# Chromodomain-mediated spreading on active genes

Alison M Hosey & Marjorie Brand

**The formation of heterochromatin involves spreading of repressor proteins along large chromosomal domains. A new study reveals that the concept of spreading also holds true for establishing domains of active chromatin. More specifically, spreading of the *Drosophila melanogaster* male-specific lethal (MSL) activator complex, which is required for dosage compensation on the X chromosome, involves interaction between the MSL3 chromodomain and histone H3 methylated at lysine 36.**



Post-translational modifications of histones are an integral part of gene regulation[1]. One role for these histone marks is to stabilize the binding of effector proteins, which in turn directly or indirectly influence gene expression by establishing additional marks, modifying chromatin structure or eventually recruiting the components of the transcription machinery. Although the initial binding occurs at defined sites and leads to local effects, it is followed at some loci by large-scale (>10 kb) spreading of chromatin-associated proteins to propagate the message to neighboring genes. Even though the exact spreading mechanism is unclear and can vary between genomic locations, this process is widely accepted as a way to generate large repressive chromatin domains[2]. A well-known case of spreading-mediated silencing is the formation of heterochromatin via specific binding of the chromodomain-containing heterochromatic protein HP1 to histone H3 methylated at lysine 9. Because HP1 also interacts directly with the H3K9 methyltransferase Su(var)3-9, the repressive signal can propagate along large chromosomal domains[3,4]. In contrast, the role of spreading in creating large domains of active chromatin had remained a contentious issue[5].

The process of dosage compensation establishes equal X-linked gene expression between males (XY) and females (XX), and it occurs by differing mechanisms in flies, worms and mammals (**Fig. 1**). In *D. melanogaster*, the activator ribonucleoprotein complex MSL (**Fig. 2**) doubles the transcriptional output from the single male X chromosome to achieve dosage compensation[6]. In a new study published in the December 2008 issue of *Nature Structural & Molecular Biology*[7], the Kuroda group demonstrates that the chromodomain of the MSL3 subunit, previously deemed dispensable for MSL complex binding to the X chromosome[8], is necessary for bidirectional spreading of the dosage-compensation complex from the chromatin entry sites (CESs) to flanking regions on the X chromosome (**Fig. 3**). Using high-resolution chromatin immunoprecipitation combined with microarray analysis (ChIP-chip), the authors observed that when the MSL3 chromodomain is mutated the dosage-compensation complex is retained at the CES and is unable to spread to neighboring active genes. Notably, this sequence-independent spreading mechanism involves a specific association of the MSL3 chromodomain with trimethylated histone H3 lysine 36 (H3K36me3), a chromatin mark that is generated during transcriptional elongation[9]. This provides important evidence that specific binding to an active histone mark is involved in bidirectional spreading of an activator complex over large (10–20 kb) chromatin domains.

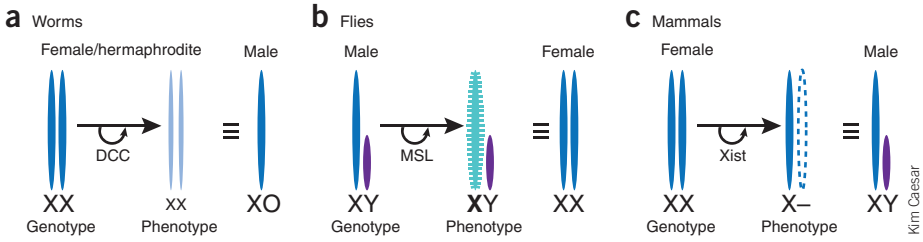
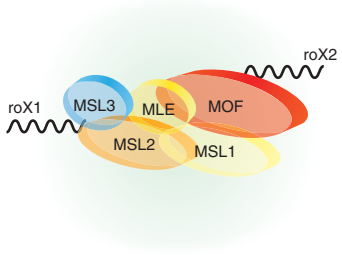



