

Minireview

Decoding dosage compensation

Xinxian Deng and Christine M Disteche

Address: Department of Pathology, University of Washington, Seattle, WA 98195, USA.

Correspondence: Christine M Disteche. Email: cdistech@u.washington.edu

Published: 26 February 2007

Genome Biology 2007, **8**:204 (doi:10.1186/gb-2007-8-2-204)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/2/204>

© 2007 BioMed Central Ltd

Abstract

How the mechanisms of dosage compensation distinguish the sex chromosomes from the autosomes has been something of a mystery. A recent study in *Caenorhabditis elegans* has identified clusters of two common DNA motifs as a *cis*-acting code for the recruitment of the DCC, the protein complex that mediates dosage compensation.

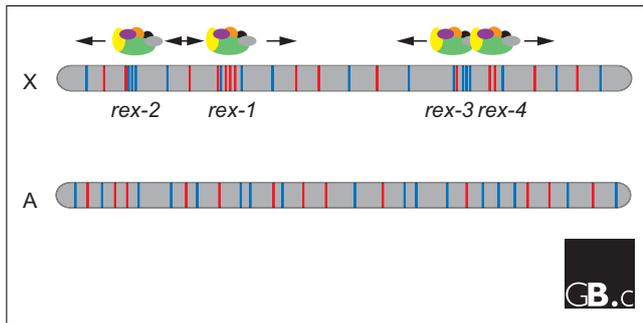
In animals in which sex is determined by a divergent pair of sex chromosomes (XY), the resulting imbalance of gene dose between autosomes and sex chromosomes is potentially fatal. This imbalance is redressed by a process termed dosage compensation, which modulates global gene expression on the X chromosome to restore a balanced network of gene expression in the diploid cells in both sexes [1-4]. In *Drosophila*, dosage compensation is achieved by a twofold upregulation of expression from the single X chromosome in males. In contrast, in mammals and *Caenorhabditis elegans*, X-chromosome upregulation occurs in both sexes, together with a reduction of X-linked gene expression in the homogametic sex by silencing one of the two X chromosomes in mammalian females or repressing both X chromosomes in *C. elegans* hermaphrodites. Despite such diverse strategies, a fundamental question in the study of dosage compensation in all these species is how specific recognition of an entire X chromosome occurs. McDonel *et al.* [5] have recently shed light on this problem. They have discovered two clustered DNA motifs on the *C. elegans* X chromosome that act as a *cis*-acting code for recognition by the dosage-compensation complex (DCC), a complex of proteins that is responsible for the repression of the X chromosomes in hermaphrodites. This breakthrough points out the importance of DNA sequence in specifying the target of the DCC.

Dosage-compensation mechanisms involve both protein complexes that act in *trans* and elements that are produced by the X chromosome itself, such as *X inactive-specific transcript (XIST)* RNA, which is essential for the onset of

mammalian X inactivation. The question is how such regulatory complexes become specifically associated with, or stay with, the X chromosome. X inactivation is random in eutherian mammals, with either the paternal or maternal X chromosome silenced; however, in early mouse embryos and in marsupials, the paternal X chromosome is always inactivated. In this type of imprinted dosage regulation the paternal X chromosome may be originally identified by its unpaired status at male meiosis [6,7]. Chromatin marks and location of the X chromosome within a specific nuclear compartment may also play a role for its identification, but all these hypotheses ultimately hinge on the ability to distinguish the DNA of the X chromosome from that of the autosomes. Although the importance of DNA elements in X-chromosome recognition is well appreciated in flies, mammals and worms, simple DNA motifs have not yet been identified. The new study, together with substantial evidence from studies of dosage compensation in flies and mammals, begins to unveil the mystery of chromosome-wide recognition.

Identification of DNA motifs that recruit the DCC in *C. elegans*

Previous studies in *C. elegans* suggested that discrete DNA recognition sites for the DCC called *recruitment element on the X (rex)* are widely distributed along the X chromosome [8]. To find the signature of *rex* sites, McDonel *et al.* [5] first tested extrachromosomal arrays comprising multiple copies of large DNA fragments sampled from regions known to bind the DCC, followed by deletion mapping of the DNA

**Figure 1**

Clustered short DNA motifs recruit the DCC in *C. elegans*. It is proposed that the DCC (shown as the multicolored globular structure above the X chromosome (X)) recognizes and binds at four specific sites (*rex-1*, *rex-2*, *rex-3* and *rex-4*) and then spreads or diffuses to the rest of the X chromosome. The *rex* sites have high densities of at least two degenerate DNA motifs (motif A, red; motif B, blue) and this clustering is required to bind the DCC. The same motifs occur on autosomes (A), but they are not clustered and thus do not bind the DCC.

fragments to refine the identity of candidate binding sites. In this way they identified for the first time four minimal *rex* sites for binding, whose sizes range from 115 to 411 bp of the *C. elegans* X chromosome (Figure 1). When present in extra-chromosomal arrays in multiple copies these *rex* sites have full DCC-binding activity.

