

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Interaction of the
Dosage Compensation Complex
with DNA



Torsten Fauth
aus
Göppingen

November 2009

Dissertation eingereicht am: 26.11.2009
1. Gutachter: Prof. Dr. Peter B. Becker
2. Gutachter: Prof. Dr. Dirk Eick
Mündliche Prüfung am: 04.02.2010

Im Gedenken an meine Inge-Oma.

Table of contents

<i>Table of contents</i>	4
<i>Summary</i>	8
<i>Zusammenfassung</i>	9
1 Introduction	10
1.1 Dosage Compensation	11
1.2 Modulation of chromatin structure	12
1.3 Dosage Compensation in <i>C. elegans</i>	13
1.3.1 Composition of the <i>C. elegans</i> DCC	13
1.3.2 Recruitment-and-spreading model	14
1.3.3 Autosomal binding sites	16
1.3.4 Mechanism of X chromosomal repression	16
1.4 Dosage compensation in mammals	17
1.4.1 Selection of the inactive X chromosome	17
1.4.2 Mechanisms of X chromosome silencing.....	17
1.5 Dosage compensation in <i>Drosophila melanogaster</i>	20
1.5.1 Composition of the <i>Drosophila melanogaster</i> DCC	20
1.5.2 Mechanism of two-fold activation.....	21
1.5.3 Principles of X chromosome targeting	23
1.5.4 High affinity sites and DNA recruitment elements	26
1.5.5 The MSL2-MSL1 targeting machinery	28
1.6 Protein-DNA interactions	30
1.6.1 DNA binding domains.....	30
1.6.2 Principles of DNA recognition	32
1.6.3 Zinc finger proteins	34
1.7 Objectives	38
2 Results	39
2.1 Purification of recombinant MSL proteins	39
2.2 MSL2 is the DNA binding factor in the MSL2-MSL1 complex	41
2.3 The CXC domain of MSL2 is required but not sufficient for DNA binding	42
2.4 Point mutations within the CXC domain affect DNA binding	43
2.5 Double stranded DNA is the preferred binding target for MSL2	44

2.6	DNA binding of MSL2 is sequence-independent	46
2.7	Affinity of MSL2 to DNA is increased with the length of DNA.....	48
2.8	Reporter gene activation in cells requires the CXC domain	49
2.9	The CXC domain is necessary for targeting the DCC to the X chromosome	50
2.10	Binding of MSL2 to RNA-DNA triplex structures	52
2.11	MSL2 is not sufficient for reporter gene activation in human cells.....	54
2.12	Enrichment of a native DCC-containing fraction.....	55
2.13	DNA binding of an enriched native DCC fraction	57
3	<i>Discussion</i>.....	58
3.1	Multimerization of the MSL2-MSL1 complex.....	58
3.2	The DNA binding factor of the MSL2-MSL1 complex	59
3.3	CXC domains as DNA binding modules.....	60
3.3.1	DNA binding by the CXC domain of MSL2.....	60
3.3.2	CXC domains and their function in DNA binding	61
3.3.3	The CXC domain – a conserved DNA binding module.....	62
3.4	MSL2 – an additional RNA binding protein within the DCC	64
3.5	Recognition of high affinity sites	64
3.6	Alternative binding mechanisms	65
3.6.1	The nature of HAS: Involvement of triple helix structures?	66
3.6.2	Factor X required for sequence-specific binding?.....	67
3.6.3	High affinity sites as allosteric effectors?.....	69
3.7	Outlook	69
4	<i>Materials and Methods</i>.....	71
4.1	Materials.....	71
4.1.1	General chemicals.....	71
4.1.2	Chemicals for tissue culture	71
4.1.3	Antibodies.....	72
4.1.4	Bacteria and cell lines.....	73
4.1.5	Dialysis, chromatographic and filtration material	73
4.1.6	Kits and enzymes.....	73
4.1.7	Oligonucleotides.....	73
4.1.7.1	Oligonucleotides for EMSA	73
4.1.7.2	Oligonucleotides for cloning and site-directed mutagenesis	74

4.1.8	Plasmids.....	75
4.1.9	Baculoviruses.....	77
4.1.10	Other materials.....	77
4.1.11	Standard buffers.....	77
4.2	Mammalian and insect tissue culture	78
4.2.1	General cell culture conditions.....	78
4.2.2	Cultivation of human HEK-293 cells.....	78
4.2.3	Cultivation of <i>Spodoptera frugiperda</i> Sf21 cells.....	79
4.2.4	Cultivation of <i>Drosophila melanogaster</i> SL2 cells.....	79
4.2.5	Freezing and thawing of SL2 cells.....	79
4.2.6	Generation of stable SL2 cell lines.....	79
4.3	Methods for manipulation of DNA	80
4.3.1	General molecular biology methods.....	80
4.3.2	Cloning of expression constructs.....	80
4.3.3	Site-directed mutagenesis.....	81
4.3.4	Preparation of long DNA fragments.....	82
4.3.5	Annealing of oligonucleotides.....	82
4.3.6	Radioactive labeling of DNA fragments.....	82
4.3.7	Formation of RNA-DNA triplexes.....	83
4.4	Methods for protein analysis	83
4.4.1	Precipitation of proteins with Trichloroacetic acid (TCA).....	83
4.4.2	SDS-polyacrylamide gel electrophoresis (SDS-PAGE).....	83
4.4.3	Silver staining of protein gels.....	84
4.4.4	Staining of protein gels by Coomassie Blue.....	84
4.4.5	Determination of protein concentration.....	85
4.4.6	Protein transfer to membranes (Western blot).....	85
4.4.7	Immunodetection of transferred proteins.....	85
4.4.8	Superose 6 gelfiltration of purified MSL proteins.....	86
4.4.9	Mass spectrometry.....	86
4.5	Expression and purification of proteins and protein complexes	87
4.5.1	Heterologous expression of MSL proteins in insect Sf21 cells.....	87
4.5.1.1	Infection of Sf21 cells with baculoviruses.....	87
4.5.1.2	Preparation of Sf21 cell extracts.....	87
4.5.1.3	Purification of FLAG-tagged MSL proteins from Sf21 cell extracts.....	88
4.5.2	Bacterial expression of the GST-tagged CXC domain.....	89
4.5.2.1	Induction of protein expression.....	89
4.5.2.2	Preparation of bacterial cell extract.....	89
4.5.2.3	Purification of GST-tagged CXC domain from bacterial cell extracts.....	90
4.5.3	Enrichment of native MOF containing fractions from SL2 cells.....	90

4.5.3.1	Preparation of nuclear extracts from SL2 cells.....	90
4.5.3.2	Purification of FLAG-tagged MOF from SL2 nuclear extracts.....	92
4.6	Reporter gene assays in insect and human cells	92
4.6.1	Reporter gene assay in SL2 cells.....	92
4.6.1.1	Transient transfection of reporter gene constructs.....	92
4.6.1.2	Preparation of cell extracts from transiently transfected SL2 cells	93
4.6.1.3	Measuring luciferase activities	93
4.6.2	Reporter gene assay in human HEK-293 cells	93
4.6.2.1	Transient transfection of reporter gene constructs.....	93
4.7	Protein-DNA interactions	94
4.7.1	Electrophoretic mobility shift assays (EMSA).....	94
4.7.2	Calculation of affinity constants.....	94
4.8	Immunofluorescence staining on SL2 cells.....	95
5	References.....	96
6	Appendix	105
6.1	Abbreviations	105
6.2	Acknowledgments.....	107
6.3	Ehrenwörtliche Versicherung.....	108
6.4	Erklärung	109
6.5	Curriculum vitae.....	110

Summary

The proper and precise regulation of gene expression is crucial for all cellular processes. In particular it is a major task for regulating the eukaryotic genome to not only control single genes but in addition to simultaneously regulate chromosomal domains that are widely spread. A powerful model to study the selective regulation of an entire chromosome is the process of dosage compensation in the common fruit fly *Drosophila melanogaster*.

The precise activation of gene expression from the single male X chromosome requires the selective recruitment of the dosage compensation complex (DCC) to the male X chromosome. However, the mechanism by which the DCC distinguishes the X from the other chromosomes is poorly understood. It is thought that a subcomplex consisting of the core proteins MSL2 and MSL1 serves as a binding module, which initially recognizes so-called high affinity sites on the X chromosome. Despite extensive *in vivo* characterization, a direct interaction of the MSL2-MSL1 subcomplex with high affinity sites has not been shown.

Therefore a purified system consisting of recombinant MSL proteins and synthetic nucleic acid structures was established to quantitatively measure their interaction *in vitro*. The MSL2 protein was identified as the DNA binding factor within the MSL2-MSL1 complex. For DNA binding MSL2 required its CXC domain, to which so far no function had been assigned. The relevance of this DNA binding function was confirmed *in vivo*: The CXC domain was required for MSL2 recruitment to a reporter gene and for the proper targeting of the DCC to the X chromosome in *Drosophila* cells. Moreover, with MSL2 being able to bind RNA, a novel RNA binding protein within the DCC has been identified.

However, MSL2 bound DNA in a sequence-independent manner, indicating that for selective binding additional factors might be required. This is supported by the finding that in contrast to the *Drosophila* system, in a heterologous reporter gene system MSL2-MSL1 could not activate transcription. In summary, the DNA binding activity by MSL2 is crucial, nevertheless, for achieving binding specificity additional targeting principles must be involved.

Zusammenfassung

Die korrekte und genaue Regulation der Genexpression ist unabdingbar für alle zellulären Prozesse. Insbesondere bei der Regulation des eukaryotischen Genoms ist es eine schwierige Aufgabe nicht nur einzelne Gene, sondern auch chromosomale Domänen, die weit verteilt sind, gleichzeitig zu regulieren. Ein geeignetes Modell, um die Selektion und Regulation eines ganzen Chromosoms zu untersuchen, ist der Prozess der Dosis Kompensation in der Fruchtfliege *Drosophila melanogaster*.

Die genaue Aktivierung der Genexpression des einzelnen männlichen X Chromosoms bedarf der selektiven Rekrutierung des Dosis Kompensations Komplexes (*dosage compensation complex*, DCC) zum männlichen X Chromosom. Jedoch ist der Mechanismus, durch den der DCC das X von den anderen Chromosomen unterscheidet, nur unzureichend aufgeklärt. Man geht davon aus, dass ein Teilkomplex bestehend aus den Proteinen MSL2 und MSL1 als Bindungsmodul dient, welches anfänglich sogenannte *high affinity sites* auf dem X Chromosom erkennt. Trotz der umfangreichen Charakterisierung *in vivo*, wurde eine direkte Interaktion des MSL2-MSL1 Teilkomplexes mit *high affinity sites* noch nicht gezeigt.

Deshalb wurde ein gereinigtes System etabliert, welches sich aus rekombinanten MSL Proteinen und synthetischen Nukleinsäure-Strukturen zusammensetzt, um deren Interaktion quantitativ *in vitro* zu messen. Das MSL2 Protein wurde als der DNA-bindende Faktor innerhalb des MSL2-MSL1 Komplexes identifiziert. Zur DNA-Bindung benötigt MSL2 seine CXC Domäne, für die bisher keine Funktion bekannt war. Die Relevanz dieser DNA-Bindungsfunktion wurde auch *in vivo* bestätigt: Die CXC Domäne war für die Rekrutierung von MSL2 zu einem Reportergen und für das korrekte *targeting* des DCC zum X Chromosom in *Drosophila* Zellen notwendig. Daneben wurde durch die Erkenntnis, dass MSL2 auch RNA zu binden vermag ein weiteres potentiell RNA-Bindungsprotein innerhalb des DCC identifiziert.

Jedoch bindet MSL2 DNA auf eine Sequenz-unabhängige Weise, was darauf hindeutet, dass für die selektive Bindung zusätzliche Faktoren gebraucht werden könnten. Dies wird durch das Ergebnis unterstützt, wonach MSL2-MSL1, im Gegensatz zum *Drosophila* System, in einem heterologen System nicht die Transkription aktivieren konnten. Zusammenfassend konnte gezeigt werden, dass die DNA-Bindungsaktivität von MSL2 notwendig ist, um jedoch Bindungsspezifität zu erreichen, bedarf es zusätzlicher Faktoren oder Mechanismen, die es noch aufzuklären gilt.

1 Introduction

The discovery of the DNA as the heritable information and the deciphering of the genetic code, which turned out to be a rather simple four ‘letter’ code, in the middle of the 20th century truly marked milestones and opened undreamed-of possibilities. The sequencing of the complete genome of various organisms, including our own at the beginning of the 21st century, allowed obtaining a genetic ‘blue print’. But on the other hand it became clear that just knowing the ‘blue print’ is by far not sufficient to understand the complexity of cellular processes. It is in fact the usage, that is the realization of the genetic information, which enables a cell to carry out its specific function. This becomes obvious when one thinks about a multi-cellular organism with all its many different cell types. They all carry the identical genetic information, but it is the differential gene expression, which leads to their different morphology and their different functions.

The proper control of gene expression is therefore a major task for all organisms. Factors, like the correct time point or time window (e.g. a specific environmental stimulus or developmental stage) or the correct place (e.g. a specific tissue), for switching a gene on or off have to be considered. In addition simply switching a gene from off-state to on-state is mostly insufficient and much too imprecise to meet the exact needs of a cell. Instead the expression of genes has to be fine-tuned and tightly regulated to achieve a precise level of expression. Another problem arises from the large number of several thousand genes, which have to be controlled. Genes whose expression has to be modulated have to be selected from the huge genome, while all others have to stay untouched. In particular for the more complex organized eukaryotic genome it is a big challenge to not only select a single or a small group of (for example functionally related) genes but in addition to simultaneously regulate chromosomal domains that are widely spread over hundreds or thousands of kilobases of DNA.

A powerful model to study two of the above-mentioned challenges, that is the selectivity of genes and their precise regulation, is the process of dosage compensation in the common fruit fly *Drosophila melanogaster*. Dissecting this process allows insight into how one particular chromosome, namely the single male X chromosome, and almost all of its residing genes are selected, while all the other chromosomes are unaffected and secondly how the expression of those selected genes is tightly limited to a two-fold activation.

1.1 Dosage Compensation

Sexual reproduction occurs quite frequently in nature. It increases genetic variability with every generation and hence allows a population to better cope with environmental changes, which is the basis for evolutionary success [1]. However, the emergence of different sex chromosomes to define the male and female sex would – without a compensatory mechanism – lead to a chromosomal imbalance: The sex with two X chromosomes would have the double amount of X-linked gene products as compared to the other sex with only a single X and a single gene-poor Y chromosome. This huge difference of several thousand gene products between the sexes of the same species would be lethal and therefore has to be corrected by a process called dosage compensation. In general three possibilities exist to compensate the effect of XX vs. XY; and indeed examples for all three mechanisms are found in nature (Figure 1-1 and [2, 3]). In mammals one of the two X chromosomes of females is inactivated and consequently genes from that X chromosome are switched off. Another possibility is realized in the nematode *Caenorhabditis elegans*, where the transcription from both X chromosomes is halved, whereas in *Drosophila melanogaster* transcription is doubled from the single male X chromosome. In all three cases the result is the same, namely equalized levels of gene products between the sexes. However, this compensation leaves the imbalance between the X and the autosomes untouched. To ensure a balance between the single X and the autosomes also in *C. elegans* and in mammals, a global upregulation of the active X compared to autosomes is postulated [4-6]. Yet, each compensation system made use of and adapted special existing systems and machineries; consequently the underlying mechanisms for the precise gene regulation and for the selection of the correct chromosome differ tremendously.

		female/ hermaphrodite	=	male
	<i>Homo sapiens</i>	$X_{[X]}$ 1	=	$X Y$ 1
	<i>Caenorhabditis elegans</i>	$x x$ $1/2 + 1/2$	=	X 1
	<i>Drosophila melanogaster</i>	$X X$ 1 + 1	=	$X Y$ 2

Figure 1-1: Strategies of dosage compensation. To equalize gene expression between the sexes, different strategies of dosage compensation have evolved: In humans and other mammals one of the two X chromosomes of females is inactivated. In *C. elegans* gene expression from both X chromosomes of the hermaphrodite sex is reduced by 2-fold. In *Drosophila melanogaster* transcription from the single male X chromosome is two-fold upregulated. Numbers represent the gene dose from the single chromosomes.

1.2 Modulation of chromatin structure

The dosage compensation systems are based on different mechanisms, nevertheless all make use of principles that allow modulating the structure and organization of chromatin in order to regulate gene expression. How this is achieved in general will be discussed in this chapter, details can be found later in the individual sections.

The organization of the eukaryotic genome into chromatin allows on the one hand packaging DNA into the cell nucleus, on the other hand it restricts accessibility of DNA for all biological processes, such as replication, DNA repair and transcription. Chromatin is not a rigid structure, on the contrary it is highly dynamic and can be altered by various principles, such as the modification of histones, the use of histone variants and the action of chromatin remodeling machines [7].

