). The modulation of X-linked gene expression is essential for viability. In **Drosophila** males, dosage compensation globally upregulates expression from the single X chromosome twofold. In one step, this strategy equalizes X-linked gene expression between males and females and restores the balance between the X chromosome and autosomes (see Glossary, Box 1). By contrast, mammalian dosage compensation, also known as X-inactivation, silences gene expression from one of the two X chromosomes in females (Box 2). Dosage compensation in *C. elegans* also occurs in the homogametic sex (XX animals are hermaphrodites), but acts on both X chromosomes to downregulate gene expression twofold (Box 2). Although these mechanisms equalize X-linked gene expression between the sexes in *C. elegans* and mammals, X chromosome expression would be effectively half that of autosomes if these mechanisms stood alone. Therefore, in *C. elegans* and mammals, X chromosome repression in homogametic animals is thought to be accompanied by a twofold upregulation in the expression of the X chromosome(s) in both sexes, thereby restoring the balance between the X chromosome and autosomes (Gupta et al., 2006; Nguyen and Disteche, 2006). Even though the individual strategies differ, the dosage compensation machineries in these organisms share a central problem: how to distinguish the X chromosome(s) from autosomes? In a more general sense, this represents a fundamental challenge in genomics: how do chromatin regulators recognize their targets within a complex genome?

This review will focus on the dosage compensation machinery in **Drosophila**, known as the male-specific lethal (MSL) complex, and on how this ribonucleoprotein complex recognizes the male X chromosome. The giant polytene chromosomes (see Glossary, Box 1) from larval salivary glands, consisting of up to 1024 copies of each chromosome aligned in register, have provided a powerful system for investigating MSL localization. Recently, high-resolution technologies have enabled the identification of MSL targets at the molecular level. Together, these approaches are beginning to uncover a complex X chromosome targeting mechanism that involves the recognition of DNA elements and transcriptionally active genes. Here, we explore the relative contributions of sequence elements and transcriptional activity to the complete pattern of MSL complex binding, and discuss how the individual RNA and protein components of the complex contribute to this process.

**Dosage compensation in Drosophila: components and models**

The MSL complex, which is essential for male viability, consists of at least five protein subunits and two non-coding RNAs (ncRNAs). These components co-localize to hundreds of discrete sites along the male X chromosome (Lucchesi et al., 2005) (Figs 2 and 3; Table 1).

---

**Box 1. Glossary**

- **Autosomes**: Non-sex chromosomes.
- **Chromocenter**: Densely staining body in **Drosophila** polytene chromosomes where under-replicated heterochromatic regions from all chromosomes converge.
- **Dimorphic sex chromosomes**: Sex chromosomes that have two distinct forms, i.e. the X and the Y.
- **Heterogametic sex**: The sex that has different sex chromosomes, i.e. XY males.
- **Homogametic sex**: The sex that has identical sex chromosomes, i.e. XX females.
- **Polytene chromosomes**: Aligned chromosome fibers from the salivary glands of **Drosophila** larvae that are formed by multiple rounds of DNA replication without cell division. Euchromatic regions are present in as many as 1024 copies, whereas heterochromatin is generally under-replicated.
MSL1 and MSL2 form a core protein complex that targets a subset of sites on the X chromosome. (Fig. 3A,B). However, the protein components MSL3 (Fig. 3C), Males absent on the first (MOF) (Fig. 3D), and Maleless (MLE) (Fig. 3E), as well as the ncRNAs RNA on the X (roX) 1 and 2 (Fig. 3F,G), are required for the full MSL localization pattern on the X chromosome (Palmer et al., 1994; Lyman et al., 1997; Gu et al., 1998; Meller and Rattner, 2002). roX1 and roX2 are encoded on the X chromosome and are functionally redundant for male viability, despite significant differences in sequence and size (Amrein and Axel, 1997; Meller et al., 1997; Meller and Rattner, 2002). MLE and MOF are enzymes: MLE is an RNA/DNA helicase related to human RNA helicase A; MOF [also known as K-acetyltransferase (KAT) 8] is a histone acetyltransferase that specifically modifies histone H4 at lysine 16 (H4K16) (Hilfiker et al., 1997; Lee et al., 1997; Akhtar and Becker, 2000; Smith et al., 2000).

H4K16 acetylation by MOF is linked to transcriptional upregulation. Like other histone acetylation marks, H4K16 acetylation stimulates transcription on chromatin templates (Akhtar and Becker, 2000), and the MSL complex is required for the enrichment of this modification on the Drosophila male X chromosome (Lucchesi et al., 2005). H4K16 acetylation also has special properties: it antagonizes the ISWI family of ATP-dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.

An alternative view is that the MSL complex acts in response to the increased expression of all chromosomes caused by monosomy of the X chromosome (the inverse dosage effect) (Birchler et al., 2003). The central idea of this model is that the MSL complex sequesters dependent chromatin remodeling enzymes and is the only histone modification known to decondense chromatin structure globally (Corona et al., 2002; Shogren-Knaak et al., 2006; Robinson et al., 2008). Therefore, by depositing this histone mark, the MSL complex is thought to enhance transcription along the male X chromosome.
presence of the corresponding transcriptional activator, providing the first evidence that transcription might be involved in attracting the MSL complex to sites along the X chromosome.