The new data support the model of *C. elegans* DCC action in which the condensin-like DCC is first recruited to individual sites and then spreads or diffuses to neighboring regions to repress gene transcription over the entire X chromosome [5]. McDonel *et al.* [5] also define two short degenerate DNA motifs (A and B) that are clustered within the *rex* sites (Figure 1). They showed by mutation analysis that motif A is critical for DCC binding, whereas motif B is also important, but to a lesser extent. Intriguingly, these motifs are not X-chromosome-specific and are not even enriched on the X chromosome. Their co-occurrence at high density is essential for full DCC recruitment activity in *rex* arrays (Figure 1), indicating that it is their clustered distribution on the X chromosome that is critical for their function.

X-chromosome-specific recognition motifs in other species

Global recognition of X chromosomes may have evolved in a similar way among different species. In the somatic cells of *Drosophila*, the single male X chromosome is bound and upregulated by a dosage-compensation complex called the male-specific lethal complex (MSL), which includes specific proteins and two noncoding RNAs transcribed from genes on the X chromosome - *RNA on the X 1* and *2* (*roX1* and *roX2*) [1]. A number of X-linked sites have affinity for MSL, but to a variable extent [9,10]. It has been proposed that densely clustered sites with a range of affinities for MSL

work cooperatively at short and medium distances to recruit the complex to the whole X chromosome [11,12].

Two recent analyses, based in one case on the sequence of high-affinity binding sites for MSL and in the other on a compilation of MSL-binding sequences found by whole-genome chromatin immunoprecipitation analyzed by high-resolution genome tiling arrays (ChIP-chip), have identified several recurrent motifs made up of various combinations of degenerate short DNA sequences [10,13]. Notably, most of these short elements contain GA or CA dinucleotides, and mutation of GAGA repeats in the weakly conserved sequence of around 200 bp in *roX1* and *roX2* genes, which encode the RNA components of the DCC, disrupts MSL binding to the *roX* genes [14]. Although the presence of these short DNA sequences partially explains MSL binding on the X chromosome, not all binding sites can be predicted. Interestingly, in *Drosophila* there is no overall enrichment on the X chromosome of the degenerate short DNA sequences that recruit the MSL complex [10,13]; it is the combination of these sequences that seems to be important - a situation similar to that reported for *C. elegans* by McDonel *et al.* [5].

In mammals, the mechanism of X upregulation is not known [15], whereas silencing of one X chromosome in females by X inactivation is well characterized. X inactivation is initiated at the *XIST* locus located within the X-inactivation center, from which silencing spreads in *cis* along the chromosome [2]. Repeats of the LINE-1 element are enriched on the X chromosome and might act as 'relay elements' to facilitate the spreading of gene silencing [16]. Enrichment in specific dinucleotides has also been reported for the mammalian X chromosome [17], but it remains to be determined whether these play any role in silencing, or possibly in upregulation. Segregation of the inactive X chromosome as a body of condensed chromatin (the Barr body) could also facilitate the spreading of silencing, which may be initiated by nucleation of heterochromatin at the center of the Barr body [18,19].

How accurate is the *C. elegans* DCC recognition code?

The model of McDonel *et al.* [5] predicts that a cosmid with at least one motif A and one motif B clustered within a window of 600 bp should have DCC-binding activity. Using this simple model, the authors showed that the distribution of predicted 30 kb segments of this type corresponded well to the DCC-recruiting activity of large regions of the X chromosome mapped in previous work [8]. Although this model predicted DCC-binding cosmids with 44% accuracy, it is notable that a proportion of the predictions were incorrect, suggesting that other DNA motifs might be involved in DCC binding. Because of the limitation of extra-chromosomal array analysis, it will be difficult and time-consuming to search for other *rex* sites and DNA motifs

using this approach. As mentioned above, whole-genome ChIP-chip in *Drosophila* has revealed the complete DCC-binding pattern along the X chromosome [13,20]. Such an analysis in *C. elegans* would generate global DCC-binding profiles and provide clues for the identification of additional *rex* sites with high binding activity. Additional factors might play a role in DCC recruiting: for example, active transcription and nuclear pore proteins have been shown to affect DCC targeting to the *Drosophila* X chromosome [21,22]. Highly expressed X-linked genes are frequent DCC targets, implying either the need for dosage compensation for these active X-linked genes or facilitation of DCC binding due to the open chromatin structure [13,20].

Another intriguing finding by McDonel *et al.* [5] is that motifs A and B are partially functionally redundant despite their distinct DNA sequences. Furthermore, the binding activity of a single type of *rex* site in an extrachromosomal array is closely correlated with the number of motifs, suggesting that these motifs might work cooperatively to recruit the DCC. How are these clustered short motifs recognized? Do they interact with multiple domains of one protein or with different proteins of the *C. elegans* DCC simultaneously? It will be important to identify these proteins, or possibly novel proteins, as well as their order of recruitment. Other potential candidates that may recognize the *cis* motifs are noncoding RNAs. Large noncoding RNAs, such as *XIST* in mammals and *roX* in flies, play an important role in directing the DCC to the X chromosome. Interestingly, both *XIST* and *roX* RNAs contain functionally redundant elements, which might interact collaboratively with degenerate DNA elements via imprecise base pairing [2,12,23]. It will be of great interest to search for noncoding transcripts from the *C. elegans* chromosomal regions containing *rex* sites.