**Figure 1** Different mechanisms for gene-dosage compensation. Across numerous organisms, sex-determining mechanisms result in the two sexes differing in their number of X chromosomes, yet both require equivalent levels of X-chromosome gene products. The regulatory processes collectively known as dosage compensation function to neutralize the difference in gene doses from the X chromosome between males and females. Although the mechanisms of dosage compensation vary widely across species, common to all organisms is the regulation of this process at the level of chromatin. (**a**) In worms, the dosage-compensation complex (DCC) represses X-chromosome dosage by half in each X chromosome of the female (XX) to equal expression of the male (XO) X chromosome. (**b**) In flies, the male (XY) doubles transcription from the X chromosome to achieve the levels of gene expression reached by two actively transcribed X chromosomes in the female (XX). Hypertranscription of the single male X chromosome is mediated by the MSL ribonucleoprotein complex, which causes a two-fold transcriptional upregulation of most X-linked genes. As females lack a functional MSL complex, hypertranscription of the X chromosome occurs solely in males. (**c**) In mammals, dosage compensation involves the random inactivation of one X chromosome of the female (XX) to meet expression levels from one X in the male (XY). This is achieved through spreading of the X-linked large noncoding RNA Xist. The essential role of dosage-compensation mechanisms is underlined by the lethality incurred with failure to accomplish equalization of gene expression of the X chromosome. Thickness of chromosomes indicates expression level from the X chromosome.

The question of how the MSL complex bypasses all autosomes to specifically target the X chromosome had remained a mystery for many years. Particularly puzzling was the fact that no DNA sequence or motif could be identified that uniquely defines the X chromosome[10]. The first indication of an entry site for the MSL complex on the X chromosome came with the finding that the MSL complex can nucleate at the two X-linked noncoding RNA *roX* genes[11]. Studies of the MSL binding pattern using partial MSL complexes that lack specific subunits (initially using cytological methods and later by higher-resolution ChIP approaches) led to the identification of hundreds of additional CESs[12,13]. Interestingly, it was only with the recent development of the ChIP-Sequencing technology to attain sufficient dynamic range that high-affinity MSL binding sites became distinguishable. This ultimately

*Alison M. Hosey and Marjorie Brand are at the The Sprott Center for Stem Cell Research, Regenerative Medicine Program, Ottawa Health Research Institute, 501 Smyth Road, Ottawa ON K1H 8L6, Canada.*
*e-mail: mbrand@ohri.ca*