**High-resolution analysis of MSL binding on the X chromosome**

The recent molecular identification of MSL targets has provided novel insight into the mechanisms by which the MSL complex recognizes discrete sites along the X chromosome. Chromatin immunoprecipitation to enrich for MSL-associated DNA fragments was coupled with microarrays (a technique known as ChIP-on-chip) to determine the distribution of the MSL complex along chromatin at high resolution. The MSL complex was found to be specifically enriched on the X chromosome in embryos, larval salivary glands, and in two ‘male’ cell lines, consistent with cytology (Alekseyenko et al., 2006; Gilfillan et al., 2006; Legube et al., 2006).

In contrast to many known transcriptional regulators, the MSL complex does not bind to promoters, nor does it generally bind over large domains along the X chromosome. Instead, the MSL complex binds specifically to genes, broadly associating with the middle and 3' ends of active genes on the X chromosome (Alekseyenko et al., 2006; Gilfillan et al., 2006). Even though the mechanism of transcriptional enhancement is currently unknown, the localization of the MSL complex suggests that it acts downstream of initiation, potentially at the level of elongation or of the recycling of RNA polymerase II (RNAP II) back to the promoter for reinitiation (Smith et al., 2001; Alekseyenko et al., 2006; Gilfillan et al., 2006).

---

**Table 1. MSL complex components**

<table>
<thead>
<tr>
<th>Component</th>
<th>Protein/RNA type</th>
<th>Mutant phenotype</th>
<th>Proposed basis for chromosomal phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSL1</td>
<td>Scaffold for MSL complex assembly</td>
<td>Male lethality; no X chromosome recognition</td>
<td>MSL1 and MSL2 are required together for CES recognition; MSL1 is unstable in the absence of MSL2</td>
<td>(Palmer et al., 1994; Lyman et al., 1997; Chang and Kuroda, 1998; Scott et al., 2000; Morales et al., 2004)</td>
</tr>
<tr>
<td>MSL2</td>
<td>RING finger protein</td>
<td>Male lethality; partial MSL complex at CES</td>
<td>MSL2 is the limiting MSL complex component in males</td>
<td>(Palmer et al., 1994; Kelley et al., 1995; Lyman et al., 1997)</td>
</tr>
<tr>
<td>MSL3</td>
<td>Chromodomain protein</td>
<td>Male lethality; partial MSL complex at CES</td>
<td>Chromodomain required for spreading</td>
<td>(Palmer et al., 1994; Buscaino et al., 2006; Sural et al., 2008)</td>
</tr>
<tr>
<td>MOF</td>
<td>H4K16 acetyltransferase</td>
<td>Male lethality; partial MSL complex at CES</td>
<td>Histone acetyltransferase activity required for spreading</td>
<td>(Hilfiker et al., 1997; Gu et al., 1998; Akhtar and Becker, 2000; Gu et al., 2000; Smith et al., 2000)</td>
</tr>
<tr>
<td>MLE</td>
<td>RNA/DNA helicase; prefers dsRNA or RNA/DNA hybrid substrates</td>
<td>Helicase activity required for spreading</td>
<td>(Palmer et al., 1994; Lee et al., 1997; Gu et al., 2000; Izzo et al., 2008; Morra et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>roX1</td>
<td>Non-coding RNAs; encoded on the X chromosome</td>
<td>Male lethality in double mutant; MSL complex relocalized to chromocenter</td>
<td>Internal redundancy within each RNA contributes to X chromosome specificity and spreading</td>
<td>(Meller and Rattner, 2002; Stuckenholz et al., 2003; Park et al., 2007; Kelley et al., 2008; Park et al., 2008)</td>
</tr>
<tr>
<td>roX2</td>
<td>Non-coding RNAs; encoded on the X chromosome</td>
<td>Male lethality in double mutant; MSL complex relocalized to chromocenter</td>
<td>Internal redundancy within each RNA contributes to X chromosome specificity and spreading</td>
<td>(Meller and Rattner, 2002; Stuckenholz et al., 2003; Park et al., 2007; Kelley et al., 2008; Park et al., 2008)</td>
</tr>
</tbody>
</table>

CES, chromatin entry site; dsRNA, double-stranded RNA.
Although all ChIP-on-chip studies published to date report preferential MSL binding to active genes on the X chromosome, the extent of this correlation varies between studies, and two different interpretations have been put forward to explain how this localization is achieved. The sequence model posits that DNA elements play a gene-specific role in recruiting the MSL complex. In two studies, sequence motifs were derived from the MSL localization pattern, although these were not predictive of the full pattern of MSL binding (Gilfillan et al., 2006; Legube et al., 2006). On the basis of a third study, in which no sequence elements specific to MSL binding sites were identified, the transcription model was proposed instead, which postulates that the MSL complex localizes along the X chromosome by recognizing features of active genes (Alekseyenko et al., 2006).

The sequence and transcription models are not mutually exclusive. As described below, there is a general consensus that sequence-dependent high-affinity sites promote the distinction between the X chromosome and autosomes. Following the recognition of the X chromosome, the MSL complex may spread to active genes on the X chromosome by recognizing marks of transcription and possibly additional DNA elements. The roles of sequence elements and transcriptional marks in targeting the MSL complex to the X chromosome and in its spreading along this chromosome are considered in turn.