Histones, which represent as nucleosomes (histone octamer with DNA that is wrapped around) the basic unit of chromatin (Figure 1-2), can be heavily modified, especially at their N-terminal tails [8]. Known posttranslational modifications are among others phosphorylation, methylation, ubiquitylation and acetylation. These modifications can be set (by ‘writers’) and also deleted by histone-modifying enzymes (‘erasers’) and might either influence directly the compaction of chromatin, as discussed later for the acetylation at lysine 16 of histone H4, or indirectly by serving as recognition signals to attract specific effector proteins (‘readers’) (Figure 1-2 and [9]). Additionally, the conventional core histones can be replaced by histone variants. Some histone variants differ in their N- or C-terminal tails, making them differently susceptible to histone modifications. Histone variants can also affect the structure and properties, like mobility or stability of the nucleosome and in this way might influence the overall dynamics of chromatin [7]. Nucleosomes that contain the histone variant macroH2A, which is found on the inactive X chromosome in mammals (see below), were shown to be less mobile and hence might be an explanation for its ability to block the access of transcription factors to binding sites *in vitro* [10]. Another possibility to change chromatin structure is based on the action of chromatin remodeling enzymes. Those molecular machines, often organized in protein complexes with several subunits, use ATP-hydrolysis to alter the histone-DNA interactions [11, 12]. This can lead to a variety of effects, such as repositioning of nucleosomes along the DNA, exchange, assembly or eviction of histones. Hence, the intrinsic inhibition of nucleosomes can be overcome and allow access of the transcription machinery to its target sites.

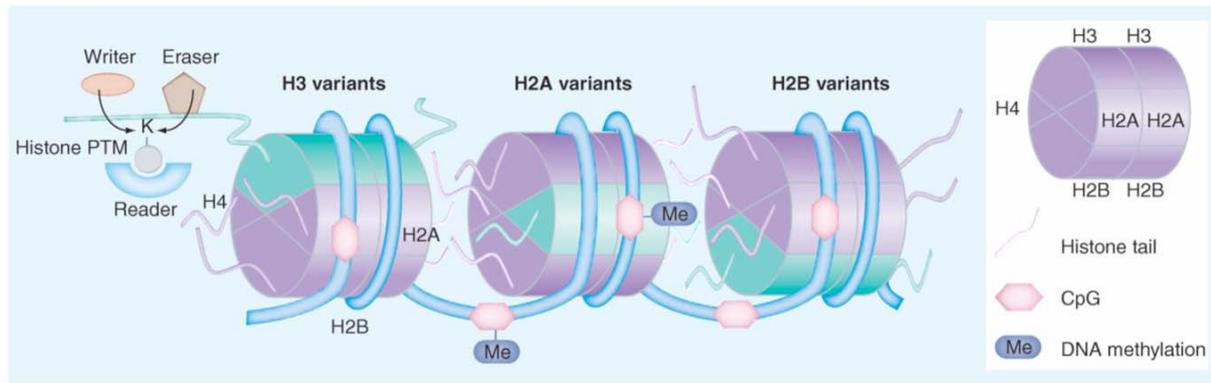


Figure 1-2: Principles that modulate chromatin structure and function. The nucleosome, which is composed of DNA (shown in blue) wrapped around the histone octamer (shown in purple) can be heavily modified. Posttranslational modifications (PTM) of histones are ‘written’, ‘erased’ and recognized by specific chromatin modifying factors. Histone variants (shown in green) can replace the conventional core histones (shown in purple) (modified, [8])

1.3 Dosage Compensation in *C. elegans*

1.3.1 Composition of the *C. elegans* DCC

Dosage compensation in *C. elegans* is brought about by a multi-protein complex called the dosage compensation complex (DCC). At least eight proteins are known to form the DCC and in contrast to the fly DCC non-coding RNAs are not known to be involved. Interestingly, several of those proteins (DPY-26, DPY-27 and DPY-28) show homologies to and even share (MIX-1) subunits of the condensin complex, an evolutionary conserved complex, which is needed for proper chromosome compaction and segregation during meiosis as well as during mitosis (Figure 1-3A and [13]). This homology suggests that also the DCC might act in a similar way: it could impose changes in chromatin structure, bringing different chromosomal compartments together and leading to their compaction, which consequently could reduce gene expression genome-wide [14, 15]. The DCC also includes subunits that are unique for this complex: SDC-1, SDC-2, SDC-3 and DPY-21. SDC-2 can be regarded as the key player in dosage compensation. Its expression decides whether the DCC is formed or not [16]. An X:A ratio of 1 (XX hermaphrodites) leads to repression, a ratio of 0.5 (XO males) to activation of expression of XOL-1, which in turn represses SDC-2 expression. SDC-2 is therefore only present in hermaphrodites. There it localizes specifically to the X chromosomes and not to the autosomes [16, 17]. Interestingly, SDC-2 acts independently, meaning it does not need any other DCC proteins for X chromosome binding [18]. The idea is that SDC-2 is the factor that initially distinguishes X from other chromosomes and subsequently promotes the assembly of the other DCC subunits. Because biochemical and

structural data are missing the mechanism of specific binding of SDC-2 remains unknown. Another protein also localizes to the X chromosomes, SDC-3, but it requires SDC-2, while all the other DCC subunits are dispensable [17]. Essential for dosage compensation (but not for its second role in sex determination) are two zinc finger domains of SDC-3. The presence of zinc fingers led to the assumption that SDC-3 might be able to bind to DNA via those domains and therefore be responsible for association of the DCC with the X chromosomes [17]. However, a direct involvement of SDC-3 and its zinc fingers in DNA binding has not been shown.

1.3.2 Recruitment-and-spreading model

The mechanism by which the DCC recognizes the X chromosomes is only poorly understood. Initially, huge X chromosomal fragments (1 - 5 Mbp) that were bound by the DCC in the native context were tested for their ability to recruit the DCC when detached from the X [19]. Regions were discovered that could indeed recruit the DCC autonomously, but some bound the DCC only weakly or not at all, despite the fact that they were able to tether the complex in the native X chromosomal environment. These findings lead to the 'recruitment-and-spreading' model (Figure 1-3B): According to that the DCC binds to few recruitment sites dispersed along the X chromosomes and then spreads *in cis* from those sites to adjacent regions, which by themselves lack recruitment ability [19]. In a follow-up study, Meyer and colleagues dissected these regions further and found small, discrete recruitment sites, termed rex sites (recruitment element on the X), which occur widely spread on the X and share at least two distinct DNA sequence motifs. Mutations in those motifs disrupted the recruitment of the DCC *in vivo*. But surprisingly, the rex sites are not uniquely found on the X chromosome, neither are they enriched in comparison to the autosomes. However, the rex sites occur highly clustered on the X [20]. Taken together it could be shown that DNA sequence is indeed a determinant for X chromosome binding, but it seems that it is the clustering of motifs, which creates a DCC binding site. The model of the rex sites was further refined by high-resolution ChIP-on-chip mapping [21]. A 10 bp long consensus sequence encompassing the two previously proposed motifs was identified, which is again not exclusive to but more clustered on the X. Moreover, the DCC was found to be preferentially bound at the 5'-end of genes (promoters), leading to the idea that the complex represses transcription of the genes it binds, maybe by interfering with transcription initiation or with

polymerase-coupled spreading [21]. Accordingly, the complex would first recognize rex sites and then locally ‘hop’ to close-by promoters of active genes [22].

However, recent findings obtained from another genome-wide analysis challenge this mode of action [23]: Unexpectedly, the proposed direct correlation between binding of the DCC and gene repression was not found. As a consequence, the authors propose an alternative model: The DCC is indeed recruited to rex sites, but then does not ‘hop’ and bind to 5’-end of genes where it represses transcription. It rather distributes to the newly identified dox sites (dependent on X sites), sites which are found preferentially in expressed genes but by themselves are not able to recruit the DCC. Because of the long distance that separates rex and dox sites (2 - 90 kbp), the DCC must somehow bring both kind of sites in physical vicinity, likely by altering the architecture of the X chromosome, for example by DNA looping. In that way the DCC would be able to spread and achieve gene repression over long distances and not locally at the genes it binds. Still, such a rearrangement of chromosomal domains has to be shown. Despite the progress that has been made, some important questions remain unanswered: First of all, it has not been shown which protein of the DCC is in fact binding to the consensus sequence. And secondly, since there are a lot of rex sites showing only weak similarity to the consensus sequence, it is clear that other determinants than DNA sequence must exist for forming a DCC binding site.

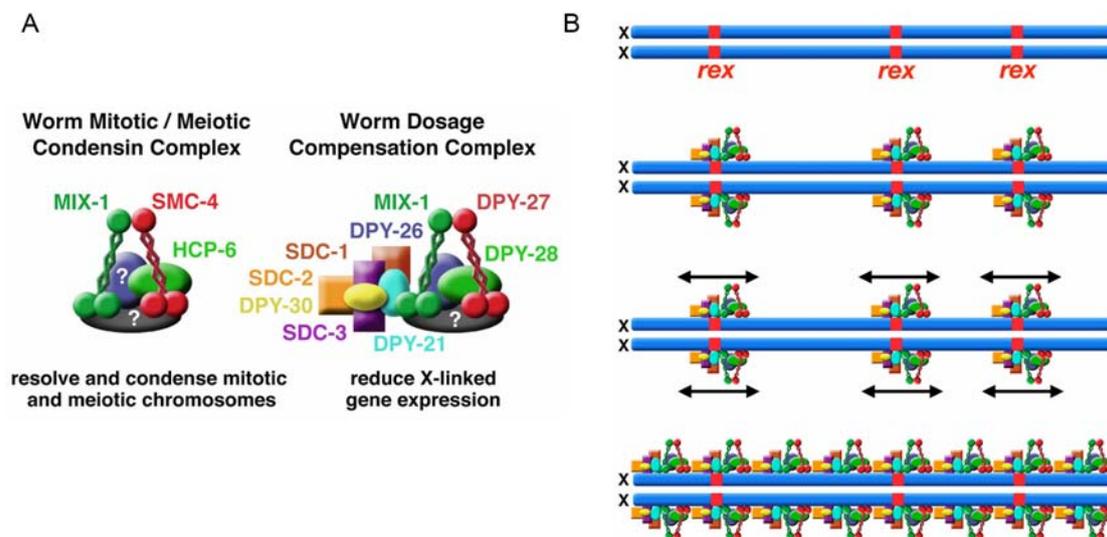


Figure 1-3: Composition of the *C. elegans* DCC and the ‘recruitment-and-spreading’ model. (A) The condensin complex of *C. elegans* is specialized for compaction and segregation of mitotic and meiotic chromosomes, whereas the *C. elegans* DCC is specialized for dosage compensation. Proteins of both complexes are color-coded to show homologous proteins. (B) Multiple rex sites appear clustered on the X compared to autosomes and hence recruit the DCC specifically to both hermaphrodite X chromosomes. The DCC then spreads in *cis* to adjacent X regions that lack rex sites and establishes X-chromosome-wide repression. Alternatively, the DCC might distribute from rex to dox-sites bringing distant chromosomal regions together leading to gene repression over long distances (not shown in this figure) (modified, [24]).

1.3.3 Autosomal binding sites

Genome-wide mapping revealed many more autosomal sites than previously judged from immunofluorescence studies [23]. The DCC binds to discrete, dispersed sites (mostly on promoters), but the typical consensus sequence and its clustering is lacking on the autosomes, demonstrating that the binding mode must differ from the one on the X chromosomes [23]. By depleting the SDC-2 subunit of the DCC, also autosomal gene expression was affected: In contrast to the X chromosomal genes, whose expression is decreased, the autosomal genes behave the other way round and get upregulated. Thus, the DCC functions not only on the X to adjust gene expression; it affects gene expression throughout the whole genome [23].

So far only one autosomal site has been studied in detail, the *her-1* gene. Since it promotes male development, it has to be shut down in hermaphrodites. This task is managed by the DCC, which binds to the promoter region of the *her-1* gene and represses its transcription around 20 fold [25]. This of course raises the question how the DCC can bind to all those different sites (X vs. autosomal) and additionally evoke different effects (2-fold and 20-fold down- and upregulation as well). It is speculated that the DPY-21 protein might be needed for fine-tuning transcription on the X. Its absence from the DCC on the *her-1* region might unleash the full repressive potential of the DCC and reduce transcription around 20-fold [22]. Whether these are common features important for all autosomal sites or just a special case for *her-1* has to be seen.

1.3.4 Mechanism of X chromosomal repression

The mechanism by which the DCC represses gene activation is not known. The study by Ercan et al. [21] suggests that the DCC is targeted gene-by-gene and accumulates at the promoter region and transcriptional start sites. Consequently, the authors propose that the regulation by the DCC occurs at the level of transcription initiation [22], for example by preventing accessibility of the promoter to the transcription machinery. It is known that negative supercoils in the DNA can support transcription initiation [26]. The DCC could, like it was shown for the condensin complex [27], introduce positive supercoils (or eliminate negative supercoils) and that way hinder transcription initiation [22]. On the other hand, Jans et al. [23] proposed that the DCC might act over long distances, which requires chromosomal rearrangements. Gene repression would therefore be achieved not gene-by-gene, but globally, by specific condensation of the entire X chromosome.

1.4 Dosage compensation in mammals

In contrast to fruit flies and the nematode *C.elegans*, dosage compensation in mammals does not involve a specialized protein complex, but is based on the combined action of the non-coding RNA Xist (X inactive specific transcript) and histone modifications [1, 28]. Such modifications influence the structure and state of chromatin, which consequently influences gene expression – in both directions: repression and activation as well. Especially post-translational modifications on the histones, such as methylation, acetylation or ubiquitinylation, or exchange of histone variants, as well as chromatin remodeling machines and DNA methylation are known key players (see above and [7]).

1.4.1 Selection of the inactive X chromosome

Usually the selection of one of the two X chromosomes to become inactivated happens randomly early in development. After the choice has been made, expression of the non-coding Xist RNA is increased. The transcribed Xist RNA binds *in cis* and coats specifically that X chromosome from which it is expressed [29], leading to its inactivation [30, 31]. The second X chromosome, however, has to stay active and therefore Xist transcription from the active X (Xa) is prevented by the action of another non-coding RNA called Tsix [32]. Tsix is transcribed antisense to Xist on the active X; conversely Tsix is downregulated on the inactive X (Xi). Interestingly, the expression of Tsix, in contrast to Xist, declines over time, meaning that Tsix is not required to keep Xist downregulated on the Xa [32]. The question arises how Tsix is able to regulate Xist. There is evidence for a role of Tsix in changing the chromatin configuration at the Xist promoter. It could be shown that the chromatin structure at the promoter adopted a more open and transcriptionally permissive state in absence of Tsix [33]. In addition, in the absence of Tsix, several histone modifications, which are known to be related with transcriptional activation, were found to be accumulated at the promoter [34]. However, the machinery to do so was not identified. And it is also not solved, whether simply anti-sense transcription or the RNA itself triggers these changes.

1.4.2 Mechanisms of X chromosome silencing

The presence of Xist RNA is a mark for the inactive X chromosome. But how this labeling translates into silencing a chromosome is only poorly understood. In general it is

thought that Xist might serve as a scaffold to bring repressive complexes to the Xi. However, not much is known about the factors that interact with Xist. Interestingly, Xist RNA was shown to have distinct domains that exhibit different functions. For silencing so-called A-repeats at the 5'-end are required, while coating is brought about by several domains dispersed over the entire RNA [35]. An interesting exception is the Xist RNA from humans, where the A-repeats carry out both functions, coating and silencing [36]. Still, silencing requires additional factors than just Xist. It was demonstrated that expression of Xist could not always induce silencing; it was dependent on the correct time and developmental stage, showing that the cellular context is critical for X inactivation.

Factors known to be involved in silencing are the polycomb repressive complexes PRC1 and PRC2. Both complexes were found on the Xi, where they set their characteristic histone modifications (H3K27me3 and H2AK119ub1) (Figure 1-4A). Surprisingly, both complexes and their marks were shown not to be essential for X inactivation [37, 38]. In addition, PRC1 and PRC2 were still to some extent found on the Xi, when the A-repeats of Xist (which are required for silencing) were deleted, demonstrating that the marks are not sufficient to trigger silencing. Several other repressive histone modifications are also associated with the Xi, while active marks like H3K4me and H4 acetylation are excluded [39]. Moreover, the histone variant macroH2A1 [40] gets incorporated, while H2A-Bbd is absent from the inactive X [41] and promoters of X-linked genes are methylated. DNA methylation is the modification that is in particular important for maintaining the X inactive through multiple rounds of cell divisions over the entire lifetime of the organism [42]. In summary, it became clear that it is not a single, but a whole series of different modifications with likely redundant functions that contribute to X chromosomal silencing (Figure 1-4A, C). It also seems likely that the responsible complexes identify the Xi via the Xist RNA, however nothing is known about the mechanism of their interaction.

Besides changes in chromatin state, X inactivation also involves a change in the nuclear organization (Figure 1-4). It was reported that the Xist RNA is involved in setting up a special silencing compartment, from which RNA-Polymerase II and the transcription machinery is excluded. During silencing X-linked genes move into the Xist-coated compartment, while genes that escape X inactivation are located outside it. Quite recently, two new players have been identified: SATB1 and SATB2 are known to be involved in the nuclear organization of chromosomal domains during T-cell development [43]. Now it was shown that Xist's restricted ability to induce silencing outside a certain developmental context

is dependent and can be overcome by the re-expression of those proteins [44], which links silencing and chromosome organization.