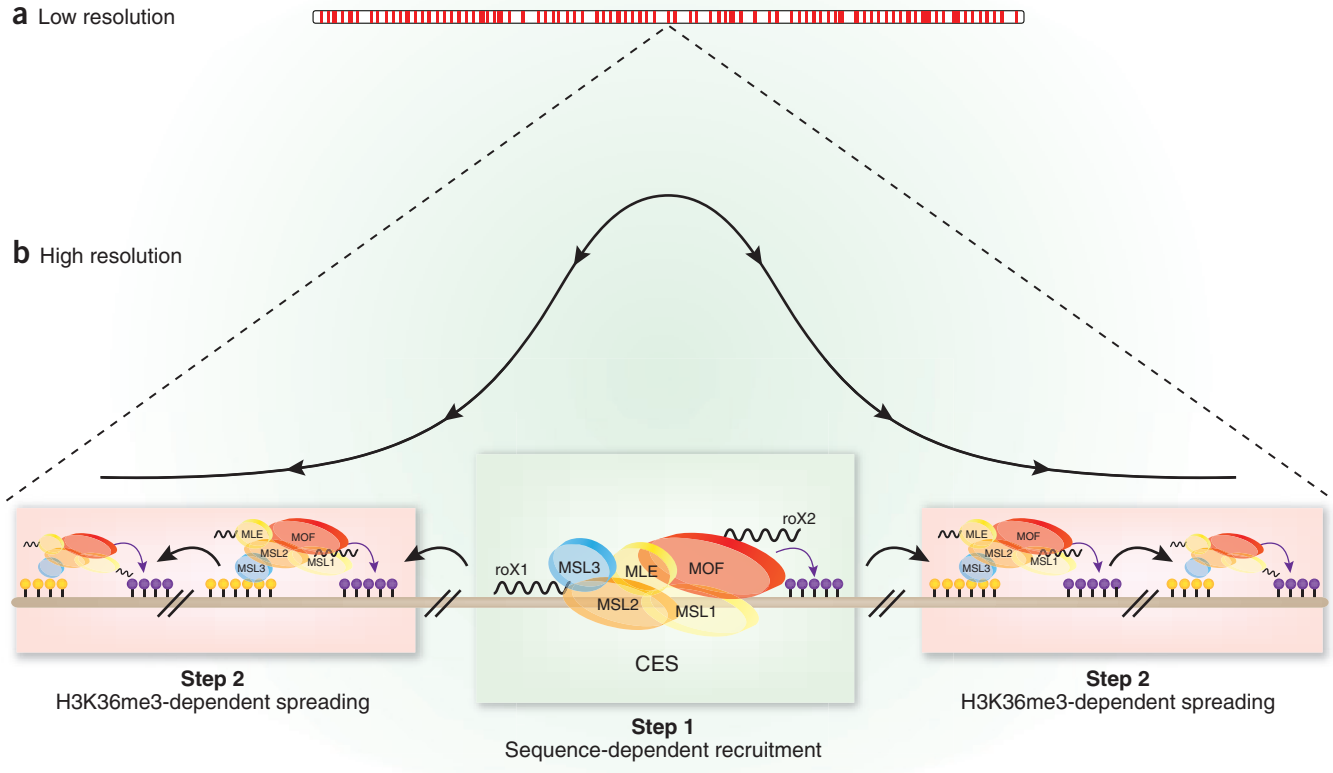

Kim Caesar

**Figure 2** The MSL complex. The MSL complex is composed of two X-linked noncoding RNAs—RNA on the X 1 and 2 (roX1 and roX2)—and five proteins—the H4K16 acetyltransferase MOF, the ATP-dependent DEXH-box RNA and DNA helicase Maleless (MLE) and three additional subunits (MSL1, 2 and 3).

led to identification of the long-awaited MSL-recognition DNA element[13]. However, despite the identification of these high-affinity entry sites, the process leading to MSL binding to the remaining sites on the X chromosome (representing 75% of the total MSL binding sites) was still unresolved. More precisely, it was not clear whether this binding outside of the CES is also mediated via a sequence-dependent mechanism involving degenerate, lower-affinity DNA motifs[14,15] or via a sequence-independent spreading from the CES.

Sural *et al.*[16] showed using ChIP-chip that MSL3 proteins with mutations within their chromodomain retain binding to the CES but have a reduced ability to spread. This provides evidence that spreading *in cis* is dependent on the MSL3 chromodomain and illustrates the resolving power of ChIP-chip over polytene chromosome immunostaining methods that could not detect the loss of 75% of MSL3 binding sites in a MSL3Δ chromodomain mutant[7,8]. Importantly, the authors observe that MSL3 chromodomain mutants are specifically defective in their binding to H3K36me3-containing nucleosomes. This, together with the previous finding that the MSL complex spreads to active genes marked by H3K36me3 in a Set2 methyltransferase–dependent manner[16,17], provides evidence that the MSL3 chromodomain induces spreading via specific interaction with the H3K36me3 mark (**Fig. 3**).

Interestingly, H3K36 trimethylation is localized to the 3′ end of active genes. The discontinuity in this mark across the genome may well suggest a 'hopping' mechanism for spreading of the MSL complex, similar to that previously proposed for the spreading of the repressive *D. melanogaster* PRC1 and PRC2 Polycomb complexes moving away from the Polycomb responsive element (PRE)[2]. Notably, the chromodomain-containing Polycomb subunit of PRC1 has been reported to spread outside the PRE in a pattern strikingly similar to that of MSL3 (high-affinity binding Polycomb is centered on the PRE and tails gradually on both sides of the PRE), whereas the PSC subunit of the same PRC1 complex is tightly localized at the PRE with no spreading[18]. These results suggest that