Finally, it should be noted that the dissection of *C. elegans* *rex* sites by McDonel *et al.* [5] was carried out mostly using extrachromosomal arrays, which might behave differently from the native *rex* sites. For example, in mutants of SDC-3, a core protein component of the DCC, another core protein component, SDC-2, still binds to the *rex-1* arrays but not to the X chromosome containing native *rex* sites [5]. The nature of these *cis*-acting DNA elements will need to be ultimately addressed by mutation analysis of endogenous *rex* sites. It would be interesting to test whether deletion of any *rex* site is lethal to the animal, affects spreading or recognition of DCC to adjacent regions, or derepresses neighboring X-linked genes. Such studies will greatly expand our understanding of the basis of chromosome-wide recognition of the X chromosome in dosage compensation.

Acknowledgements

This work was supported by grants from the National Institutes of Health.

References

1. Lucchesi JC, Kelly WG, Panning B: **Chromatin remodeling in dosage compensation.** *Annu Rev Genet* 2005, **39**:615-651.
2. Heard E, Distèche CM: **Dosage compensation in mammals: fine-tuning the expression of the X chromosome.** *Genes Dev* 2006, **20**:1848-1867.
3. Larsson J, Meller VH: **Dosage compensation, the origin and the afterlife of sex chromosomes.** *Chromosome Res* 2006, **14**:417-431.
4. Straub T, Becker PB: **Dosage compensation: the beginning and end of generalization.** *Nat Rev Genet* 2007, **8**:47-57.
5. McDonel P, Jans J, Peterson BK, Meyer BJ: **Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex.** *Nature* 2006, **444**:614-618.
6. Huynh KD, Lee JT: **Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos.** *Nature* 2003, **426**:857-862.
7. Namekawa SH, Park PJ, Zhang LF, Shima JE, McCarrey JR, Griswold MD, Lee JT: **Postmeiotic sex chromatin in the male germline of mice.** *Curr Biol* 2006, **16**:660-667.
8. Csankovszki G, McDonel P, Meyer BJ: **Recruitment and spreading of the *C. elegans* dosage compensation complex along X chromosomes.** *Science* 2004, **303**:1182-1185.
9. Oh H, Bone JR, Kuroda MI: **Multiple classes of MSL binding sites target dosage compensation to the X chromosome of *Drosophila*.** *Curr Biol* 2004, **14**:481-487.
10. Dahlsveen IK, Gilfillan GD, Shelest VI, Lamm R, Becker PB: **Targeting determinants of dosage compensation in *Drosophila*.** *PLoS Genet* 2006, **2**:e5.
11. Straub T, Dahlsveen IK, Becker PB: **Dosage compensation in flies: mechanism, models, mystery.** *FEBS Lett* 2005, **579**:3258-3263.
12. Deng X, Meller VH: **Non-coding RNA in fly dosage compensation.** *Trends Biochem Sci* 2006, **31**:526-532.
13. Gilfillan GD, Straub T, de Wit E, Greil F, Lamm R, van Steensel B, Becker PB: **Chromosome-wide gene-specific targeting of the *Drosophila* dosage compensation complex.** *Genes Dev* 2006, **20**:858-870.
14. Park Y, Mengus G, Bai X, Kageyama Y, Meller VH, Becker PB, Kuroda MI: **Sequence-specific targeting of *Drosophila* roX genes by the MSL dosage compensation complex.** *Mol Cell* 2003, **11**:977-986.
15. Nguyen DK, Distèche CM: **Dosage compensation of the mammalian X chromosome.** *Nature Genetics* 2006, **38**:47-53.
16. Lyon MF: **The Lyon and the LINE hypothesis.** *Semin Cell Dev Biol* 2003, **14**:313-318.
17. McNeil JA, Smith KP, Hall LL, Lawrence JB: **Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]_n in the X escape region.** *Genome Res* 2006, **16**:477-484.
18. Chaumeil J, Le Baccon P, Wutz A, Heard E: **A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced.** *Genes Dev* 2006, **20**:2223-2237.
19. Clemson CM, Hall LL, Byron M, McNeil J, Lawrence JB: **The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences.** *Proc Natl Acad Sci USA* 2006, **103**:7688-7693.
20. Alekseyenko AA, Larschan E, Lai VVR, Park PJ, Kuroda MI: **High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome.** *Genes Dev* 2006, **20**:848-857.
21. Sass GL, Pannuti A, Lucchesi JC: **Male-specific lethal complex of *Drosophila* targets activated regions of the X chromosome for chromatin remodeling.** *Proc Natl Acad Sci USA* 2003, **100**:8287-8291.
22. Mendjan S, Taipale M, Kind J, Holz H, Gebhardt P, Schelder M, Vermeulen M, Buscaino A, Duncan K, Mueller J, *et al.*: **Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*.** *Mol Cell* 2006, **21**:811-823.
23. Wutz A: **RNAs templating chromatin structure for dosage compensation in animals.** *BioEssays* 2003, **25**:434-442.