### Targeting the MSL complex to the X chromosome
For over a decade, it has been known that partial MSL complexes can recognize a subset of sites along the X chromosome. No binding to the X chromosome is detected in msl1 or msl2 mutant backgrounds. In the absence of msl3, mle or mof, however, MSL1 and MSL2 co-localize to a reproducible set of ~35-70 sites on the X chromosome (Palmer et al., 1994; Lyman et al., 1997; Gu et al., 1998) (Fig. 2B). A similar set of sites is observed when the availability of the MSL complex is limited, which suggests that...
they constitute high-affinity MSL binding sites (Kelley et al., 1997; Demakova et al., 2003). Termined chromatin entry sites (CESs), these were postulated to be the locations of MSL complex assembly and/or nucleation sites that enable the MSL complex to access the X chromosome (Lyman et al., 1997; Kelley et al., 1999). Interestingly, the roX1 and roX2 loci were the first of these sites to be identified, based on their ability to recruit the MSL complex to an autosomal insertion in an msl3 mutant background (Kelley et al., 1999).

**MSL complex assembly along nascent roX RNAs**

The fact that the ncRNAs that are associated with the MSL complex are themselves encoded on the X chromosome suggests that the roX loci play a role in X chromosome recognition (Fig. 4A). One model is that the roX RNAs are incorporated into the MSL complex as they are synthesized. This idea of co-transcriptional MSL complex assembly at roX genes is supported by autosomal roX transgenes that nucleate ectopic bidirectional spreading of the MSL complex over 1 Mb flanking their site of insertion (Kelley et al., 1999). The frequency and the extent of local spreading are highly sensitive to MSL1 and MSL2 levels, to the number of roX loci in the genome and to the level of roX transcription (Park et al., 2002; Demakova et al., 2003; Oh et al., 2003; Kelley et al., 2008).

However, roX RNAs can integrate into the MSL complex and facilitate the recognition of the X chromosome even when encoded on an autosome: autosomal roX transgenes can rescue the viability and MSL targeting defects of males that lack roX RNAs (roX−) (Meller and Rattner, 2002). High-level expression of an autosomal roX transgene favors movement of the MSL complex to the X chromosome, whereas low-level expression favors local spreading flanking the insertion site (Kelley et al., 2008). It has been proposed that co-transcriptional assembly of a complete MSL complex favors local retention of the complex, whereas partial complexes that diffuse away complete their assembly in the nucleoplasm and are then recruited to the X chromosome (Park et al., 2002; Oh et al., 2003; Kelley et al., 2008).

Even though the presence of roX genes on the X chromosome is not necessary for MSL localization, we suspect that their location is not a coincidence. roX genes on the X chromosome may influence the efficiency of MSL targeting by assembling the complex co-transcriptionally (Deng and Meller, 2008). Alternatively, the location of the roX RNAs on the X chromosome might have played a more crucial role early in the evolution of *Drosophila* dosage compensation, before the acquisition of a sufficient number of DNA targeting sequence elements (Box 3).

**Sequence-dependent MSL targeting to the X chromosome**

The ability of the MSL complex to recognize the roX X chromosome when roX RNAs are encoded on an autosome implies that additional mechanisms exist to identify the X chromosome. Fine mapping of the roX1 and roX2 CESs has identified short fragments (<250 bp) that are sufficient for recruiting partial MSL complexes in an autosomal context (Kageyama et al., 2001; Park et al., 2003). These fragments are thought to function as DNA elements that recruit the MSL complex, although it is unknown whether the MSL complex recognizes these sequences directly or if an accessory factor is required. Conserved sequence elements required for CES function have been identified (Park et al., 2003), and a handful of other CESs have been mapped (Oh et al., 2004; Dahlsveen et al., 2006; Gilfillan et al., 2007), but a genomic approach was necessary to derive unifying principles that characterize CESs (Alekseyenko et al., 2008; Straub et al., 2008). ChIP-on-chip of MSL2 in msl3

---

**Fig. 4. Two-step model for MSL complex targeting.** (A) The MSL complex specifically recognizes the X chromosome (X), not autosomes (Au), by co-transcriptional assembly at roX genes (blue) and by recognition of MSL recognition elements (MREs) at chromatin entry sites (CESs, red). Dashed lines indicate the possibility that roX RNAs are assembled into the MSL complex before the complex targets other CESs. (B) Spreading from these sites requires H3K36me3 (stars) and perhaps other features of active genes. It has been demonstrated that the MSL complex can recognize active genes (tan) that lack X chromosome-specific sequences. However, there is evidence that MSL binding might be reinforced in some cases by the presence of additional sequence elements (green bars). Dashed lines indicate hypothetical paths of MSL spreading, although other paths are possible. Spreading might involve a linear scanning mechanism, or it may reflect a capture-and-release mechanism in three-dimensional space. In either case, the MSL complex is able to ‘skip over’ inactive genes (gray).
Box 3. Evolution of dosage compensation in Drosophila

How do dosage compensation systems evolve to compensate for the progressive loss of genes from the Y chromosome? The targets of the male-specific lethal (MSL) complex on the X chromosome are thought to be acquired gene-by-gene or block-by-block (Bachtrog, 2006). An intermediate stage of this process is observed in Drosophila miranda, in which only some regions along the neo-X/chromosome have acquired MSL binding and dosage compensation (Bone and Kuroda, 1996; Marin et al., 1996). It will be interesting to compare the sequence composition of the neo-X/chromosome in regions that do and do not undergo dosage compensation. Furthermore, identification of MSL targets at high resolution on the neo-X/chromosome should provide key insights into chromatin entry site (CES) function and into the evolution of spreading on the X chromosome.