Kim Caesar

**Figure 3** Targeting of the MSL complex to the male X chromosome in *D. melanogaster*. (**a**) The MSL ribonucleoprotein complex forms a finely banded pattern on the male X chromosome at low resolution. (**b**) High-resolution two-step model for MSL binding to the X chromosome. In a first step (shaded in green), initial binding occurs at chromatin entry sites (CESs). This step is DNA sequence dependent and seems to be independent of the MSL3 chromodomain. At the CES, MSL locally activates nearby target genes via MOF-mediated acetylation (purple arrows) of H4K16 (purple lollipops). In a second step (shaded in pink), MSL complex spreading (black arrows) from the CES occurs through MSL3 chromodomain–mediated recognition of H3K36me3 (yellow lollipops). This step is dependent on the MSL3 chromodomain binding to H3K36me3 but independent of DNA sequence. The bell-shaped curve represents the ChIP distribution over 10 kb of bound MSL complex at the CES and in bidirectional flanking regions. The MSL complex subunit MOF is responsible for catalyzing acetylation of H4K16.

spreading might entail disruption of complexes, with only some subunits actually involved in spreading. In this context, it is interesting to note that, on active X-linked genes, the Male absent on the first (MOF) acetyltransferase does not colocalize perfectly with MSL3. Instead, MSL3 binding is biased toward the 3′ end of genes, whereas MOF shows a bimodal distribution (enriched at both the promoter and the 3′ end)[19]. Therefore it will be important to determine whether MOF binding is dependent on the chromodomain of MSL3. Adding to the complexity of the spreading mechanism is the finding that MOF and MSL1 (**Fig. 2**) are required for MSL3 binding to low-affinity sites[19]. This indicates that there might be additional mechanisms that regulate MSL3 binding to H3K36me3.

Regardless of the spreading mechanism, the Sural *et al.* paper[7] adds to recent studies that have revealed surprising parallels between repressive and active chromatin domains. First, formation of heterochromatin requires the active transcription of noncoding RNA. In addition, proteins (such as HP1) and histone marks (such as H3K9me) that were once thought to be unique to repressive chromatin have since been found within coding regions of active genes[20,21]. Therefore, it seems that mechanisms used to generate repressive heterochromatin might be conserved, at least in part, for the formation of large domains of active chromatin. Clearly, spreading does not occur at every locus, but it might be particularly important for coordinated regulation of clustered genes such as Hox or β-globin genes. Interestingly, spreading of the H3K4 methyltransferases MLL1 and MLL2 have been observed at these loci[22,23]. Although MLL spreading could be mediated via interaction with the elongating polymerase[24], the study by Sural *et al.* suggests that it could also entail interaction with methylated histones. Indeed, the MLL family of proteins contains PHD domains that can recognize specific methyl marks. In fact, proteins containing domains of recognition for specific histone modifications, including bromo- and chromodomains, WD40 repeats or ankyrin domains[25,26], are all potential candidates to mediate complex spreading.

Although many questions remain about the exact molecular mechanism of spreading, as well as its role in coordinating gene expression, Sural *et al.* have paved the way for future discoveries related to the structure and function of large domains of active chromatin.

ACKNOWLEDGMENTS

We thank F.J. Dilworth and M. Groudine for critical reading of the manuscript. The work in the Brand group is supported by the Canadian Institutes of Health Research. M. Brand holds the Canadian Research Chair in the regulation of gene expression.