The fact that some X-linked genes are able to escape inactivation is mechanistically interesting. Besides the genes in the pseudoautosomal region, which have their counterparts on the male Y chromosome and therefore do not have to be compensated, there are reports that about 15% of human X-linked genes are escapees [45]. A more recent study, however, claimed that inactivation is almost complete [46]. An interesting explanation for how escaping is achieved is based on the LINE hypothesis: LINE-1 elements (long interspersed nuclear element) occur with different densities on the X and on the autosomes. Since the density on the X chromosomes is higher LINE-1 elements are thought to facilitate the spreading of Xist RNA specifically along the Xi and not on the autosomes (Figure 1-4); they consequently help to recruit genes into the silencing compartment, which leads to their inactivation [47]. Accordingly, having a low LINE-1 density would accordingly prevent genes from being silenced [48, 49].

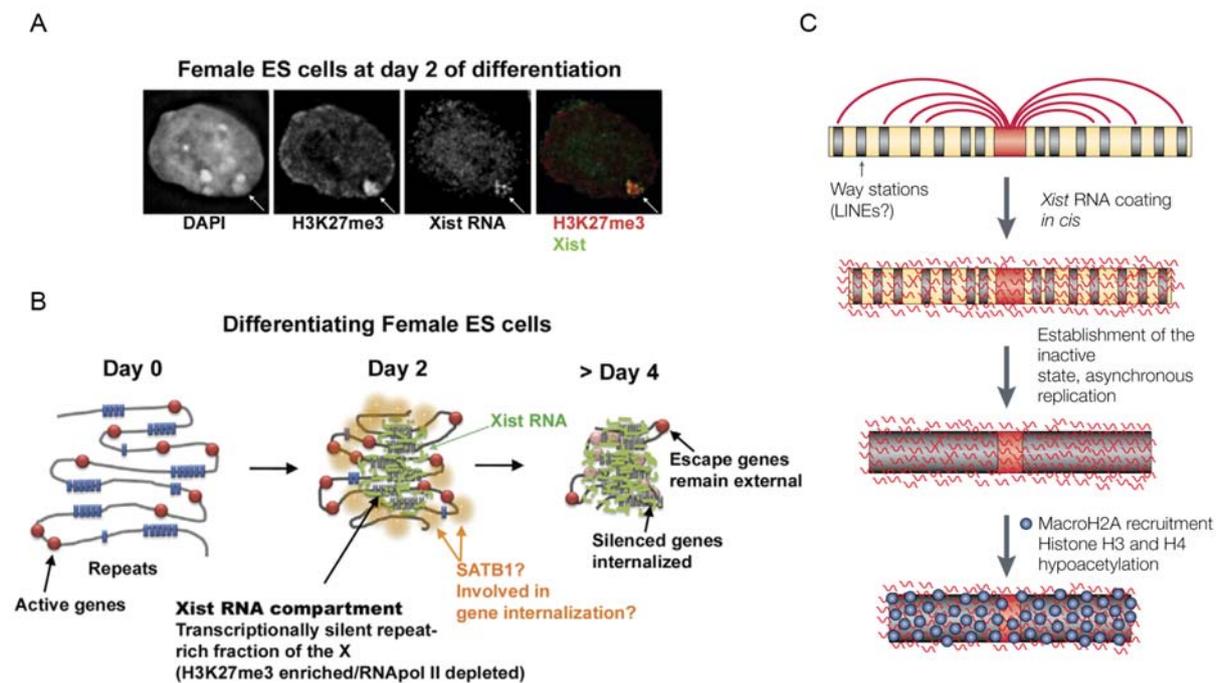


Figure 1-4: Mechanism of X inactivation in mammals. (A) Combined IF and RNA FISH on two days differentiated female mouse ES cells revealed that the Xist-coated transcriptionally silent compartment is enriched for marks such as H3K27me3. (B) Coating of the X chromosome with Xist RNA leads to formation of a transcriptionally silent compartment, which might require the matrix-associated proteins SATB1/2. Genes that are silenced are recruited into, whereas those that escape silencing are excluded from this compartment. (C) The Xist RNA spreads along the X chromosome, which might be facilitated by LINE-1 elements, to finally coat the entire X. Various histone modifications and incorporation of histone variants transform the Xist-coated X into an inactive chromatin state (modified, [28, 50]).

1.5 Dosage compensation in *Drosophila melanogaster*

1.5.1 Composition of the *Drosophila melanogaster* DCC

In the fruit fly dosage compensation is mediated by a large ribonucleoprotein complex, the dosage compensation complex (DCC) or alternatively called MSL complex (Figure 1-5). It consists of at least five proteins, which are essential for male but not for female viability, and were for that reason termed male-specific lethal proteins (MSL). For two of the MSL proteins, the histone acetyltransferase MOF and the helicase MLE, enzymatic activities were assigned, which are also important for non-dosage compensation functions [51]. The other MSL proteins (MSL2, MSL1 and MSL3) seem to be rather structural components of the DCC [52].

Several other associated proteins, which have additional, more general functions not only related to dosage compensation, were identified, for example the JIL1 kinase [53, 54] or the supercoiling factor SCF [55]. But how they mechanistically contribute to dosage compensation is not known. Besides protein subunits, the DCC also incorporates non-coding RNAs, roX1 and roX2 (RNA on the X). The two RNAs differ extremely in size and sequence [56]; nonetheless both are able to substitute for one another, suggesting that they have redundant functions [57]. Their contacts with the DCC occurs via MSL3, MOF or MLE – MSL proteins for which RNA binding activity has been shown *in vitro*.

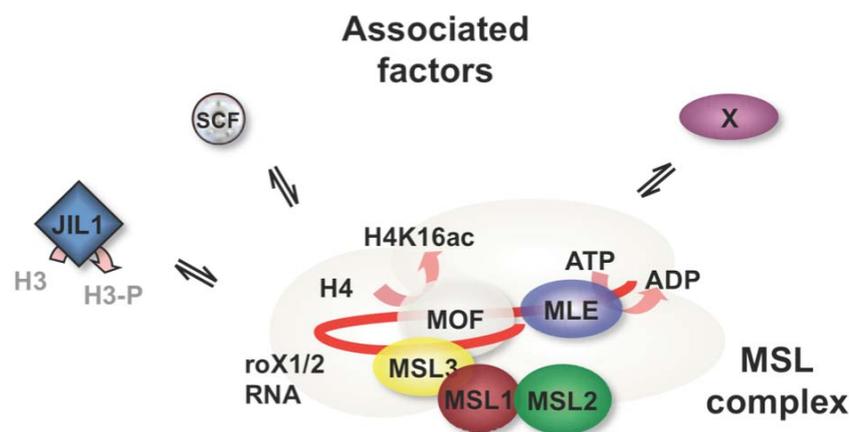


Figure 1-5: Schematic representation of the components of the DCC. The DCC or MSL complex is composed of five male-specific lethal proteins and two non-coding RNAs roX1 and roX2. Associated factors, like the super-coiling factor (SCF), the JIL1 kinase and nucleoporines (not shown in the scheme) and maybe other not yet discovered factors (X) are discussed to be involved in dosage compensation.

All the MSL proteins are also present in females, but since the expression of the key player MSL2 is inhibited by the regulatory protein SXL (sex lethal) in females the DCC assembles exclusively in the male sex [58, 59]. The DCC then covers almost the entire single male X chromosome leading to specific acetylation of histone H4 at lysine 16 [60], which is thought to be at least in part responsible for increased transcription of X-linked genes.

1.5.2 Mechanism of two-fold activation

It is of major interest to understand how the action of MOF in dosage compensation, that is the specific acetylation of H4K16, regulates transcription. The prevailing and most intuitive model suggests that MOF and the DCC work as direct activators of gene expression. Another idea termed ‘inverse dosage effect’ is that the DCC functions to down-regulate autosomal genes rather than to activate the X-linked genes [61]. This model is based on the idea that global regulators, such as MOF, which is also a general regulator in females [62, 63], would be removed from the autosomes and tethered to the X chromosome by the DCC. The result would be an accumulation of global activators on the X, which consequently would reduce transcription from the autosomes (Figure 1-6 and [64]). However, it was shown that the loss of DCC targeting directly correlated with reduced transcription of X-linked genes, while no effect was observed for the autosomal genes, which strongly favors the role of the DCC as a direct activator [65, 66].

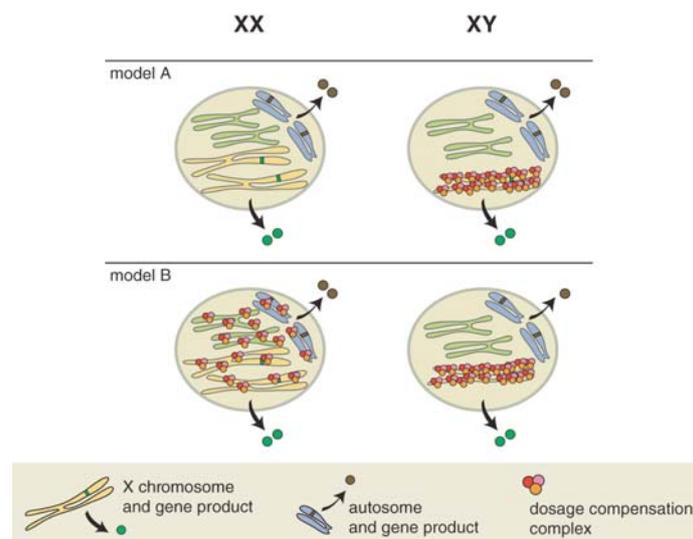


Figure 1-6: Different models of dosage compensation in flies. In model A the DCC directly activates gene expression on the male X chromosome leading to equal levels of gene products in males and females. In model B global activators are distributed equally on the X and the autosomes in females. By contrast, in males global activators are removed from the autosomes and accumulated on the X chromosome by the DCC (modified, [64]).

In general, gene activation is known to be linked to histone acetylation. This histone mark by itself is able to decondense chromatin structure, leading to a more permissive state for interacting factors [67]. The organization of higher-order chromatin, however, is also regulated by additional factors, such as chromatin remodeling complexes. For example, interfering with the function of ISWI, an ATPase found in several chromatin remodeling complexes of *Drosophila*, caused decondensation of the male X chromosome *in vivo*, which is accompanied by reduced levels of linker histone H1 that is incorporated into chromatin [68, 69]. On the other hand, acetylation of H4K16 antagonizes ISWI function *in vivo* and *in vitro* and therefore it is thought that H4K16ac is important to counteract the compaction of chromatin caused by ISWI [70]. With regards to X-linked genes H4K16ac levels are relatively low at promoters compared to 3'-end of genes. Also the binding of the DCC is preferred at 3'-ends, as shown by genome-wide mapping [71, 72]. For that reason it was proposed that gene expression is regulated at the level of facilitating transcription elongation and not by increasing the accessibility of promoters [71-73].

Still, it remains unclear how the H4K16ac mark translates into a precise two-fold activation. First insights into the mechanism came from biochemical analysis. Recombinant MOF protein was able to specifically acetylate H4K16 in nucleosomal arrays *in vitro*. As a consequence the chromatin-mediated repression of transcription could indeed be overcome but transcription was boosted far more than two-fold [74]. When MOF was in a sub-complex with MSL1 and MSL3, the substrate-specificity was improved [75]. In addition the histone acetyltransferase activity was even further increased, indicating that for MOF's full function the other MSL proteins are required. Interestingly, the roX RNAs are not involved in acetylation *in vitro* [75], suggesting a role for the RNAs rather in complex assembly or targeting but not in activation. Nevertheless, for achieving two-fold activation, a feedback mechanism would be required, perhaps by an additional repressive function within the DCC. Clues for such a repressive function could come from dissecting other MOF containing complexes. In fact MOF was shown to be present not only in male but also in female flies and even in human cells [62, 63]. Biochemical purifications show that MOF is also present outside the DCC in other complexes ([63] and M. Prestel, P. Becker lab, manuscript submitted). Moreover, it was postulated that the two distinct MOF-complexes (in flies) also exhibit different binding patterns: While the classical MOF complex, the DCC, is found preferentially at the 3'-end of genes on the X, an alternative MOF complex prefers binding to promoters of autosomal genes [76]. This indicates that the two MOF complexes might also mechanistically function in different ways. By comparing their composition and their

operating principle one might gain more insight into how gene activation by MOF can be restricted to two-fold.

1.5.3 Principles of X chromosome targeting

The mechanism by which the DCC selects specifically the X and not the other chromosomes is not well understood. Genome-wide high-resolution mapping revealed that the DCC does not bind within promoter or enhancer sequences, but is localized to the coding sequence of active genes (Figure 1-8A), indicating that the complex acts rather at the level of transcriptional elongation than initiation [71, 72]. The similar binding patterns of the DCC and of a histone modification known to be involved in active transcription, the trimethylation of H3K36, and the requirement of this modification for recruitment of target genes [77, 78], led to the idea that this histone modification might serve for the DCC as a recognition feature. Since almost all active genes are compensated [4], an active mark would indeed be a simple signal for the DCC to find its target genes. On the other hand this recognition mark is not specific for the X chromosome, and consequently, other determinants must be involved.

Early studies on dosage compensation favored the idea of an involvement of DNA sequences. Normally, the DCC assembles only in the male sex, but if the key player MSL2 is ectopically expressed in females, then the DCC forms also there and binds to the X chromosomes [59]. All the information necessary for binding must therefore be represented within the X chromosomes, leading to the idea of specific DNA sequences that are unique for the X and that serve as recognition sites for the DCC. Hundreds of sites where the complete DCC is bound can be visualized (but in low resolution) by immunostaining of giant polytene chromosomes from the salivary gland of larvae. The number of sites is dramatically reduced to 30 - 50 when MSL3, MLE or MOF are absent (Figure 1-7), showing that a subcomplex consisting of only MSL2 and MSL1 is sufficient for recognizing X chromosome binding sites [59, 79, 80]. These sites were termed 'chromosome entry sites' (CES) and were thought to be the initial recruitment sites for the MSL2-MSL1 subcomplex. Subsequent to this recruitment the association of the other MSLs and of the roX RNAs leads to the formation of the full complex, which spreads to neighboring chromosomal regions and finally covers all secondary sites on the entire X [81].

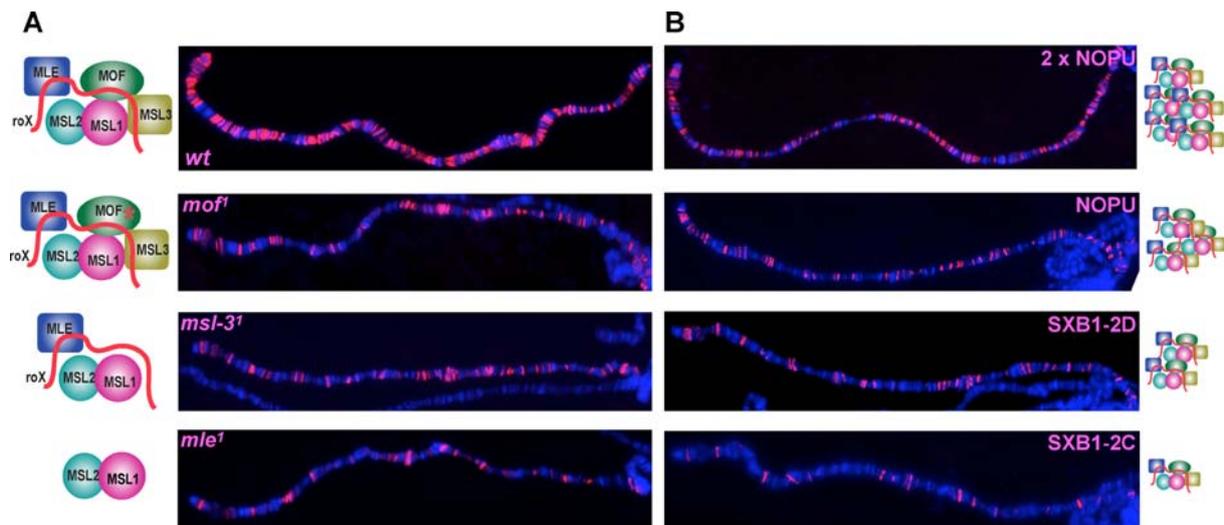


Figure 1-7: Hierarchy of DCC binding sites. Polytene chromosomes from *Drosophila melanogaster* larvae stained with MSL1 antibodies show binding of the DCC to a defined pattern of sites. (A) Partial or nonfunctional complexes bind to a subset of sites, with the MSL2-MSL1 subcomplex being the minimal requirement for binding to HAS. (B) Different MSL2 expression constructs (NOPU and SXB1-2) give rise to decreasing levels of MSL2 in females, which leads to decreasing levels of complex and to the same binding pattern as in (A), suggesting different affinities for the binding sites (modified, [82]).