What about the evolution of MSL proteins? MSL3, Maleless (MLE) and Males absent on the first (MOF), which are required for spreading and chromatin modification in Drosophila, are ancient proteins (Marin and Baker, 2000; Sanjuan and Marin, 2001). However, the key initial targeting subunits, MSL1 and MSL2, are less conserved (Marin, 2003), suggesting that targeting has evolved to serve diverse purposes. Human MSL complexes composed of MSL1, MSL2, MSL3 and MOF (MYST1) have been identified, but their downstream targets are not known (Smith et al., 2005; Mendjan et al., 2006).

Given that dosage compensation is essential for the viability of Drosophila males, one would predict that the MSL complex is subject to purifying selection in Drosophilids. However, two recent studies have demonstrated surprising evidence for adaptive evolution in MSL protein-coding genes in the Drosophila melanogaster lineage and have proposed the co-evolution of MSL binding sites, an idea supported by the recent analysis of three CESs (Levine et al., 2007; Rodriguez et al., 2007; Bachtrog, 2008). Altered specificity of the MSL complex has been hypothesized to play a role in male hybrid inviability in Drosophila (Orr, 1989).

mutant embryos identified a set of 150 CESs distributed along the X chromosome (Alekseyenko et al., 2008). ChIP of the intact MSL complex from ‘male’ Clone8 cells followed by Solexa sequencing of the enriched DNA fragments (ChIP-seq) revealed that the tallest peaks of MSL enrichment overlap strikingly with CESs, suggesting that CESs are bona fide high-affinity MSL binding sites. Motif searches revealed a 21 bp GA-rich (or TC-rich) motif, named the MSL recognition element (MRE) (Alekseyenko et al., 2008), which includes sequence motifs found in previously mapped CESs (Park et al., 2003; Gilfillan et al., 2007). At least one copy of the MRE is present in 91% of CESs (Alekseyenko et al., 2008). Importantly, in functional assays, MRE mutations abolish MSL recruitment, whereas scrambling the surrounding sequences has no effect. Therefore, MREs are both necessary and sufficient for MSL recruitment in the contexts that have been tested, demonstrating that they play a key role in MSL recognition of the X chromosome.

Independently, 130 high-affinity sites (HASs) were mapped in the ‘male’ SL2 cell line by ChIP-on-chip of MSL1 and MSL2 following the depletion of msl3, mof or mle by RNAi, or under mild cross-linking conditions (Straub et al., 2008). Strikingly, 69% of HASs overlap with the CES identified in msl3 mutant embryos, and HASs are enriched for sequence motifs, including a GA-rich element highly similar to the MRE. These high-resolution mapping studies collectively identify many more CESs/HASs than have been predicted previously from cytology (Fig. 2B), and these are probably still underestimates. We now predict that as many as 240–300 of these sites are located on the X chromosome (Alekseyenko et al., 2008; Sural et al., 2008), perhaps accounting for the incomplete overlap between the datasets.

The identification of a functional DNA sequence element at high-affinity MSL binding sites is a significant advance towards understanding how the MSL complex distinguishes the X chromosome from autosomes (Fig. 4A). However, these motifs are only ~twofold enriched on the X chromosome as compared with autosomes, and many motifs on the X chromosome are not associated with a CES/HAS (Alekseyenko et al., 2008; Straub et al., 2008). Therefore, additional features, such as nucleosome occupancy, are likely to distinguish the MREs that are recognized in vivo. Indeed, CESs/HASs are characterized by histone depletion, in contrast to unoccupied MREs on the X chromosome (Alekseyenko et al., 2008; Straub et al., 2008). Furthermore, CESs/HASs appear to be generally associated with active chromatin domains, thereby recruiting the MSL complex to regions of the X chromosome enriched for its preferred targets – transcribed genes. Restricting MRE searches to active genes, characterized by histone H3 lysine 36 trimethylation (H3K36me3), increases the proportion of MREs found in CESs and enriches MREs on the X chromosome fourfold over autosomes (Alekseyenko et al., 2008). The clustering of MREs in three-dimensional space and the binding of accessory proteins might also influence which MREs are recognized. Further experiments are required to determine the relative contributions of these factors in allowing the MSL complex to utilize MREs to distinguish the X chromosome from autosomes.

Spreading along the X chromosome

High-affinity MSL binding sites probably play a key role not only in distinguishing the X chromosome from autosomes, but also in the spreading of the MSL complex along this chromosome. The idea of spreading is generally associated with the assembly of silencing factors along chromatin to form repressive heterochromatin domains. The term ‘spreading’ is often used to describe the linear propagation of proteins along chromatin, as shown for the Sir proteins in S. cerevisiae, but spreading may also skip regions along the chromatin fiber (Talbert and Henikoff, 2006). In the case of the MSL complex, spreading results in the discontinuous distribution of the complex along the X chromosome, with a bias for the bodies of active genes (Fig. 4B). It is unknown whether the spreading of the MSL complex is a linear process, whereby the complex scans along chromatin and is only stabilized at active genes. Alternatively, the MSL complex may spread via a capture-and-release mechanism, whereby it samples nearby chromatin in three-dimensional space. In either scenario, the recruitment of the MSL complex to CESs generates high local concentrations of this complex along the X chromosome that drive spreading to nearby sites of lower affinity. The identification of at least 150 CESs on the X chromosome and their locations within active gene clusters suggest that spreading might be a local phenomenon (Alekseyenko et al., 2008; Straub et al., 2008). Indeed, the majority of MSL complex bound on the X chromosome is within 5–10 kb of a CES (Sural et al., 2008). The roX genes might be unique in their nucleation of long-range spreading owing to co-transcriptional assembly of the MSL complex at roX loci.