1. Li, B., Carey, M. & Workman, J.L. *Cell* **128**, 707–719 (2007).
2. Talbert, P.B. & Henikoff, S. *Nat. Rev. Genet.* **7**, 793–803 (2006).
3. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. *Nature* **410**, 116–120 (2001).
4. Elgin, S.C. & Grewal, S.I. *Curr. Biol.* **13**, R895–R898 (2003).
5. Bulger, M. *J. Biol. Chem.* **280**, 21689–21692 (2005).
6. Lucchesi, J.C., Kelly, W.G. & Panning, B. *Annu. Rev. Genet.* **39**, 615–651 (2005).
7. Sural, T.H. *et al. Nat. Struct. Mol. Biol.* **15**, 1318–1325 (2008).
8. Buscaino, A., Legube, G. & Akhtar, A. *EMBO Rep.* **7**, 531–538 (2006).
9. Krogan, N.J. *et al. Mol. Cell. Biol.* **23**, 4207–4218 (2003).
10. Straub, T. & Becker, P.B. *Curr. Opin. Genet. Dev.* **18**, 175–180 (2008).
11. Kelley, R.L. *et al. Cell* **98**, 513–522 (1999).
12. Kelley, R.L. *Dev. Biol.* **269**, 18–25 (2004).
13. Alekseyenko, A.A. *et al. Cell* **134**, 599–609 (2008).
14. Gilfillan, G.D. *et al. Nucleic Acids Res.* **35**, 3561–3572 (2007).
15. Kind, J. & Akhtar, A. *Genes Dev.* **21**, 2030–2040 (2007).
16. Larschan, E. *et al. Mol. Cell* **28**, 121–133 (2007).
17. Bell, O. *et al. Mol. Cell. Biol.* **28**, 3401–3409 (2008).
18. Schwartz, Y.B. & Pirrotta, V. *Nat. Rev. Genet.* **8**, 9–22 (2007).
19. Kind, J. *et al. Cell* **133**, 813–828 (2008).
20. Vakoc, C.R., Mandat, S.A., Olenchock, B.A. & Blobel, G.A. *Mol. Cell* **19**, 381–391 (2005).
21. Huisinga, K.L., Brower-Toland, B. & Elgin, S.C. *Chromosoma* **115**, 110–122 (2006).
22. Guenther, M.G. *et al. Proc. Natl. Acad. Sci. USA* **102**, 8603–8608 (2005).
23. Demers, C. *et al. Mol. Cell* **27**, 573–584 (2007).
24. Eissenberg, J.C. & Shilatifard, A. *Curr. Opin. Genet. Dev.* **16**, 184–190 (2006).
25. Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. & Patel, D.J. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
26. Collins, R.E. *et al. Nat. Struct. Mol. Biol.* **15**, 245–250 (2008).



# Alternative splicing: regulation without regulators

Brenton R Graveley

**Alternative splicing is typically thought to be controlled by RNA binding proteins that modulate the activity of the spliceosome. A new study not only demonstrates that alternative splicing can be regulated without the involvement of auxiliary splicing factors, but also provides mechanistic insight into how this can occur.**

The field of alternative splicing is moving forward at a breathtaking pace. Only a little more than 10 years ago, it was thought that most human genes encoded only a single mRNA isoform. However, it is now known that alternative splicing is the rule, not the exception—~90% of human genes encode at least two isoforms, and most known alternative-splicing events are regulated[1,2]. This implies that there is an extensive network of splicing regulators to control the plethora of alternative-splicing events. However, the number of known splicing regulators (<50) and even the number of known RNA binding proteins encoded by the human genome (<300) cannot alone be responsible for controlling all of the alternative-splicing events we know about. Are there other types of RNA binding proteins that we are unaware of? Are there mechanisms for controlling alternative splicing that fall outside our traditional view? Although the answers to both questions are likely to be yes, a recent paper in *Cell* by Nilsen and colleagues[3], provides strong evidence for the latter—alternative splicing can be controlled in a manner that is independent of known splicing regulators.

Traditional models of regulated alternative splicing involve auxiliary splicing factors—proteins that are not core components of the spliceosome, but rather bind to the pre-mRNA and either enhance or repress the ability of the spliceosome to recognize particular splice sites. Two important classes of auxiliary splicing factors are SR proteins and heterogeneous nuclear ribonucleoproteins (hnRNP) proteins.

*Brenton R. Graveley is in the Department of Genetics and Developmental Biology, University of Connecticut Stem Cell Institute, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030-3301, USA.*
*e-mail: graveley@neuron.uchc.edu*