The entry site model was based on the characterization of two CES within the *roX1* and *roX2* genes. Indeed, when the *roX* genes were inserted into autosomes, the DCC was bound to those insertion sites and moreover spreading of the DCC to neighboring, secondary sites on the autosomes was observed [83, 84]. However, by analyzing more binding sites it became clear that also X chromosomal sites that do not have the DCC bound (which are consequently not CES) could recruit the complex when inserted into autosomes and that the other way round, parts of the autosomes when translocated to the X to a nearby entry site could not attract the DCC [85, 86]. These findings led to a refinement of the characteristics of binding sites. By continuously reducing the level of the complex *in vivo* it was shown that fewer and fewer sites were bound by the DCC, leading to the idea of a hierarchy of binding sites [82, 85, 87]. According to this affinity model (Figure 1-7), the DCC or a subcomplex binds first to sites with highest affinity and as the levels of complex increase, sites with middle and low affinity get successively bound. In the absence of the full complex, the partial MSL2-MSL1 subcomplex can still bind to the X, but only to the high affinity sites, whereas the sites with lower affinity are bound when the complex is fully assembled and its concentration increases [64, 82]. However, some X-chromosomal sites that are strongly bound by the DCC were not able to recruit the complex when inserted on an autosome [88]. In these cases it is possible that the strong binding on the X is not due to the fact that these sites are real high affinity sites, but instead are low affinity sites that profit from high

concentrations of the DCC on nearby high affinity sites. A consequence of the affinity-based targeting is that at a certain concentration of complex (achieved by overexpression of MSL proteins) also non-target sites, such as autosomal sites get bound, showing that the level of MSL proteins and roX RNAs has to be tightly regulated to restrict the action of the DCC on the X chromosome [87, 89, 90].

Recent data proposed an additional feature required for targeting, namely the involvement of transcription. When high affinity sites from the X were inserted into autosomes, the DCC could be recruited, but only when either transcription through these inserted genes was allowed or when the insertion site was inserted as multiple copies. This suggests that transcription is not absolutely required, but it facilitates the recruitment by increasing the accessibility to a binding site (Figure 1-8 and [91]).

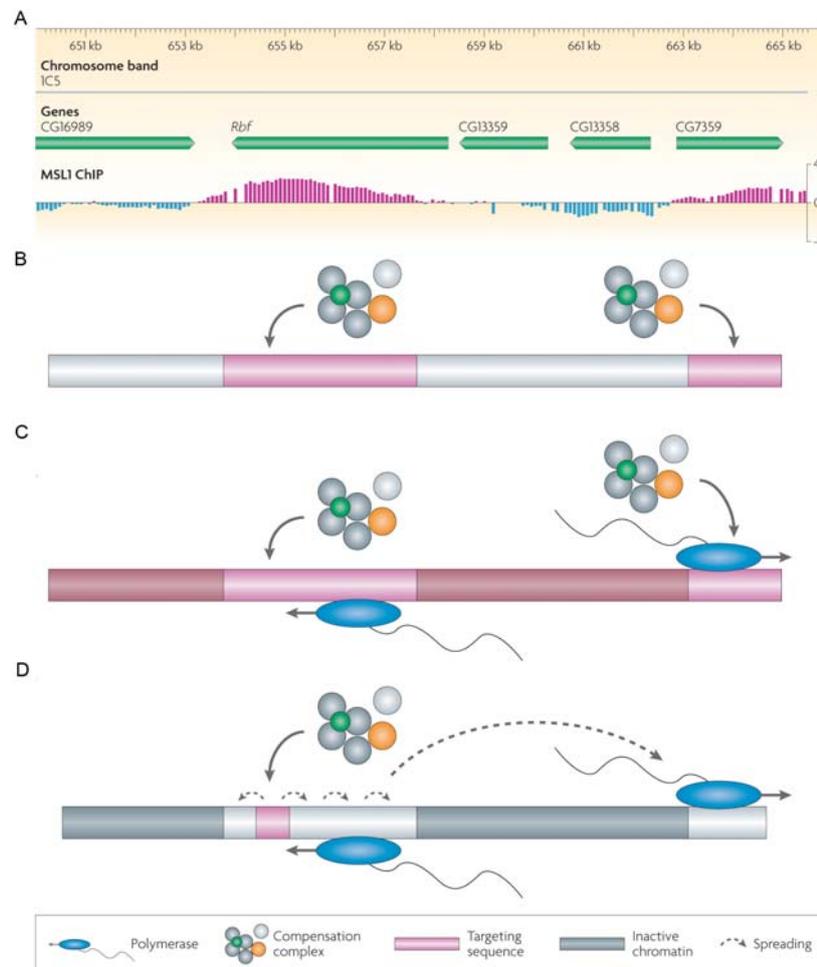


Figure 1-8: Principles of X chromosome targeting by the *Drosophila melanogaster* DCC. (A) Genome-wide high-resolution mapping of DCC binding sites on the male X chromosome. Targeting might be achieved by the following models: (B) The DCC recognizes specific DNA sequences (HAS), which carry a consensus motif (e.g. MRE), based on binding affinity. (C) Factors associated with active transcription, e.g. RNA Polymerase II or H3K36me3 are required to expose specific binding sites. (D) A two-step model: The DCC initially recognizes primary target elements and then spreads *in cis* into neighboring regions independent of sequence, instead using the transcription machinery (modified, [3]).

1.5.4 High affinity sites and DNA recruitment elements

The concept of a hierarchy of binding sites raises the question of what might be the determinants and common features characteristic for high affinity sites (HAS). First description came from *in vivo* studies of single HAS. Well-characterized binding sites are the nucleosome-free DNase hypersensitive (DHS) sites within the *roX* genes [83, 84], the 18D HAS [86] and the *Smr* and *Tao-1* gene [82]. Common for all those sites is their ability to recruit the DCC when they are inserted into autosomes, which suggests that DNA sequence could be a major prerequisite for defining a HAS. Nevertheless, the relatively large size (about 200 bp) and the small number of identified sites made it impossible to derive a common DNA sequence motif, indicating that there might be different classes of binding sites [82, 86]. The two *roX* DHS anyway seem to be rather atypical HAS, because they require the MLE helicase and one of the two *roX* RNAs for recruitment of the complex [83, 84], whereas the other high affinity sites can be bound in the absence of *roX* RNAs [82, 86]. Also transcription seems not to be a feature of HAS, since some sites, e.g. the 18D HAS, are not transcribed at all [86], while others, e.g. the *roX* genes, indeed produce transcripts (the *roX* RNAs), on the other hand the transcription *per se* from *roX* genes is not needed for DCC recruitment to the *roX* HAS [57, 83].

The dissection of HAS in more detail allowed to further narrow down the HAS to smaller X chromosomal fragments, so-called DCC binding fragments (DBF), which are sufficient to recruit the DCC *in vivo*. From the around 15 DBFs it was possible to determine several GA-rich DNA elements, which appear clustered within HAS [82]. Nevertheless, the still relatively large size of the DBFs (several thousand kilobases) complicated further refinements of potential consensus sequences.

In order to identify minimal binding elements within the DBFs that are recruiting the DCC, a reporter gene assay was developed (Figure 1-9 and [92]). In this assay several candidate binding sites derived from the DBFs were cloned in front of a minimal promoter driving transcription of a *luciferase* reporter gene. The plasmid is transfected into male *Drosophila melanogaster* SL2 cells together with a plasmid that allows expression of an MSL2-VP16 fusion protein. Upon binding of the MSL2 fusion protein to the potential binding element, the VP16 activation domain leads to the expression of the *luciferase* reporter gene. Because of the VP16 activation domain transcription is even boosted, allowing the detection of also weak protein-DNA interactions. The screening identified several minimal targeting elements, but sharing only little similarity and with no obvious consensus sequence. Among others, a small 40 bp sequence derived from DBF12 within the *Smr* gene, the so-

called DBF12-L15 element was identified as a recruitment site. It contains a GA repeat and interestingly, mutating the GA motif to a stretch of thymines (DBF12-L18) destroyed the function as a DCC recruitment site (Figure 1-9). Furthermore when trimerized and inserted into autosomes it created one of the strongest recruitment sites seen so far [92]. Consequently, it is thought that DCC binding sites might be composed of clusters of several degenerative sequence motifs.

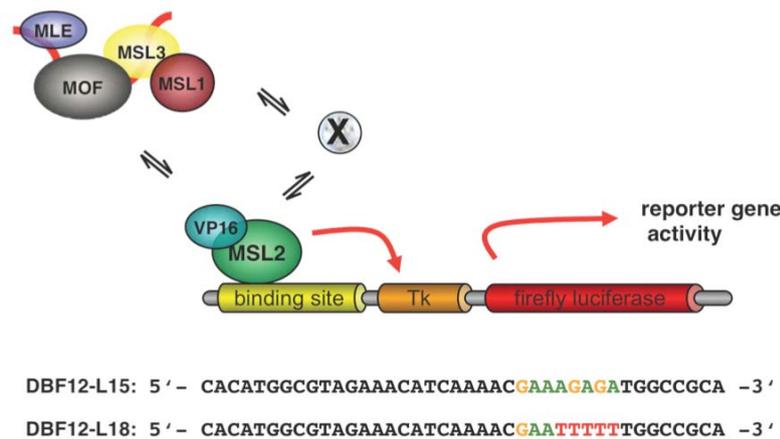


Figure 1-9: Reporter gene assay to identify DCC binding sites. MSL2-VP16 activator construct and a *luciferase* reporter gene, which carries a candidate binding site, are transfected into male *Drosophila melanogaster* SL2 cells. Upon binding of MSL2-VP16 to the binding site expression of the reporter gene is activated. Binding of MSL2-VP16 might either be direct or indirect (via an unknown factor X) and might require the DCC subunits. Among others a 40 bp DNA sequence (DBF12-L15) was identified to be the minimal element necessary for MSL2-VP16-mediated reporter gene activation. An DNA element (DBF12-L18) carrying mutations in the GA-rich sequence abolished reporter gene activation.

In order to facilitate the discovery of common features, the high-resolution analysis were recently extended to map exclusively HAS [93, 94]. To ensure that binding is restricted preferentially to HAS, the complex concentration in SL2 cells was reduced by applying RNAi against individual MSL proteins, combined with less stringent crosslinking conditions [94]. Alternatively, MSL3 mutant flies were used, where the lack of MSL3 allows only the formation of a subcomplex consisting of MSL2-MSL1 and roX-bound MLE [93]. Both strategies led to the identification of a subset of around 150 binding sites, which were still bound under the mentioned conditions and therefore meet the criteria of a HAS. By using bioinformatic analysis, again a GA-rich motif, which is similar to the one discovered in the reporter gene screen [92], was found to be common in HAS and was proposed as a MSL recognition element (MRE). However, the MRE was only slightly enriched on the X chromosome compared to autosomes, and also not every MRE found on the X was bound by

the DCC [93, 94]. This indicates that such a motif alone might not be sufficient to create a HAS. Interestingly, the identified motif occurs on sequences that are depleted of histone H3, indicating that initial DCC binding requires a naked, histone-free DNA template [93, 94].

1.5.5 The MSL2-MSL1 targeting machinery

Not only are the requirements for a HAS not fully understood, but in addition the factors that recognize such binding sites are not clearly identified. Not all subunits of the DCC are needed for initial recruitment. The roX RNAs seem to be dispensable since the overexpression of MSL proteins can overcome the lack of roX RNAs and consequently successful localization of the DCC to the X and acetylation of H4K16 was observed [89]. On the other hand the absence of roX RNAs led to increased DCC binding to autosomes and to the chromocenter, showing that roX RNAs are needed to ensure proper targeting [89]. However, it is clear that the targeting domains reside within the protein components of the DCC. The DCC subunits MSL3, MOF and MLE were shown to be not required; a subcomplex of MSL2 and MSL1 is sufficient to recognize the HAS on the X chromosome. Binding to the HAS depends on both proteins, the individual subunits by themselves are not able to bind to the X [80]. One exception is known for the *roX* genes, where MSL2 does not need any other MSL proteins and is alone sufficient to drive the transcription of roX RNA [95]. How this function outside of the DCC is achieved is unknown, but the mechanism seems not to rely on autonomous binding of MSL2 to the roX HAS, since deleting it had no effect on roX transcription [95]. The MSL2-MSL1 subcomplex and the HAS do not only ensure the proper selection of the X chromosome, they were recently found to lead to a clustering of HAS, which in turn sets up a male-specific X chromosome conformation [96]. An involvement of nucleoporines, which were previously thought to act in dosage compensation [63], could be excluded.

It is also mysterious how the direct contact to the X chromosome is achieved, whether only one protein establishes contacts to DNA or whether both proteins together form a common binding domain. Neither for MSL1 nor MSL2 a typical DNA binding domain is known or predicted (Figure 1-10). MSL1 is composed of several domains, which mediate the interactions with the other MSL proteins, showing its central role in complex assembly [97]: The coiled-coil domain mediates interaction with MSL2, the PEHE domain with MOF and the C-terminal part with MSL3 [75]. The short N-terminal region was suggested to be involved in X chromosome association *in vivo*: When the first 84 aa were deleted, MSL1

could still interact with MSL2, but MSL1 was not recruited anymore to the X chromosome, but strikingly, MSL2 was still bound to the X [97, 98]. The domain within MSL2 that could be responsible for X chromosome binding is not known. The N-terminal RING finger was shown in yeast two-hybrid assays to mediate the interaction with MSL1 [79], whereas a conserved CXC domain has no described function yet, and a sequence rich in prolines and basic amino acids (Pro/Bas patch) is thought to be important for incorporation of roX RNA by an unknown principle [99]. Taken together, these findings strongly favor the idea that indeed both proteins are co-dependent for, but on the other hand it is the MSL2 protein that mediates the interaction with the X chromosome. Such a contact to DNA might be based either on typical protein-DNA contacts characterized by an equilibrium between DNA-bound and unbound state, or involve a stable, topological ring formation around the X chromosomal DNA, as suggested from the very slow turnover of MSL2 measured by photobleaching experiments [100]. Whether the interactions are direct or involve an unknown additional factor has to be addressed, as well as the biochemical proof of the MSL interactions with DNA.

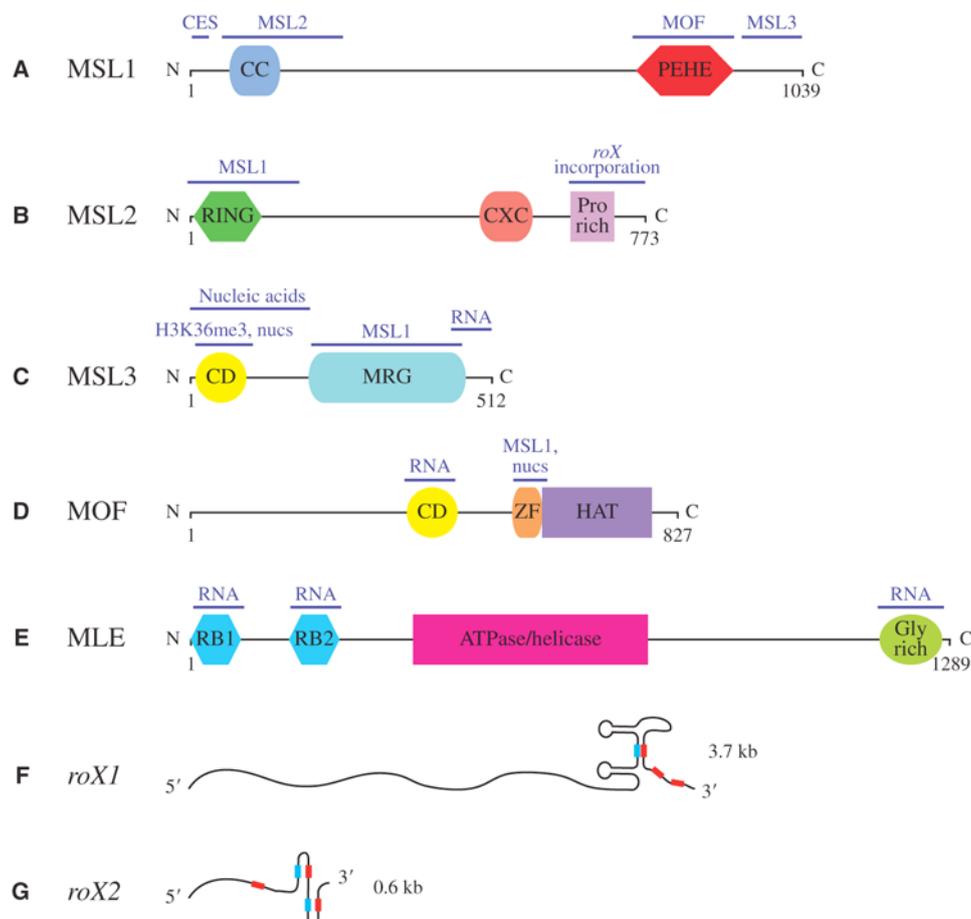


Figure 1-10: Components of the DCC. (A)-(E): The protein subunits, their domain organization and function of the individual domains are shown. (F)-(G): The two non-coding roX RNAs and conserved roX box elements are depicted in red and blue (inverse roX box) (modified, [52]).

1.6 Protein-DNA interactions

The selection of the correct DNA binding site from the vast amount of other binding sites in the genome is a mechanistically challenging task for all DNA binding proteins. For some aspects, like the organization of DNA by chromosomal proteins, it is, however, not desirable to have a special target sequence, but instead the protein should bind to DNA with minimal sequence specificity [101]. To manage the problem of specificity, numerous different DNA binding domains have evolved and are not only found in the family of transcription factors, but also in proteins that mediate recombination or DNA cleavage [102]. In the following section basic principles of DNA recognition are illustrated by examples.