The mechanism for recognizing active genes may involve sequence elements, features of transcription, or both. At one extreme, the strict sequence or ‘affinities’ model postulates that CESs are not qualitatively different from other MSL targets. They
represent the highest affinity MSL binding sites, whereas other sites have sequences that bind to the MSL complex at lower affinities (Demakova et al., 2003; Fagegaltier and Baker, 2004; Dahlsveen et al., 2006; Legube et al., 2006). At the other extreme, CESs are considered to be distinct from other MSL binding sites because they contain DNA elements that specify the X chromosome. The transcription model proposes that after the MSL complex identifies the X chromosome, it recognizes active genes by properties of transcription, such as histone modifications or by nascent transcripts themselves (Alekseyenko et al., 2006; Larschan et al., 2007). These two models are explored below.

**The affinities model: a continuum of sequences along the X chromosome**

The idea that CESs are sites of MSL entry into the X chromosome was originally called into question because several regions of the X chromosome that lack cytologically mapped CESs are competent to recruit the MSL complex in an autosomal context (Fagegaltier and Baker, 2004; Oh et al., 2004; Dahlsveen et al., 2006). To explain this observation, it was proposed that these regions contain multiple low-affinity MSL binding sites that act together to recruit the MSL complex. Such sequences are expected to be degenerate and, therefore, have evaded molecular definition. Given that recent evidence suggests that there are more CESs than previously thought (Alekseyenko et al., 2008; Straub et al., 2008), it is possible that some of these X-derived fragments might actually contain a CES.

However, this is not always the case. Two X chromosome-derived transgenes that do not appear to function as CESs recruit the MSL complex in a transcription-dependent manner when inserted onto an autosome (Kind and Akhtar, 2007). A 339 bp fragment from one of these genes is capable of recruiting the MSL complex in the absence of transcription, but only when multimerized. Therefore, it was proposed that transcription increases the accessibility of X chromosome-specific sequences within genes and is no longer necessary when multiple copies of the sequence are present. This result supports the affinities model, but the nature of the sequence element has not yet been defined.

A requirement for X chromosome-specific sequence motifs in MSL spreading has also been indicated by the observation that the MSL complex binds in a wild-type pattern to X chromosome sequences translocated onto an autosome, but does not spread visibly to adjacent autosomal sequences (Fagegaltier and Baker, 2004; Oh et al., 2004). Furthermore, the MSL complex does not spread visibly to autosomal sequences translocated onto the X chromosome, indicating that autosomes lack sequences that promote MSL complex spreading. As discussed above, spreading is potentially a short-range phenomenon, indicating that high-resolution analyses will be required to assay for local spreading that might have been missed by cytology.

Evidence for a hierarchy of MSL binding sites on the X chromosome was demonstrated by changes in the number of MSL-positive sites on polytene chromosomes in response to alterations in the levels of the MSL complex (Demakova et al., 2003). Given the new estimates of CES density on the X chromosome, perhaps these variations reflect a range of CES affinities. Whereas only a small number of variable, weak autosomal sites are detectable in wild-type males, reproducible sites of autosomal binding are observed when the MSL complex is overexpressed (Demakova et al., 2003). Therefore, MSL complex levels must be carefully modulated to avoid autosomal targeting. It would be interesting to identify the autosomal binding sites molecularly to determine whether they resemble MREs.

Together, these results raise the question of how many sequences with affinity for the MSL complex exist along the X chromosome. Until these elements are better defined, this will remain an open question.

**The transcription model: recognition of active genes**

The localization of the MSL complex to active genes along the X chromosome suggests that the MSL complex might spread from CESs to sites of active transcription. A comparison of MSL binding in two cell lines revealed rare examples of differential binding (Alekseyenko et al., 2006). In these cases, the genes were also differentially expressed and binding correlated with gene activity, suggesting a causal link between transcription and MSL recruitment.

Additional evidence indicates that X chromosome-specific sequences are not necessary to recruit the MSL complex to active genes. The high-resolution analysis of ectopic spreading of the MSL complex to autosomal sequences that surround a roX transgene has revealed that a strong correlation exists between spreading and gene activity (Larschan et al., 2007). Even though roX genes are unique in their ability to support long-range spreading, CES insertions on an autosome have recently been shown to nucleate short-range spreading to the 3’ end of an active gene ~10 kb away (Alekseyenko et al., 2008).

The localization of the MSL complex to active genes on the X chromosome, or on autosomes when ectopically recruited, indicates that the MSL complex recognizes features of transcribed genes. Further support for this idea came from the observation that the distribution of the MSL complex along the X chromosome strongly resembles the pattern of H3K36me3, a histone modification associated with active genes on all chromosomes (Larschan et al., 2007). SET2 [also known as K-methyltransferase (KMT) 3] is the enzyme responsible for H3K36me3 and is required for robust MSL recruitment to target genes, but not to CESs, thus distinguishing sequence-dependent from H3K36me3-dependent modes of recruitment (Larschan et al., 2007; Bell et al., 2008). These findings clearly demonstrate that sequence elements are not sufficient to recruit the MSL complex to active genes and that H3K36me3, and perhaps other features of transcription, are also involved (Fig. 4B).

The recruitment of the MSL complex to active genes prevents the inappropriate activation of repressed genes, and may explain why autosomal genes can be dosage compensated when inserted onto the X chromosome (Scholnick et al., 1983; Spradling and Rubin, 1983). One advantage of this mechanism is that any transcribed gene on the X chromosome would be dosage compensated without the need to evolve a specific sequence to recruit the complex. However, even though experimental evidence argues against a strict sequence model for MSL spreading, it is possible that genes on the X chromosome have evolved sequence elements that stabilize MSL complex binding after recruitment (Fig. 4B).

**Exceptions to the rule**

One prediction of a strict transcription model is that the MSL complex binds to all active genes on the X chromosome. However, differences in MSL complex and RNAP II localization have been observed in cytological and high-resolution analyses (Giffiian et al., 2006; Kotlikova et al., 2006; Legube et al., 2006). One potential explanation for this is that the imperfect correlation between MSL binding and transcription results from the application of somewhat arbitrary thresholds. Using H3K36me3 as a proxy for transcriptional activity, an unbiased computational analysis of the data from a ‘male’ cell line estimates that nearly 80% of active genes on the X chromosome are clearly bound by
the MSL complex, whereas less than 1% are clearly unbound (Larschan et al., 2007) (S. Peng, P. J. Park and M.I.K., unpublished). The remaining genes are in an intermediate category that cannot be unambiguously defined.

Although the precise numbers are undefined, a set of transcribed genes does not recruit the MSL complex. Why are these genes skipped? Are they dosage compensated by long-range effects of MSL action, by an MSL-independent mechanism, or do they escape dosage compensation altogether? Nearly all transcribed genes on the X chromosome are dosage compensated (Gupta et al., 2006). The best-studied exception is Lsp1α, a gene expressed in the Drosophila larval fat body (the predominant immune-responsive tissue) that does not appear to recruit the MSL complex in this tissue and that escapes dosage compensation (Weake and Scott, 2007). However, an Lsp1α promoter-lacZ transgene can recruit the MSL-associated H4K16 acetylation mark in other locations on the X chromosome. Why does the endogenous Lsp1α gene fail to recruit the MSL complex? There is no evidence to suggest a repressive chromatin environment or the presence of insulators that block MSL recruitment. It has been proposed that Lsp1α lacks a sequence element necessary to recruit the MSL complex in its endogenous location. Alternatively, it is conceivable that the developmentally regulated expression of the endogenous Lsp1α gene lacks features that are required to recruit the MSL complex. For example, it is possible that some active genes are not associated with H3K36me3. In this scenario, the altered regulation of Lsp1α transgenes might account for their ability to recruit the complex. Genes like Lsp1α will need to be evaluated on a case-by-case basis to investigate the complexities of MSL recruitment.

The nature of spreading: the dynamics of MSL localization

What is the nature of MSL spreading from high-affinity MSL binding sites to the remainder of the X chromosome? As discussed above, recent evidence suggests that spreading might be a short-range phenomenon. It will be interesting to learn whether the loss of an individual CES has a local effect on MSL targeting to nearby genes. Alternatively, other CESs and lower-affinity MSL sequences might compensate for its absence. At the cytological level, removing the roX2 CES has no detectable effect on MSL localization (Meller and Rattner, 2002).

If the MSL complex generally recognizes transcribed genes, how is spreading restricted to the X chromosome? A tight regulation of MSL levels is likely to be involved (Demakova et al., 2003). Once a local concentration of the MSL complex is established on the X chromosome, how is spreading to nearby autosomal regions prevented? One possibility is that the MSL complex is confined to a subnuclear domain that prevents spreading to other chromosomes. The nuclear pore component NUP153 and the associated protein MTO have been implicated in MSL localization to the discrete subnuclear territory of the X chromosome (Mendjan et al., 2006). An investigation into the spatial organization of the X chromosome, particularly of CESs, might provide insights into the link between the MSL complex and nuclear architecture.

Little is known about the dynamics of MSL binding to the X chromosome. The MSL complex remains associated with the X chromosome throughout mitosis, suggesting that its localization on the X chromosome might need to be established only once during development (Lavender et al., 1994; Straub et al., 2005b). Photobleaching experiments indicate that the association of MSL2 with the X chromosome territory is surprisingly stable, suggesting little MSL turnover on the X chromosome during interphase (Straub et al., 2005b). This experiment is limited to a period of several minutes, but it raises the question of whether the MSL binding pattern can change over time.

The transcription model predicts that the MSL complex can be recruited to a newly activated gene, although this has not yet been demonstrated. The induction of an autosomal roX gene by the Gal4 activator causes a shift from ectopic local spreading on the autosomal to X chromosome targeting within a few hours (Kelley et al., 2008). This result suggests that recognition of the X chromosome might be possible late in development, although a small amount of the MSL complex may still persist on autosomes prior to roX induction. In the same experiment, the loss of the MSL complex along the autosome indicates a high turnover at these sites. Does this reflect dynamic MSL binding on the X chromosome? Alternatively, is there a qualitative difference in MSL binding on the X chromosome and on autosomes, whereby sequence elements serve to reinforce MSL binding to genes on the X chromosome and lead to lower turnover?