1.6.1 DNA binding domains

One of the most frequently found motifs is the helix-turn-helix (HTH) structure, occurring for example in the λ repressor (Figure 1-11A), λ Cro protein or in the *E.coli* CAP protein [103]. Two helices, which vary in length between the individual family members) are linked by a tight bend, whereby the second helix inserts into the major groove of the DNA duplex, where both base and DNA backbone contacts are formed [104, 105]. The orientation and position of the second helix in the major groove varies drastically within the HTH family, leading to different ways of docking of the HTH against the DNA and establishment of different interactions (Figure 1-12A). Not only helix 2 and 3 make contacts with the major groove, but also regions outside the HTH motif are needed for binding, e.g. the flexible linker arm at the N-terminus of the HTH motif of the λ repressor [106]. It is therefore not surprising that the HTH motif is not able to function by itself and is always a part of a larger DNA binding domain [103].

The classes of basic region-leucine zipper (bZIP, Figure 1-11B) and helix-loop-helix (HLH) proteins are characterized by two distinct subdomains: A basic region (rich in arginines and lysines), which forms contacts with the DNA and a dimerization region, either an HLH motif or a leucine zipper [107-109]. The leucine zipper shows repeats of leucines over 30 - 40 residues and forms two parallel helices in a coiled-coil arrangement [108]. Different bZIP proteins are able to form also heterodimers via their leucine zipper helices. Such mixed dimers, for example the proteins Fos and Jun form the transcription factor AP-1, expand the repertoire of target sequences and allows regulating protein function [110]. By contrast, HLH proteins have a different dimerization interface composed of two helices

separated by a loop [109]. For both protein families the interaction with the DNA is achieved by the basic region. It adopts a long helical structure, which inserts into the major groove. This helical arrangement is only structured in the presence of DNA, otherwise it is disordered [111]. Interestingly, the basic region of GCN4 does not need the classical leucine zipper motif, also a simple disulfide bond between the two helices is sufficient to allow dimerization and consequently binding of the specific target sequence is achieved [111].

Besides the major groove also the minor groove can serve for interaction, whereby sufficient distortion of DNA is a prerequisite (Figure 1-11C), which is at first energetic costly but can be compensated by the favorable contacts that are achieved [102]. Examples are found in the family of lacI proteins, e.g. the lac repressor or the purine repressor dimer (PurR). Each PuR monomer forms two separate binding modules (Figure 1-11C): A classical HTH motif to contact the bases in the major groove and a two-turn hinge helix to contact the bases in the minor groove [112]. The hinge helices associate via hydrophobic interactions and lead to unwinding and bending of the DNA away from the protein, opening the minor groove to enable the contact between side chains of the hinge helix pair and bases in the minor groove. Hydrophobic side chains (leucines) of the hinge helices assist in keeping the 45° kink in the DNA open. An even more extreme example of promoting DNA distortion is the TATA binding protein (TBP). Its huge ten-stranded β -sheet induces drastic unwinding and bending, which allows the concave surface to interact with bases in the now opened minor groove [113, 114]. Similar to PurA, the distortion is mainly due to the intercalating hydrophobic side chains (phenylalanine) into the DNA.

Other major structural motifs – the zinc finger proteins – can have additional functions, which are not only restricted to DNA binding. Examples and their mode of action are discussed below in more detail (Figure 1-11D, E and Figure 1-13).

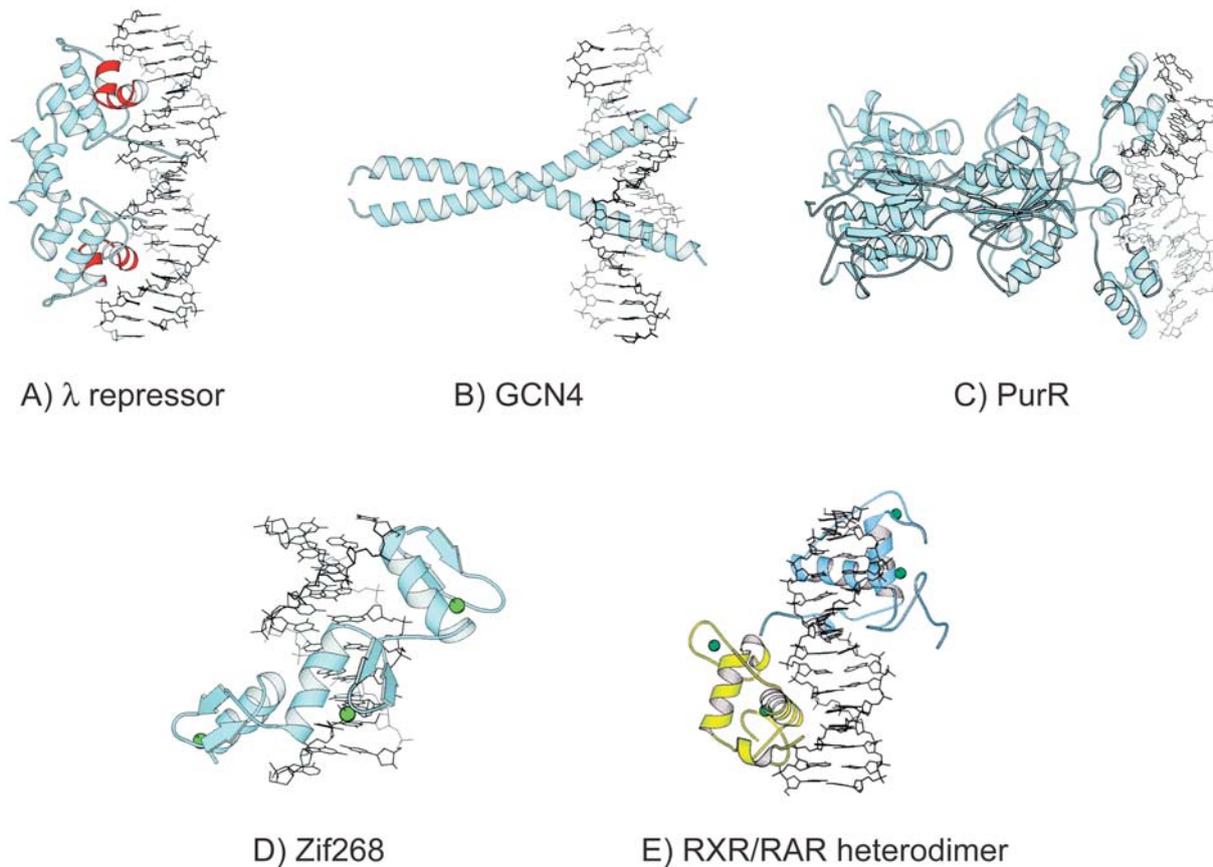


Figure 1-11: Representative examples of different DNA binding folds. (A) Helix-turn-helix structure (in red) of the bacterial λ repressor. (B) Basic region-leucine zipper of GCN4. (C) LacI member: purine repressor dimer. (D) Cys₂His₂-type zinc finger (coordinated zinc ion in green) of zif268 transcription factor. (E) Two Cys₂Cys₂-type zinc fingers (coordinated zinc ion in green) in each monomer of the DNA binding domain of the nuclear receptor heterodimer RXR (in blue) and RAR (in yellow) (modified, [102]).

1.6.2 Principles of DNA recognition

The contacts with the DNA bases are central for site-specific binding. As seen above, these contacts occur mainly with the bases of the major groove, which is not only larger, but offers also more potential sites for hydrogen bonding than the minor groove. In the minor groove, the combinations of hydrogen bond donors and acceptors are quite limited, since A-T pair is similar to T-A and G-C is similar to C-G. However, it is impossible to derive a simple recognition code [115]; especially when one considers the different ways of how an α -helix can be positioned in the major groove (Figure 1-12A), the flexibility of many side-chains and the modest energetic costs of small DNA distortions. Nevertheless, some recurring themes of recognition can be observed [115]: The majority of sequence-specific interactions are thought to come from hydrogen bonding. In contrast to a single hydrogen bond, bidentate interactions, in which a single side chain (e.g. of arginine) establishes two hydrogen bonds (e.g. with the

guanine base), contribute to a higher specificity (Figure 1-12B, C). Since guanine is the only base in the major groove that contains two hydrogen bond acceptors, a substitution by any base would reduce DNA binding affinity. Other pairings like glutamine or asparagine with adenine and lysine with guanine are also found to be important for site-specific recognition (Figure 1-12C and [116]). Van der Waals interactions (e.g. between thymine methyl group and protein side-chains) play a minor role, since they lack a directional requirement (Figure 1-12B and [102]). The DNA backbone itself does not contain sequence information; nevertheless the contacts with the DNA backbone (achieved by basic or neutral hydrogen-bonding side chain that contact phosphodiester oxygens) also contribute to sequence-specificity, mainly by positioning the protein in order to allow the establishment of the correct hydrogen bonds [103].

Contacting the DNA backbone and the minor groove are also main features of sequence-independent DNA binding proteins. Those proteins, e.g. the chromosomal protein HMG-D or the archaeal DNA chaperone Sac7d, have to avoid sequence-specific contacts and instead establish sequence-neutral interactions with invariant DNA positions, such as the sugar-phosphate backbone, to obtain affinity without specificity [117, 118]. Especially water molecules are involved in allowing variation in the recognized sequence, since they can cushion the interface, allowing electrostatic interactions to dominate. Moreover, key residues (see above) can be bridged by water molecules (Figure 1-12D), so that space is filled up and additional hydrogen bonds are offered, which add stability but not specificity [101].

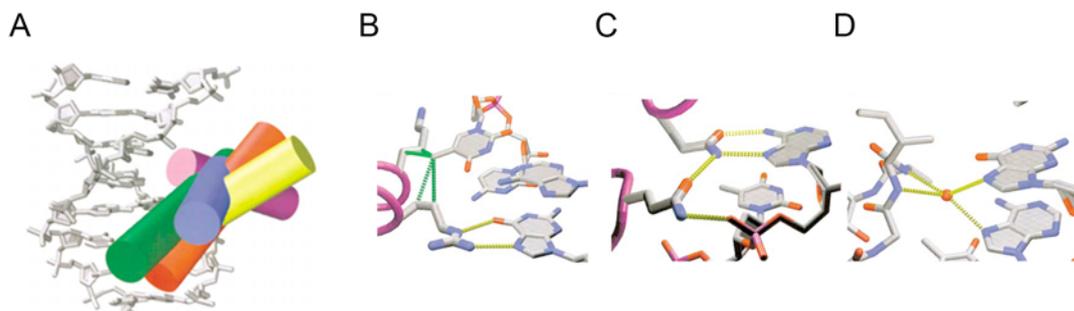


Figure 1-12: Principles of protein-DNA interactions. (A) Different orientations of α -helices within the helix-turn-helix family inserted in the major groove of DNA. (B)-(D) Details of protein-DNA contacts. (B) Bidentate hydrogen bonds between arginine side chain and guanine base (yellow dashed lines). Hydrophobic contacts between amino acid side chains to a thymine methyl group (green dashed line). (C) Glutamine side chain forms bidentate contacts with adenine base and additional contacts to another glutamine side chain, which in turn contacts a phosphate group of the DNA backbone (λ repressor-DNA complex). (D) Water-mediated hydrogen bonds within the helix-turn-helix structure of the Trp repressor bound to DNA (modified, [102]).

1.6.3 Zinc finger proteins

Zinc finger structures are quite common in higher organisms, with more than 15000 classical zinc finger domains predicted in around 1000 human proteins and more than 20 additional related classes [119, 120]. The only common property of all those very diverse structures is simply that their protein fold is maintained by coordinating zinc ions via cysteine or histidine residues (Figure 1-13). The classical Cys₂His₂-type zinc finger was first discovered in TFIIIA [121, 122] and crystal structures of related transcription factors, such as zif268, revealed that it is the most simple zinc finger: It is composed of a short α -helix, two antiparallel β -sheets and a zinc ion coordinated by two histidines (in the α -helix) and two cysteines (near the turn in the β -sheet) (Figure 1-11D and [123]). Such a motif occurs mostly as multiple copies within a protein and uses the α -helix to insert into the major groove of the DNA, whereupon the next zinc finger wraps around the double helix and makes another DNA contact [124]. The superfamily of nuclear hormone receptors binds their target DNA sequence as homo- or heterodimers. The DNA binding domain of one monomer is composed of a pair of zinc finger motifs [125]. Each motif coordinates a zinc ion by 4 cysteine residues (Cys₂Cys₂-type) ensuring the formation of a peptide loop between two of the cysteines and an adjacent α -helix (Figure 1-11E). The first helix fits into the major groove to establish base contacts, while the second helix provides on the one hand phosphate contacts with the DNA backbone and on the other hand the dimerization interface [126].

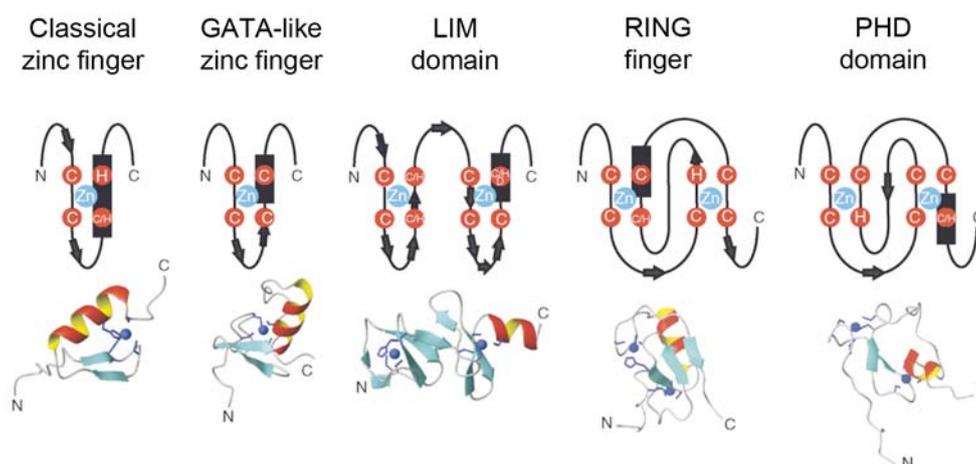


Figure 1-13: Topology and structure of different zinc finger domains. β -sheets are represented as black or light blue arrows, α -helices as black boxes or red helices and coordinated zinc ions as blue spheres. Details are found in the text (modified, [127]).

Besides those two mentioned zinc fingers, also other various cysteine or histidine-rich motifs exist, varying drastically in sequence, clustering and spacing (Figure 1-13). As a consequence, it is impossible to predict their structure or function and their role in DNA binding has to be shown experimentally. In fact, several zinc finger structures, including the classical Cys₂His₂-type, have additional functions in mediating protein-protein [127, 128] and protein-RNA interactions [129, 130]. The TFIIIA transcription factor not only recognizes the 5S rRNA gene, where it regulates transcription, but also the 5S rRNA to stabilize the RNA until usage for ribosome assembly [131, 132]. All but the fourth and sixth Cys₂His₂-type zinc finger of TFIIIA directly interact with DNA, whereas zinc finger four and six present non-DNA binding spacers to properly position the DNA binding zinc fingers (Figure 1-14A and [124, 133]). By contrast binding to the 5S rRNA requires direct contacts of the zinc finger four and six, where they recognize elements in the loop region of the RNA using the N-terminal ends of their α -helices (Figure 1-14B and [134]). The location of the α -helix and the base contacts that are made with RNA differ from those that are made with DNA by the other DNA binding zinc fingers of TFIIIA, however, some crucial base contacts are found in both RNA and DNA [124, 134]. Interestingly, the zinc finger five is involved in both DNA and RNA binding (Figure 1-14C, D). It binds DNA in the typical Cys₂His₂ zinc finger fashion using the side-chains of the α -helix to establish base and DNA backbone contacts in the major groove (Figure 1-14C and [124]). Binding to the equivalent RNA sequence, however, does not involve base contacts, instead contacts of the side chains of the α -helix with the backbone of the major groove of the RNA are established, which, notably, are the same found in complex with DNA (Figure 1-14D and [134]). Hence, the classical zinc finger five recognizes both DNA and RNA and thereby makes use of overlapping sets of side chains to contact nucleic acids. It has to be noted that the principles by which proteins recognize DNA or RNA are influenced by fundamental differences in the structure of the nucleic acids. Most importantly, the major groove of the RNA is too narrow to accommodate a α -helix and in contrast to DNA, the tertiary structure of RNA is another significant factor that contributes to the formation of protein-RNA complexes.