X chromosome targeting machinery

The results discussed above support a model in which the MSL complex first identifies the X chromosome by co-transcriptional assembly at roX genes and by recognition of DNA elements within CESs (Fig. 4A), and then spreads to active genes by recognition of a transcription-associated histone modification and possibly additional sequence elements (Fig. 4B). How does the MSL complex achieve this multi-step mechanism of targeting along the X chromosome? MSL1 serves as a platform for MSL complex assembly as it independently binds MSL2, MSL3 and MOF (Scott et al., 2000; Morales et al., 2004) (Fig. 3; Table 1). Despite its crucial role in MSL spreading, MLE appears to be only weakly associated with the MSL complex, to the extent that it has been reported to be absent from several biochemical MSL complex purifications (Copps et al., 1998; Smith et al., 2000; Mendjan et al., 2006). MLE may interact weakly with the complex via the C-terminal domain of MSL2 or through roX RNAs, as MLE, MOF and MSL3 are all implicated in RNA binding (Richter et al., 1996; Akhtar et al., 2000; Buscaino et al., 2003; Li et al., 2008). Perhaps chromatin association stabilizes these interactions, as MSL components co-localize along the polytene male X chromosome. In order to fully understand the structural organization of the MSL complex, and in particular how the roX RNAs are incorporated into the complex, biochemical order-of-assembly experiments are required.

The cytological analysis of partial MSL complexes has provided early evidence for the contributions of the different subunits to MSL complex localization. MSL1 and MSL2 are required for the association of the MSL complex with CESs, whereas the addition of MSL3, MLE and MOF is required for its spreading to the full MSL pattern (Palmer et al., 1994; Lyman et al., 1997; Gu et al., 1998). The roX RNAs also play a pivotal role in MSL targeting (Franke and Baker, 1999; Meller and Rattner, 2002). In roX males, MSL proteins localize to a small number of sites on the X chromosome that may be related to CESs; much of the complex, however, is relocalized to the chromocenter (see Glossary, Box 1), the clustered mass of under-replicated heterochromatin from all chromosomes in Drosophila salivary glands (Meller and Rattner, 2002; Deng et al., 2005). It is intriguing to speculate that, in the absence of roX RNAs, the MSL complex is mistargeted by an aberrant association with RNAs that are derived from repetitive elements in heterochromatin.
The role of MSL proteins in X chromosome recognition and spreading

MLS1 and MSL2 are expected to be the only MSL proteins required to recognize CESs and the associated MREs. However, they lack any obvious DNA-binding domains, suggesting that MRE recognition might be mediated by other factors. The GA-rich core of the MRE suggests that the GAGA factor (GAF, encoded by Trl) might be involved (Alekseyenko et al., 2008). To date, there is little evidence to support a general role for GAF in MSL recruitment (Sun et al., 2003; van Steensel et al., 2003), although Trl mutations subtly affect MSL binding on polytene chromosomes (Greenberg et al., 2004). Perhaps GAF and other GAGA-binding proteins function redundantly to recruit the MSL complex to MREs. Additional experiments are required to determine whether partial MSL complexes are capable of sequence-specific binding to MREs or to identify factors that mediate this interaction.

An N-terminal basic region of MSL1 (amino acids 1-26) is implicated in CES recognition, but is dispensable for interactions with other MSL proteins (Li et al., 2005). This same region is necessary for the chromocenter localization of MSL complexes that form in the absence of the C-terminal domain of MSL2, which is required for the efficient incorporation of roX RNAs into the complex (Li et al., 2008). Therefore, the MSL1 N-terminal region is likely to mediate common interactions that underlie localization to these distinct targets. The adjacent region (amino acids 26-84) is required for the self-association of MSL1, which might promote the cooperative binding of the MSL complex to clustered binding sites (Li et al., 2005).

How do MSL3, MOF and MLE promote spreading along the X chromosome? MSL3 contains an N-terminal chromodomain, which is required for preferential interaction with H3K36-trimethylated nucleosomes in vitro (Sural et al., 2008). Mutations in the chromodomain disrupt spreading of the MSL complex beyond ~1 kb from CESs in vivo, suggesting that MSL3 promotes spreading through the recognition of H3K36me3. Defects in spreading associated with chromodomain mutants were not detectable by cytology, demonstrating the importance of high-resolution approaches.

The enzymatic activities of MOF and MLE are also required for spreading (Lee et al., 1997; Gu et al., 2000). One attractive model is that H4K16 acetylation by MOF serves as a docking site that facilitates spreading (Lee et al., 1997; Gu et al., 2000). One attractive model is that H4K16 acetylation by MOF serves as a docking site that facilitates spreading (Lee et al., 1997; Gu et al., 2000).

Table 2. Accessory factors implicated in MSL targeting

<table>
<thead>
<tr>
<th>Factor</th>
<th>Protein type</th>
<th>Relevant mutant/RNAi phenotype</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>SET2</td>
<td>H3K36 methyltransferase</td>
<td>Mutation/RNAi reduces MSL binding at active genes on X chromosome, not CES (ChiP of third instar larvae, SL2 cells)</td>
<td>(Larschan et al., 2007; Bell et al., 2008)</td>
</tr>
<tr>
<td>GAF (TRL)</td>
<td>GAGA-binding factor</td>
<td>Mutation is associated with an increased number of MSL sites on autosomes and the loss of a single CES (immunostaining of polytene chromosomes)</td>
<td>(Greenberg et al., 2004)</td>
</tr>
<tr>
<td>UNR</td>
<td>RNA-binding protein</td>
<td>Mutation/RNAi decreases MSL staining on X chromosome (immunostaining of polytene, SL2 cells)</td>
<td>(Patalano et al., 2009)</td>
</tr>
<tr>
<td>NUP153</td>
<td>Nuclear pore protein</td>
<td>RNAi depletion disrupts MSL localization to the X chromosome (immunostaining of SL2 cells)</td>
<td>(Mendjan et al., 2006)</td>
</tr>
<tr>
<td>MTOR</td>
<td>Nuclear pore-associated protein</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CES, chromatin entry site; ChiP, chromatin immunoprecipitation; RNAi, RNA interference.
Accessory factors involved in MSL targeting

The MSL complex requires additional factors for its targeting to the X chromosome (Table 2). As discussed above, SET2 is required for the MSL complex to efficiently target active genes on the X chromosome (Larschan et al., 2007; Bell et al., 2008), and mutations in the gene that encodes GAF, Trl, subtly alter MSL localization along the polytene X chromosome (Greenberg et al., 2004). In addition, Upstream of N-ras (UNR), an RNA-binding protein, is required for MSL localization, potentially through its association with the roX RNAs (Patalano et al., 2009). Several other proteins have been linked to the MSL complex through biochemical purifications, including the JIL-1 histone H3 serine 10 kinase, nuclear pore components, interband-associated proteins, and components of the nuclear exosome (Jin et al., 2000; Mendjan et al., 2006). Of these, only the nuclear pore protein NUP153 and the associated protein MTOR have been implicated in MSL targeting thus far (Mendjan et al., 2006). These interactions suggest that the compartmentalization of the X chromosome might promote the targeting and/or spreading of the MSL complex.

Other genes, including Jil-1 and supercoiling factor (scf), have been implicated in dosage compensation based on their mutant phenotypes of preferential male lethality and of abnormal X chromosome morphology (Deuring et al., 2000; Wang et al., 2001; Ebert et al., 2004; Liu et al., 2005; Spierer et al., 2005; Furuhashi et al., 2006; Zhang et al., 2006). Although these factors are not known to affect MSL targeting, their interplay with the MSL complex remains to be determined and might shed further light onto the mechanism of dosage compensation.

In the future, the identification of novel factors in MSL targeting might require additional screening approaches. An RNAi screen for factors required for MSL targeting to the X chromosome territory in ‘male’ SL2 cells has revealed known MSL proteins and a regulator of mle splicing (Worringer and Panning, 2007). Recently, a roX2 reporter that exhibits MSL-dependent transcriptional enhancement was utilized to test for factors required for CES recruitment and dosage compensation in cell culture (Yokoyama et al., 2007). This type of strategy should be useful in order to identify novel components involved in MSL targeting and function.

Conclusions

Significant progress has been made in recent years towards understanding the mechanism by which the MSL complex recognizes the X chromosome. The utilization of high-resolution approaches has contributed greatly to these advances. The MSL complex first identifies the X chromosome by co-transcriptional assembly on roX RNAs and by sequence-specific recognition of CESs. We favor the idea that subsequent spreading of the MSL complex to active genes is largely mediated by the recognition of transcriptional marks. However, many questions still remain about how the MSL complex identifies the X chromosome and how it spreads to its active gene targets. How does the MSL complex specifically recognize MREs on the X chromosome? Are there other sequence elements that facilitate MSL recruitment or perhaps even repel the complex? Aside from H3K36me3, are there additional features of transcribed genes that the MSL complex recognizes? How do the components of the MSL complex, particularly the roX RNAs, contribute to sequence-specific binding and to the identification of active genes? What are the accessory factors that facilitate these events? How dynamic is MSL binding on the X chromosome? Can MSL recruitment be induced in response to changes in the transcriptional state of a gene and, if so, is MSL loading restricted to specific time-points, such as DNA replication?

It is critically important to continue to apply recently developed technologies to address these problems. For example, high-resolution localization analysis has enabled the detection of short-range MSL spreading along an autosome (Aleksyenko et al., 2008). Coupling high-resolution approaches to the analysis of MSL mutants can provide novel insights into the contributions of individual MSL components to the targeting of the complex. Furthermore, site-specific integration methodologies enable the comparison of mutant MSL components expressed from the same locus (Kelley et al., 2008; Sural et al., 2008). Eliminating the confounding effects of the chromatin environment is also useful for comparing the recruitment potential of X chromosome-derived fragments on the autosomes (Aleksyenko et al., 2008). In this type of experiment, it is important to remember the caveat that pairing between the autosomal insertion and the endogenous locus on the X chromosome may facilitate MSL transmission to the transgene (Kelley et al., 1999). Even though the ectopic recruitment of the MSL complex to autosomes is a stringent test of the recruitment ability of X chromosome-derived DNA fragments, high-resolution analyses are also important to investigate the behavior of autosomal or heterologous sequences in the X-chromosomal environment. Perhaps we can utilize these tools not only to describe the steady-state MSL pattern, but also to observe the establishment of MSL binding along the X chromosome over time at the molecular level. These types of experiments will trigger the next wave of discoveries regarding how the ribonucleoprotein MSL complex recognizes a complex landscape of DNA sequence and epigenetic marks to target and regulate a chromosome.

References