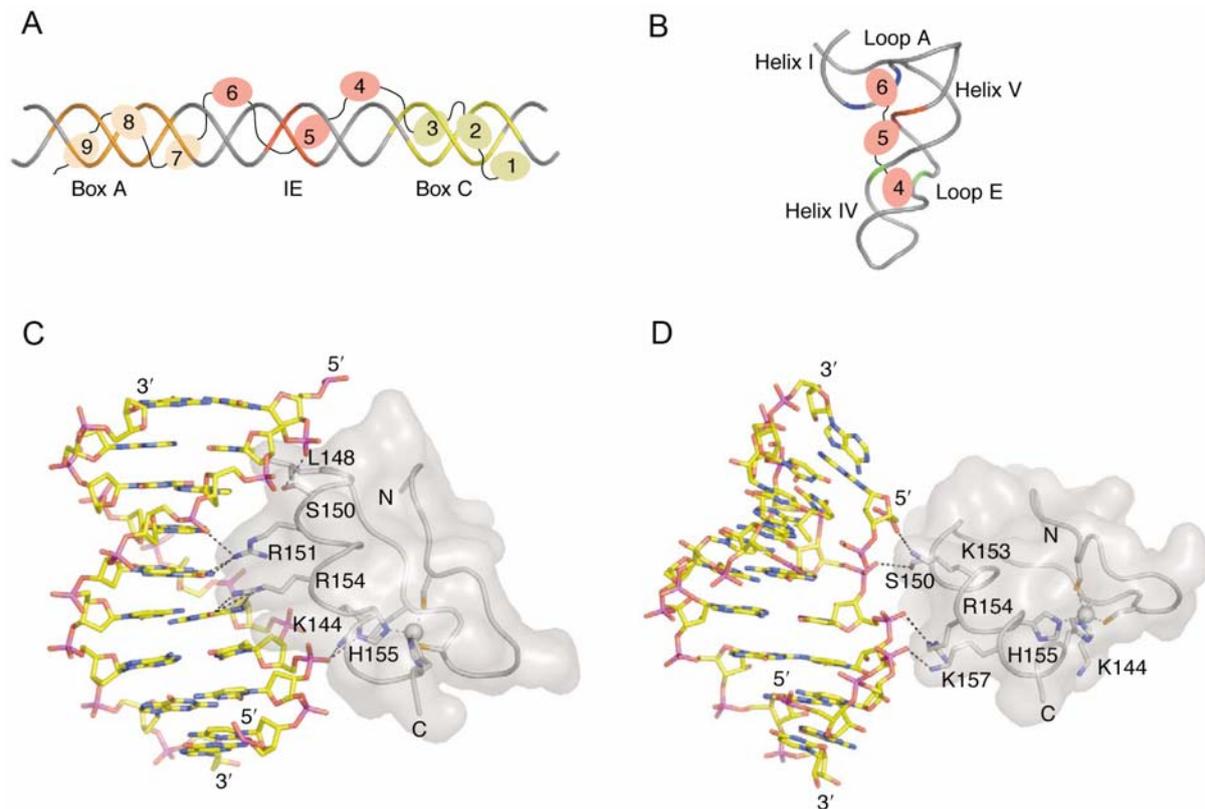


Figure 1-14: Comparison of DNA and RNA binding of the zinc finger protein TFIIIA. (A) All classical Cys₂His₂-type zinc fingers of TFIIIA, except zinc finger 4 and 6, which are needed for proper spacing, participate in direct interactions with specific DNA sequences of the 5S rRNA promoter (Box C, IE and Box A). (B) In addition to zinc finger 5, also zinc finger 4 and 6 recognize elements of the 5S rRNA (loop E, helix V and loop A). Details of contacts of zinc finger 5 (grey cartoon trace), either with DNA of the 5S rRNA gene (C) or with 5S rRNA (D). In the DNA-bound state, the α -helix (depicted as grey helix) inserts into the major groove of the DNA and establishes hydrogen bonds between amino acids side chains and bases of the DNA (C). In the RNA-bound state the side chains of the α -helix do not contact the bases of the RNA but form contacts to phosphate groups of the RNA backbone (D). Color code: Nitrogen in blue, oxygen in red, carbon in yellow, sulfur in orange and phosphorus atoms in magenta. Coordinated zinc ions as grey spheres (modified, [130]).

Zinc finger domains can also function as protein recognition motifs [127, 128]. The occurrence of classical zinc fingers as multiple copies or arrays allows the assignment of different functions for the single zinc fingers. It was shown for the Ikaros family of transcription factors that one cluster consisting of several zinc fingers overtakes the function of DNA binding, while the other cluster mediates interaction with other zinc finger domains either of the same or of a different family member, leading to high affinity DNA binding [135, 136]. Other examples are known, where DNA recognition and protein-interaction function are even harbored within the same zinc finger: The N-terminal classical zinc finger of GATA-1 uses distinct surfaces to bind simultaneously to DNA and to a zinc finger of its interaction partner FOG-1 ('Friend of GATA-1'), whose protein interaction surface, by contrast, overlaps with the classical DNA binding surface (Figure 1-15A and [137]). Other

zinc finger structures, like the LIM, RING and PHD domains function primarily as protein-protein interaction surfaces (Figure 1-13 and [127]). LIM domains are composed of two sequential zinc-binding modules that are packed against each other to form a single domain. Within this family not only cysteine and histidine, but also aspartate and glutamate residues are found to coordinate more than one zinc ions; with differing length of the loops between the zinc ions [138]. As seen for the LMO4 and the *ldb1* protein, four of the eight β -sheets of the two tandem LIM domains (LIM1 and LIM2) of LMO4 are used to pair with β -sheets of the LIM interaction domain of *ldb1*, leading either to primarily hydrophobic (first LIM domain) or to a mixture of electrostatic and hydrophobic interactions as seen in the second LIM domain (Figure 1-15B and [139]).

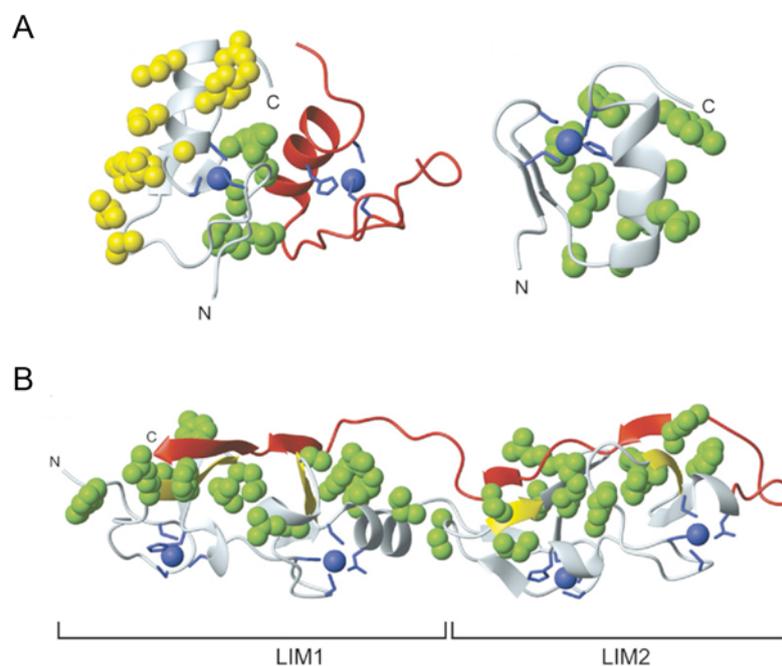


Figure 1-15: Protein-protein interfaces. (A) Interaction of the N-terminal classical GATA-1 zinc finger (in light blue) in with the FOG-1 zinc finger (in red). The DNA binding surface of the GATA-1 zinc finger (contacts depicted in yellow) does not overlap with its FOG-1 interaction surface (contacts depicted in green, picture on the left). By contrast, the FOG-1 zinc finger uses the DNA binding surface also as the protein interaction surface (contacts shown in green, picture on the right). Amino acid residues coordinating zinc ions are shown in dark blue. (B) Interaction of LIM domains (in light blue) of LMO4 with LIM interaction domains (in red) of *lab1*. Amino acid residues of LMO4 that participate in hydrophobic or electrostatic interactions are depicted in green. β -sheets of both domains that pair up are shown in yellow (modified, [127]).

In contrast to LIM domains, RING and PHD domains do not bind the two zinc ions sequentially, but interdigitated or ‘cross braced’ (Figure 1-13). RING domains are quite frequently found in E3 ubiquitin ligases, which catalyze the transfer of ubiquitin molecules from E2 transferases to the target protein. The RING finger can function as an interaction

domain with the E2 transferase, but can also possess intrinsic E3 ligase activity [127, 140], however, how this activity is mechanistically achieved by the RING structure is not solved yet. Recently, it could be shown that also the RING finger of human MSL2, which so far was only shown to function as a protein-protein interaction domain in *Drosophila*, is required for E3 ligase activity, at least in human cells [141]. In contrast to the related RING finger, PHD domains are often found in chromatin remodeling complexes and therefore a role in regulating chromatin structure is likely [142]. Indeed, PHD domains were shown to be able to bind to histones and to specifically recognize trimethylated lysine 4 of the N-terminal tail of histone H3 by forming a ‘cage’ of two or four aromatic side chains in which the trimethylated lysine side chain is trapped [143-145].

1.7 Objectives

In the field of eukaryotic gene regulation it is of major interest to understand the selective marking of an entire chromosome. A powerful model to study such a selection is the specific targeting of the X chromosome by the DCC in *Drosophila melanogaster*. Most of the knowledge about DCC targeting comes from *in vivo* studies, which frequently do not permit to dissect the underlying molecular mechanism. Especially the huge number of players and the complexity of their interactions in the *in vivo* situation cause problems. In contrast, the use of an *in vitro* system that makes use of purified components offer a way to dissect the molecular function of the single components involved. This study focuses on a single aspect of the multi-step targeting process: the interactions of MSL2 and MSL1 with nucleic acids.

In vivo studies suggest that the MSL2-MSL1 subcomplex serves as a binding module for the initial recognition of high affinity sites. However, a direct interaction of MSL2-MSL1 with candidate binding sites has not been studied. Moreover, neither the precise nature and structure of binding sites, nor a DNA binding function within the subcomplex is known. To address these fundamental questions, a purified system comprised of recombinant proteins and synthetic nucleic acids was employed. The interaction of several protein derivatives lacking important domains with different kind of nucleic acid structures was quantitatively measured by using electrophoretic mobility shift assays. The relevance of those protein-nucleic acid interactions for DCC targeting was in addition determined *in vivo* by using reporter gene assays and by localization studies of GFP-tagged proteins in cells.

2 Results

2.1 Purification of recombinant MSL proteins

In order to investigate the direct interaction of MSL proteins with candidate high affinity sites it was necessary to produce sufficient amounts of pure protein. As the heterologous expression system Sf21 cells (in combination with Baculovirus infection) were chosen and various MSL derivatives were purified using their C-terminal FLAG tags. In addition, the MSL2-MSL1 complex was coexpressed and copurified (with 1:1 ratio of both subunits) via the FLAG-tag of MSL2, whereas MSL1 was untagged. Besides expression in Sf21 cells, the CXC domain of MSL2 was also expressed in *E.coli*, but as a GST fusion protein. A schematic representation of all constructs and the Coomassie-stained protein preparations are shown in Figure 2-1.

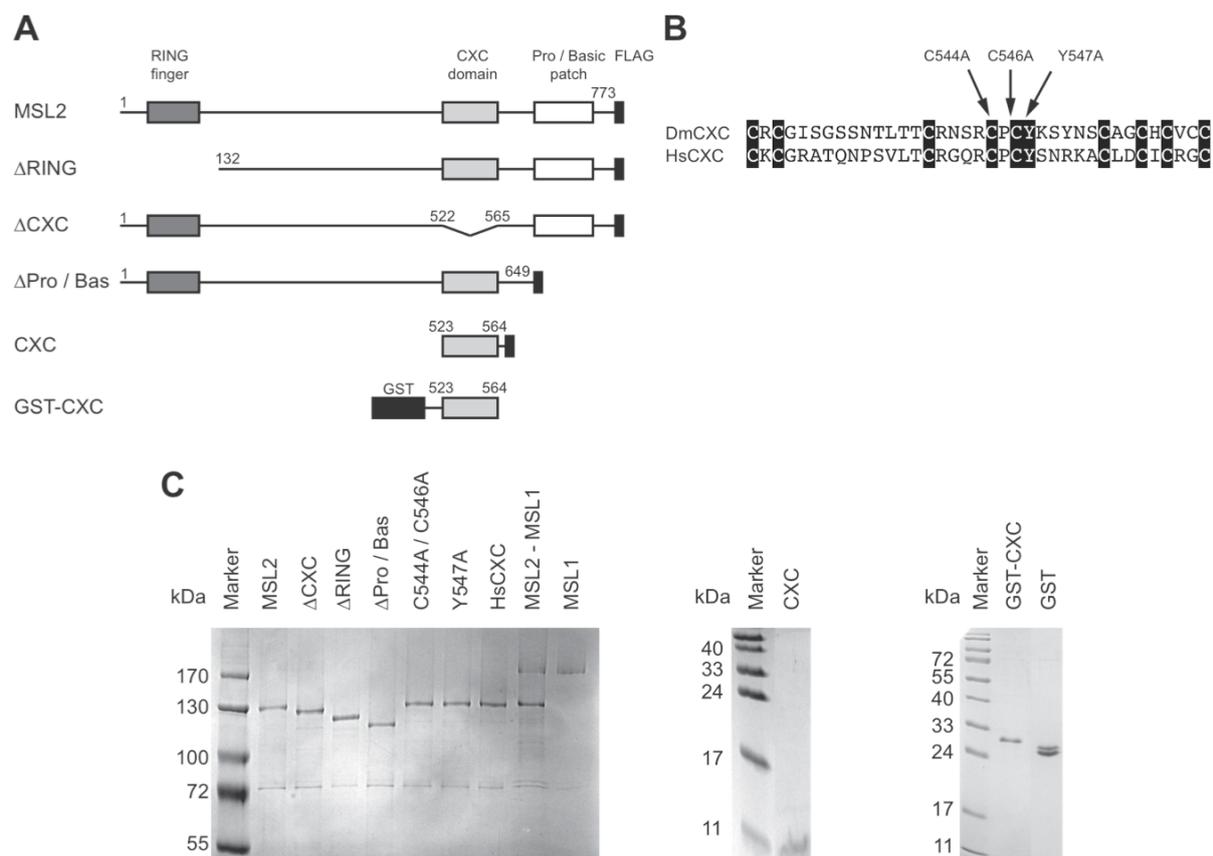


Figure 2-1: Overview of the various purified recombinant MSL proteins investigated in this study.

(A) Schematic representation of MSL2 domain organization. FLAG and GST tags are shown in black. Numbers correspond to the amino acid positions in full-length MSL2. (B) Alignment of orthologue CXC domains from the *Drosophila melanogaster* MSL2 protein (DmCXC) and from the *Homo sapiens* protein KIAA1585 (HsCXC). Black boxes show the conserved amino acids. Arrows indicate the introduced point mutations. (C) Coomassie-stained SDS-PAA gels of purified recombinant MSL proteins.

After having purified the MSL2-MSL1 complex from Sf21 cells the molecular mass and the stoichiometry of the complex was determined. For that purpose the purified complex (1:1 ratio of both subunits, Figure 2-1C) and the MSL2 protein were subjected to size exclusion chromatography (Figure 2-2). Both MSL2 and the MSL2-MSL1 complex do not elute at a defined molecular mass but rather elute over a wide range, indicating that both rather do not form a single, defined complex but exist as several complexes of different sizes. It seems that MSL2 and MSL1 are not forming a heterodimer (300 kDa) or heterotetramer (600 kDa) and that MSL2 by itself does not exist as a monomer (130 kDa). Interestingly, MSL2 and MSL1 seem to appear in equal amounts in every fraction, what suggests that a single MSL2 molecule interacts with a single MSL1 molecule. It is possible that the protein aggregates are only forming because MSL2 and MSL1 are taken out of the physiological context of the DCC, however it might well be that these aggregates are functionally important.

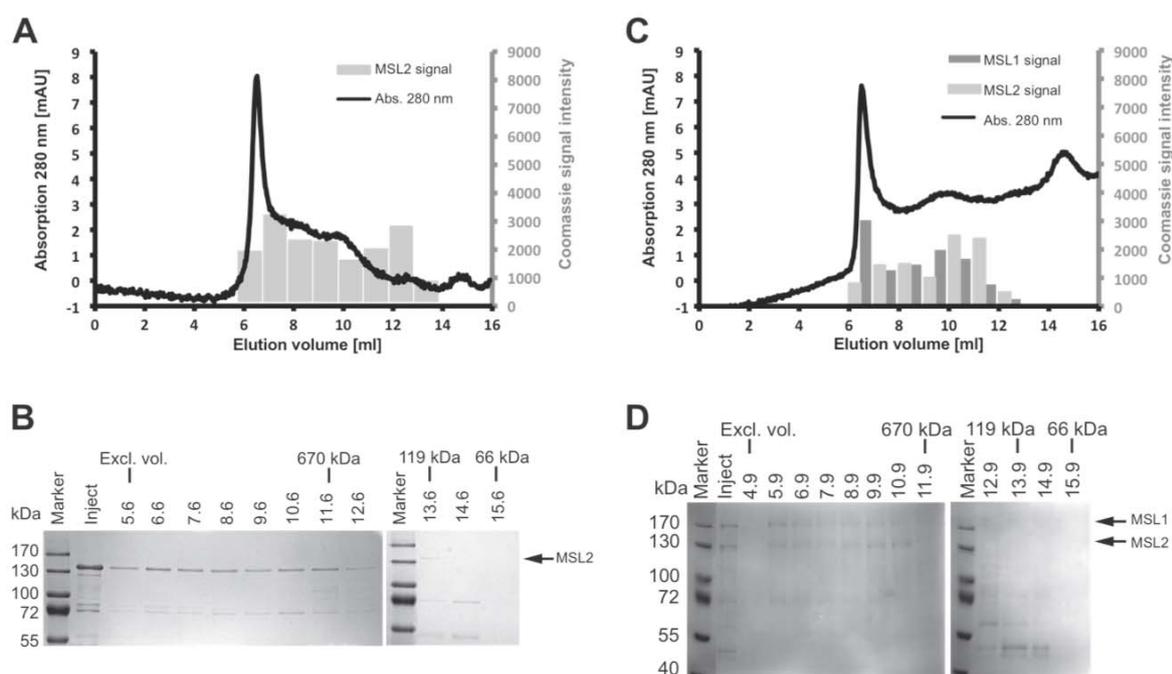


Figure 2-2: Gel filtration to analyze the stoichiometry of MSL2 and the MSL2-MSL1 complex. Chromatogram of size separation of recombinant MSL2 protein (A) and MSL2-MSL1 complex (C). The black line is the measured absorption at 280 nm during the gel filtration run (Superose 6 column). Gray bars show the signal intensity (arbitrary units) of MSL2 and MSL1 protein in the individual fractions. The signal intensity was determined by quantification of MSL2 (B) or MSL2 and MSL1 (D) after TCA precipitation of the individual fractions, separation by SDS-PAGE and subsequent staining with Coomassie blue. Numbers correspond to the elution volume in ml. The elution volume of molecular weight standards (Thyroglobulin 670 kDa, ISWI 119 kDa, BSA 66 kDa) is indicated. The exclusion volume of the Superose 6 column is marked with 'Excl. vol.'. 'Inject' represents 2% of the loaded material.

2.2 MSL2 is the DNA binding factor in the MSL2-MSL1 complex

The MSL2-MSL1 complex is considered to serve as a binding platform that initially recognizes HAS on the X chromosome [79, 80]. But so far such a direct interaction of MSL2-MSL1 with HAS has not been shown. Therefore the direct binding of MSL proteins to HAS was analyzed. As a model HAS, the recently by Gilfillan *et al.* isolated DCC binding fragment DBF12-L15 was used. This short 40 bp DNA element was found to be the minimal DNA sequence, which is sufficient to recruit the DCC in a reporter gene assay (Figure 1-9 and [92]). It contains a proposed consensus sequence (GAAAGAGA), which is thought to be important for marking a DCC binding site *in vivo* [93, 94]. The purified recombinant MSL2, MSL1 and the MSL2-MSL1 complex (Figure 2-1C) were used and tested for their ability to bind the DBF12-L15 in electrophoretic mobility shift assays (EMSA). Indeed, MSL2 directly binds and forms a stable complex with the 40 bp DNA element in EMSAs (Figure 2-3).

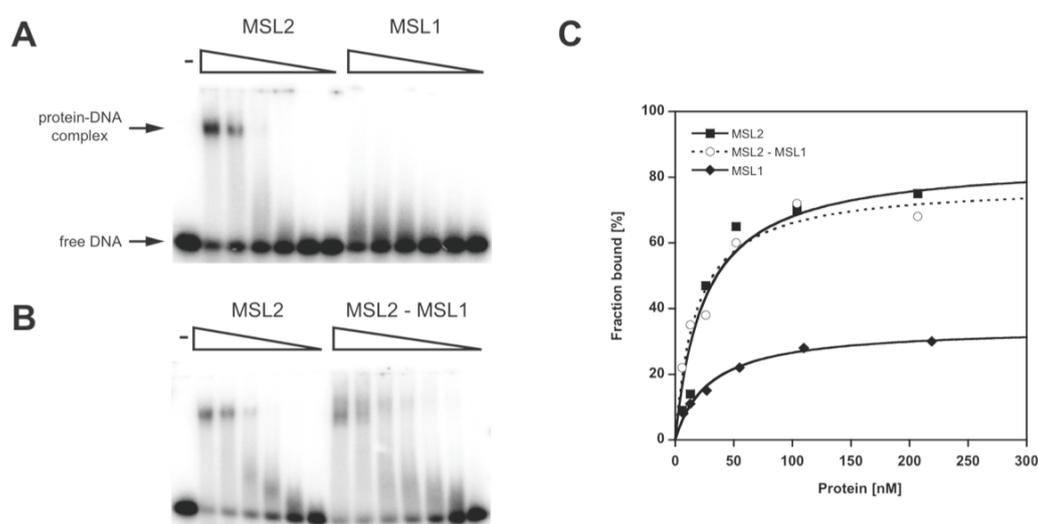


Figure 2-3: Binding of recombinant MSL proteins to a DNA HAS *in vitro*. Electrophoretic mobility shift assay. Increasing concentrations (from 5 nM to 250 nM) of MSL2 and MSL1 (A) or MSL2 and MSL2-MSL1 complex (B) were incubated with radiolabeled double stranded 40 bp DBF12-L15 DNA and protein-DNA complexes were separated from unbound DNA in non-denaturing EMSA gels. (C) Binding curves obtained from quantification of (A) and (B) and fit to a standard bimolecular model (see 'Material and Methods').

The affinity to DBF12-L15 can be determined from the binding curve and lies in the low nanomolar range (33 ± 18 nM). In contrast, MSL1 did not bind and did not form a stable protein-DNA complex with this short high affinity site (Figure 2-3A, C and Table 1). Since both MSL proteins are thought to act together *in vivo*, also the copurified MSL2-MSL1 complex in EMSAs was tested (Figure 2-3B). The affinity of the MSL2-MSL1 complex to the high affinity site was similar to the affinity of MSL2 (Figure 2-3C and Table 1), meaning that

MSL1 does not contribute to the DNA binding ability of the MSL2-MSL1 complex. This strongly suggests that it is only the MSL2 protein, which is required for the recognition of HAS.

2.3 The CXC domain of MSL2 is required but not sufficient for DNA binding

Since a typical DNA-binding motif within the MSL2 or any other MSL protein has not been identified yet, different deletion constructs of MSL2 were tested for DNA binding to identify the responsible DNA binding domain. Within the MSL2 protein several domains have been described: The RING finger is the interaction domain for MSL1 [80] and the C-terminal Pro/Bas patch was considered to be involved in RNA binding by an unknown, direct or indirect principle [99]. A function for the CXC domain, however, has not been assigned yet (Figure 1-10). Several MSL2 derivatives that lack one of the domains were constructed and purified (Figure 2-1A, C) and their affinity to the DBF12-L15 DNA element in EMSAs was measured (Figure 2-4 and Table 1).

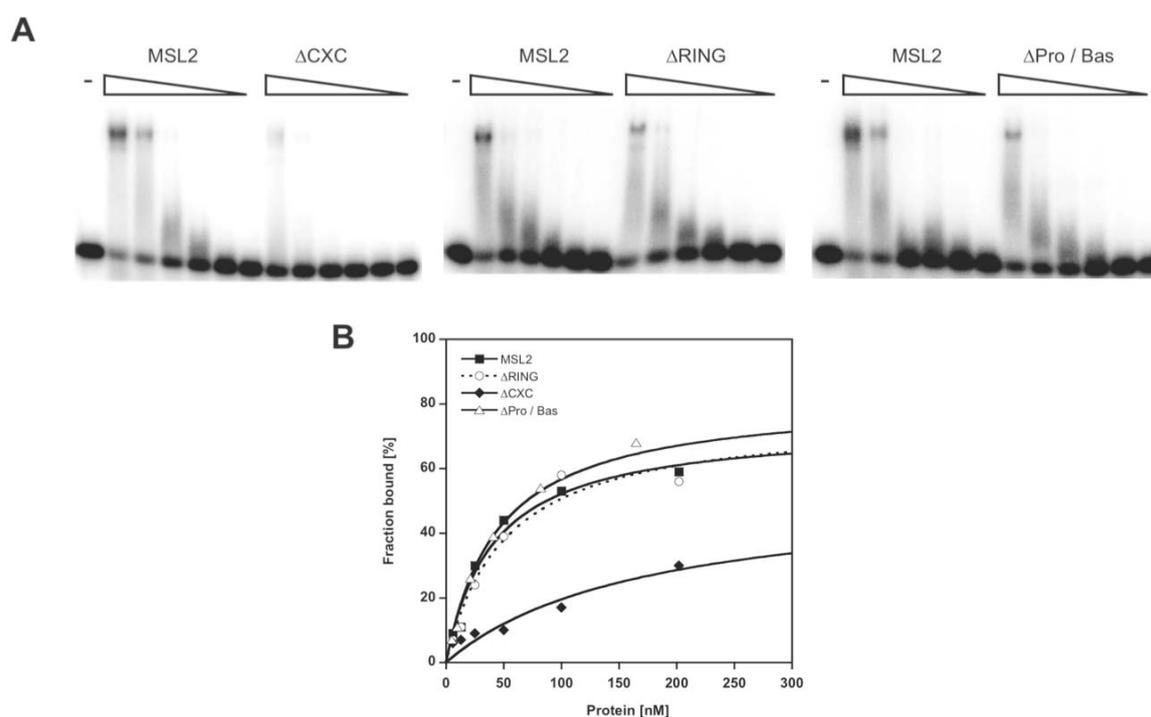


Figure 2-4: Binding of different recombinant MSL2 derivatives to a DNA HAS *in vitro*. (A) Electrophoretic mobility shift assays. Increasing concentrations of MSL2, MSL2- Δ CXC, MSL2- Δ RING and MSL2- Δ Pro/Bas were incubated with radiolabeled double stranded 40 bp DBF12-L15DNA and protein-DNA complexes were separated from unbound DNA in non-denaturing EMSA gels. (B) Binding curves obtained from quantification of EMSA gels and fit to a standard bimolecular model.

The deletion of the MSL2 RING finger did not change the affinity of MSL2 to the DNA element. The affinity also did not change drastically, when the C-terminal Pro/Bas patch was removed. On the other hand, only the deletion of the CXC domain strongly reduced the affinity to the DNA, suggesting that it is the CXC domain, which mediates binding to HAS.

The question arose whether the CXC domain by itself would still be able to bind DNA. Therefore the CXC domain was expressed alone, but fused to a GST or FLAG tag, and tested for its ability to bind DNA. Both the FLAG tagged and the GST tagged CXC domain could bind DNA by itself. However, the affinity was much lower (three orders of magnitude) compared to the full-length MSL2 protein (Figure 2-5). In other words the CXC domain is required but not sufficient for DNA binding. Additional regions within MSL2 are needed for full DNA binding.

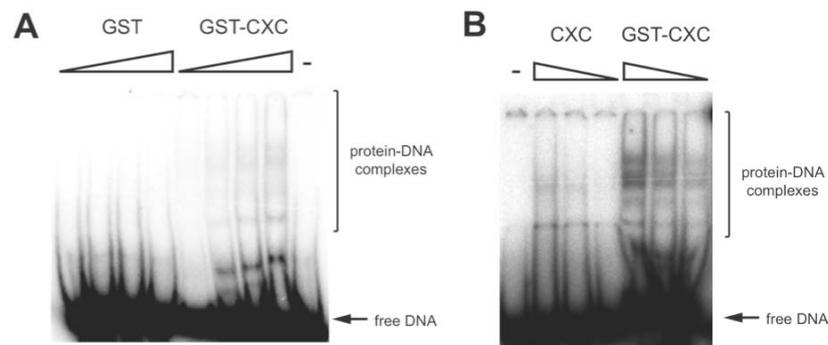


Figure 2-5: Binding of the isolated recombinant CXC domain to a DNA HAS *in vitro*. Electrophoretic mobility shift assay. Increasing concentrations (from 1 μ M to 8 μ M) of the GST tag alone and GST-CXC (A) or of CXC-FLAG (5 - 20 μ M) and GST-CXC (2 - 8 μ M) (B) were incubated with radiolabeled double stranded 40 bp DBF12-L15 DNA and protein-DNA complexes were separated from unbound DNA in non-denaturing EMSA gels.

2.4 Point mutations within the CXC domain affect DNA binding

In order to confirm the novel role of the CXC domain as a motif, which is required for DNA binding, we created several mutations within the CXC domain (Figure 2-1B) and analyzed their effect on DNA binding. On the one hand two point mutations (cysteines to alanines), which were already described to affect development and viability of male transgenic flies [80], were simultaneously introduced within the CXC domain. On the other hand, a tyrosine to alanine substitution adjacent to one of the cysteines was created. The mutated MSL2 proteins were purified (Figure 2-1C) and tested in EMSAs (Figure 2-6 and Table 1). In fact, all mutations within the CXC domain reduced the affinity to DNA (33 nM

vs. 82 nM), but the effect was milder than the complete deletion of the CXC domain (189 nM).

The CXC motif is conserved in the human MSL2 homologue [146]. To assess the DNA binding properties of the human CXC domain, a chimeric MSL2 protein was constructed (HsCXC, Figure 2-1B, C), where the CXC domain was replaced by a similar CXC motif of an orthologue *Homo sapiens* protein KIAA1585 [146]. Strikingly, the chimeric HsCXC protein exhibits the same affinity to DNA as the *Drosophila melanogaster* MSL2 protein (Figure 2-6 and Table 1), pointing out the importance of the conserved amino acids (Figure 2-1B) of the CXC motif for DNA binding.

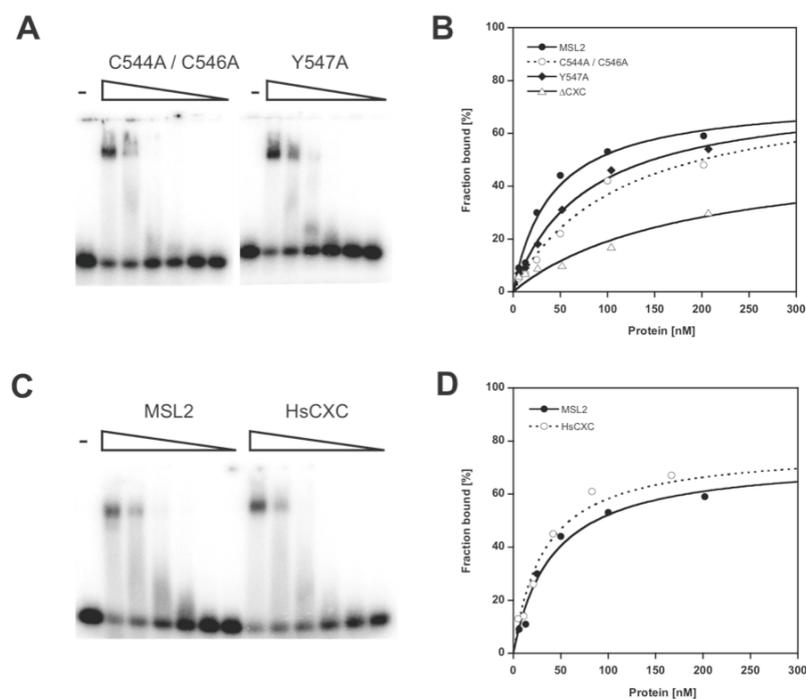


Figure 2-6: Binding of different MSL2 point mutants and of the chimeric MSL2 protein to a DNA HAS site *in vitro*. Electrophoretic mobility shift assay. Increasing concentrations of MSL2 carrying point mutations in the CXC domain (A) or of the chimeric HsCXC protein (C) were incubated with radiolabeled double stranded 40 bp DBF12-L15 DNA and protein-DNA complexes were separated from unbound DNA in non-denaturing EMSA gels. (B) and (D) Binding curves obtained from quantification of (A) and (C) and fit to a standard bimolecular model. For comparison MSL2 wildtype and MSL2 Δ CXC are displayed, too.

2.5 Double stranded DNA is the preferred binding target for MSL2

So far we demonstrated that MSL2 binds via its CXC domain to a double stranded (ds) DNA HAS. Since it is not known whether a HAS is in fact composed as a ds DNA element, we next asked if such a site would also be recognized when only one single DNA strand of the DBF12-L15 element was used or when the DBF12-L15 element was transcribed into RNA sequence, double and single stranded (ss). For this purpose a series of competition

assays was performed, where the relative affinities of MSL2 to other nucleic acids were measured (Figure 2-7). The competition curves showed that neither ssDNA nor ssRNA could compete with the MSL2-bound dsDNA. In contrast, dsRNA could compete the binding, nevertheless with a lower relative affinity compared to dsDNA (IC_{50} value 13.0 ± 6 nM vs. 36 ± 8 nM). In addition the competition curve for dsRNA has a slight sigmoid shape, indicating that there might be more than one binding site for RNA.

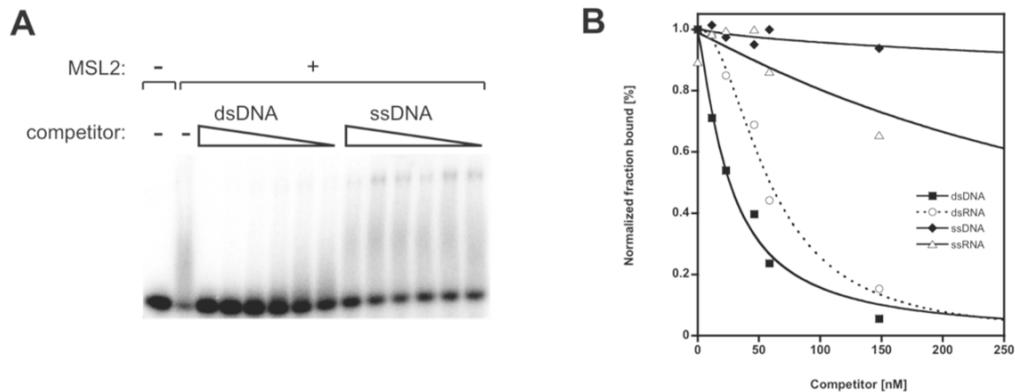


Figure 2-7: Competition assays to assess selective binding of MSL2 to different nucleic acids. (A) Electrophoretic mobility shift assay. 50 nM of MSL2 was incubated with the radiolabeled 40 bp DNA element DBF12-L15 and then increasing concentrations (from 5 nM to 2080 nM) of unlabeled competitor nucleic acids were added. Shown is a representative example of competition with ds and ssDNA. (B) Competition curves obtained from EMSAs and fit to the model described in ‘Material and Methods’.

The potential of MSL2 to directly bind RNA has not been appreciated so far and we were therefore interested to map the RNA binding domain within the MS2 protein. Different recombinant MSL2 truncations were used and their affinity to dsRNA was measured in EMSAs (Figure 2-8). Interestingly, the affinity of MSL2 to RNA was not affected by any of the truncations. Even the deletion of the CXC domain had no effect on RNA binding, which clearly demonstrates the function of the CXC domain as a specific dsDNA binding motif within the MSL2 protein. On the other hand, the region(s) responsible for RNA binding could not be located using this set of deletions. No further attempts were made since the focus of this project was on DNA binding domains.

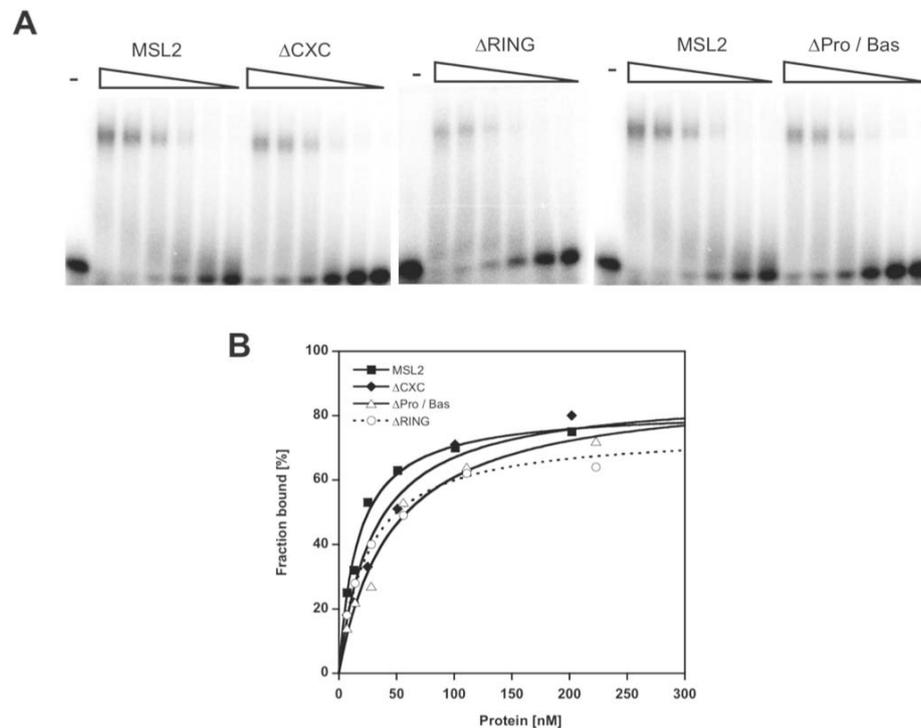


Figure 2-8: Binding of different recombinant MSL2 derivatives to RNA. (A) Electrophoretic mobility shift assays. Increasing concentrations of MSL2 versions were incubated with radiolabeled dsRNA of DBF12-L15 and protein-RNA complexes were separated from unbound RNA in non-denaturing EMSA gels. (B) Binding curves obtained from quantification of EMSAs and fit to a standard bimolecular model.

2.6 DNA binding of MSL2 is sequence-independent

According to a simple model, the recognition of the X chromosome by the DCC would involve binding of the DCC to a consensus sequence with higher affinity compared to non-target sites. Here it was shown for the first time that MSL2 has the potential to serve as a DNA recognition module. In the next step, it should be tested whether MSL2 was able to distinguish HAS from random DNA. As a model HAS the DBF12-L15 DNA element was used. Mutations of the GA-rich consensus sequence to a stretch of thymines led to the DBF12-L18 sequence that was not anymore able to recruit the DCC in a reporter gene assay in cells (Figure 1-9 and [92]). We used those two dsDNA sequences, DBF12-L15 and L18, together with recombinant MSL2 protein in EMSAs and measured their binding affinities. Surprisingly, the affinity of MSL2 to a HAS was similar to a non-high affinity site (K_D value 33 ± 18 nM vs. 23 ± 8 nM, Figure 2-9A). Conceivably for a sequence-specific binding, not only MSL2 is needed, but in addition the MSL1 protein. We therefore included the copurified MSL2-MSL1 complex in our binding studies. Like the MSL2 protein, the MSL2-MSL1 complex bound the high affinity site as well as the mutated site (data not shown), reinforcing

the fact that MSL2 is the single component required for the direct, however, sequence-independent DNA binding.

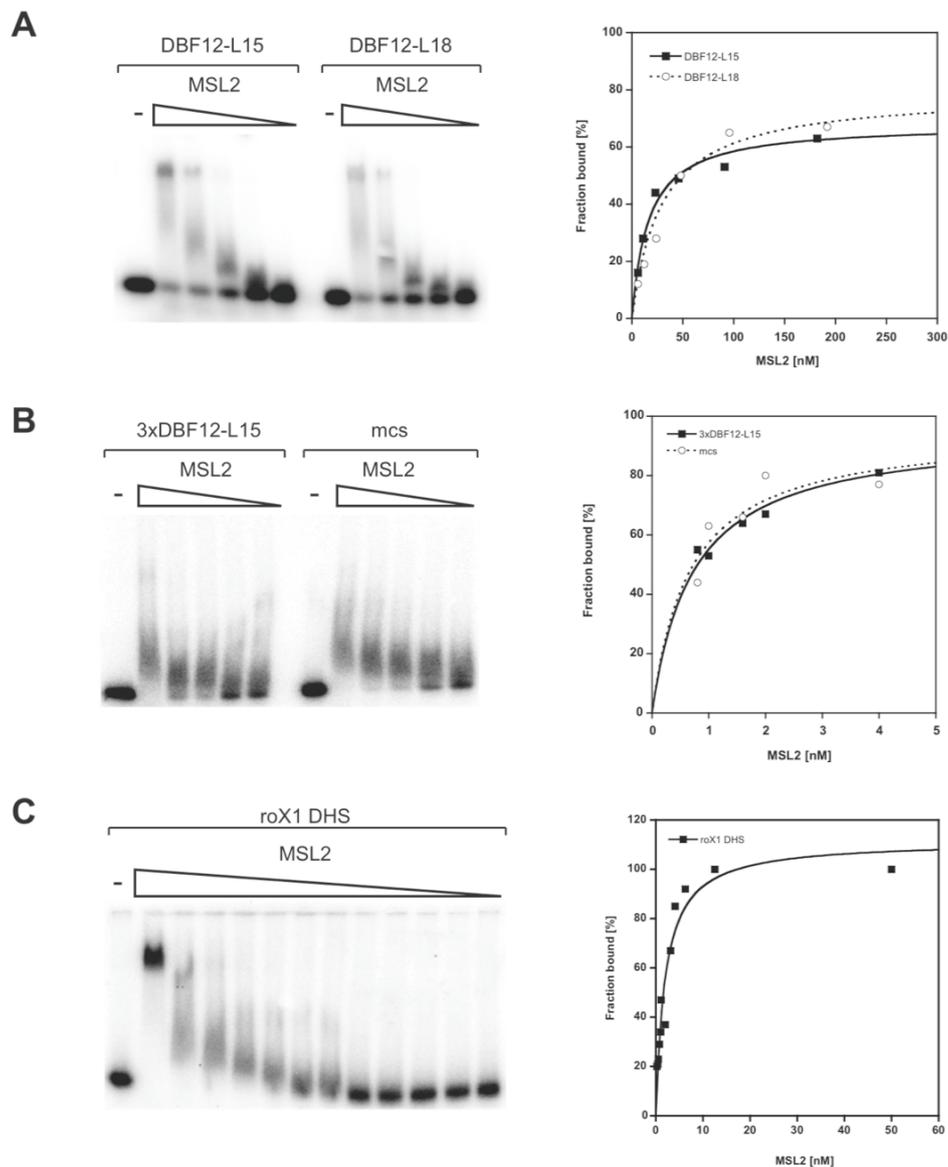


Figure 2-9: Binding of recombinant MSL2 to different DNA HAS *in vitro*. Electrophoretic mobility shift assays and binding curves obtained from quantification of EMSA gels and fit to a standard bimolecular model. Increasing concentrations of MSL2 were incubated with radiolabeled dsDNA of the model DBF12-L15 and the mutated high affinity site DBF12-L18 (A). (B) Same as (A) but with the DBF12-L15 trimer and a non-related multiple cloning site (mcs) or (C) with the roX1-DHS as DNA binding sites. Protein-DNA complexes were separated from unbound DNA in non-denaturing EMSA gels.

2.7 Affinity of MSL2 to DNA is increased with the length of DNA

Previously it was shown that the single 40 bp DBF12-L15 DNA element by itself was rather weak in recruiting the DCC to both, a reporter gene in cells and an autosomal insertion site in larvae. However, as a trimer it produced one of the strongest recruitment sites detected so far [92]. Therefore it was tested whether such a trimerized DNA element would increase the affinity or selectivity of MSL2. As measured by EMSAs the affinity of MSL2 increased dramatically, when the DNA element was trimerized (from K_D of 33 ± 18 nM to 0.59 ± 0.18 nM). Surprisingly, a control sequence (the multiple cloning site of a cloning vector) of similar length produced a similar increase of affinity (Figure 2-9B and Table 1). MSL2 also bound a 226 bp fragment derived from the *roX1* gene, which is known to contain a HAS [83] with similar affinity (Figure 2-9C). In other words, the affinity of MSL2 increases with the length of the DNA, however, the specificity for a certain DNA sequence is not altered.

Table 1: Comparison of binding affinities of different recombinant MSL2 derivatives to various nucleic acid binding sites. K_D values were calculated from binding curves obtained from quantification of EMSA gels and fitting to a standard bimolecular model. Mean values are shown plus and minus the standard deviation from several replicates obtained with at least two independent protein preparations. For details see the ‘Materials and Methods’ section.

MSL protein	K_D DBF12-L15 [nM]	K_D DBF12-L15 RNA [nM]	K_D DBF12-L18 [nM]	K_D DBF12-L15 x 3 [nM]	K_D mcs [nM]	K_D roX1 DHS [nM]
MSL2	33 ± 13	21 ± 4	23 ± 9	0.59 ± 0.18	0.61 ± 0.07	1.4 ± 0.87
Δ CXC	189 ± 18	29 ± 12				
Δ RING	51 ± 5	23 ± 5				
Δ Pro / Bas	61 ± 5	37 ± 10				
C544A / C546A	82 ± 4					
Y547A	81 ± 5					
HsCXC	33 ± 9					
MSL2 / MSL1	26 ± 15					
MSL1	no binding					

2.8 Reporter gene activation in cells requires the CXC domain

It was shown that the CXC domain is indispensable for DNA binding of MSL2 *in vitro*. To confirm the relevance *in vivo*, the previously described reporter gene assay in *Drosophila melanogaster* SL2 cells was used ([92]). In this assay DCC binding sites are placed in front of a minimal promoter that drives transcription of a *luciferase* reporter gene. The plasmid is transfected into male *Drosophila melanogaster* SL2 cells together with a plasmid that allows expression of an MSL2-VP16 fusion protein. Upon binding of the MSL2 fusion protein to the binding element, the VP16 activation domain leads to the expression of the *luciferase* reporter gene (Figure 1-9 and Figure 2-10B). In this study, the different MSL2-VP16 constructs were tested for their ability to activate the reporter gene. Importantly, all MSL2 derivatives were expressed at a similar level (Figure 2-10A). Upon recruitment of MSL2-VP16 to the DBF12-L15 binding site the *luciferase* reporter gene was activated roughly six-fold (Figure 2-10C in agreement with earlier results [92]). The MSL2 deletion Δ RING was still able to activate the reporter gene, suggesting that this domain is not important for gene activation. Since the RING finger interacts with MSL1, the MSL1 protein seems to be not required for DNA binding *in vitro*, nor for transactivation *in vivo*. By contrast, MSL2 lacking the CXC domain or carrying point mutations in the same domain were not able to activate. Evidently, affecting the DNA binding function of the CXC domain also affects gene activation *in vivo*. Even lowering the affinity to DNA by only 3-fold (Table 1) completely prevents the activation of the reporter gene. The deletion of the Pro/Bas patch reduced reporter gene activity, however, still a robust two-fold activation was measured. Remarkably, the chimeric HsCXC protein, which binds DNA with wildtype affinity, did not mediate transactivation. Together these data suggest that the DNA binding function of MSL2 via its CXC domain is a prerequisite for gene activation, but that additional *Drosophila* factors may contribute.

When the consensus DNA sequence in the DBF12-L15 DNA element was replaced by the mutated sequence in DBF12-L18 reporter gene activation with MSL2 as activator [92] as well as with all other MSL2 derivatives was abolished (Figure 2-10C), showing the importance of the GA-rich consensus sequence for MSL2 binding and / or gene activation in cells. *In vitro*, however, MSL2 bound both sequences with the same affinity. Again, all these findings suggest that DNA binding via the CXC domain is absolutely required, but other factors are needed to achieve specificity for a certain DNA sequence.

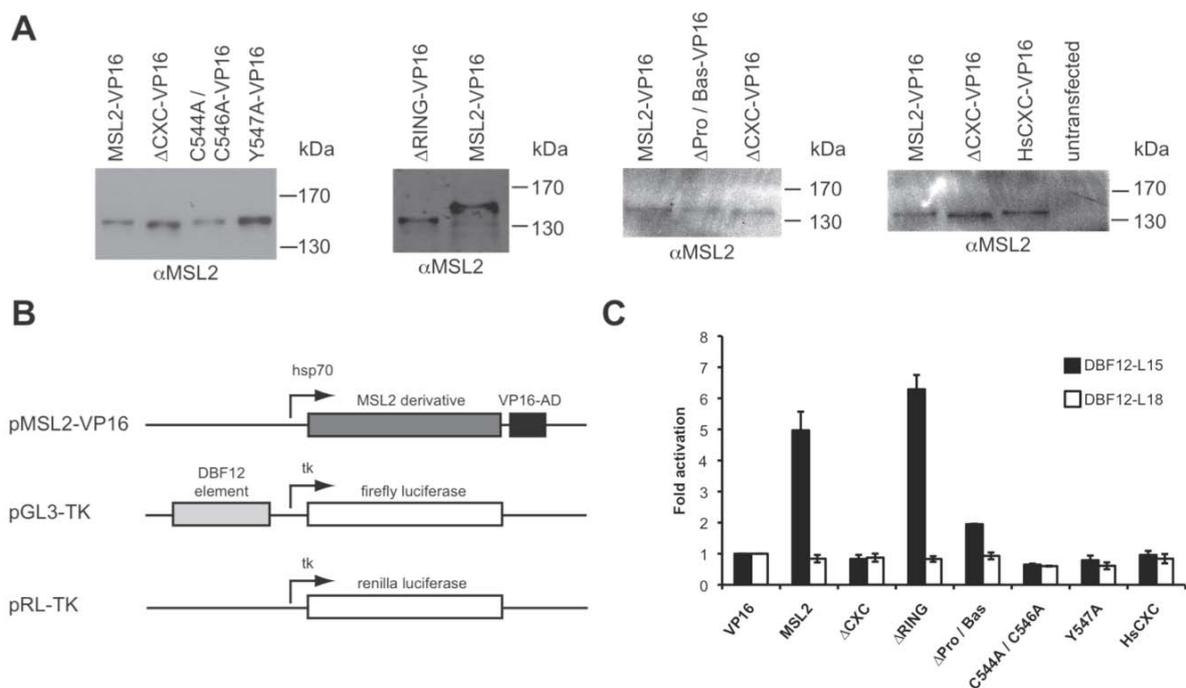


Figure 2-10: Reporter gene assay to measure transactivation potential of different MSL2 constructs *in vivo*. *Drosophila* SL2 cells were transiently cotransfected with different MSL2-VP16 constructs together with *luciferase* reporter gene constructs (pGL3-TK) carrying the model DBF12-L15 or the mutated site DBF12-L18 (40 bp elements). A *Renilla* luciferase expression vector (pRL-TK) served as normalisation control. (B) Schematic representation of the cotransfected plasmids that constitute the reporter gene system (drawing is not to scale). (C) Transactivation potential of each MSL2 derivative is displayed as fold activation (for details see Materials and Methods). The black bars indicate a reporter gene bearing MSL2 binding sites (DBF12-L15), the white bars represent (lack of) activation in the presence of DBF12-L18.

2.9 The CXC domain is necessary for targeting the DCC to the X chromosome

Targeting of the DCC to the X chromosome is known to depend on the MSL2-MSL1 heteromer. This study suggests that the direct interaction with DNA is brought about not by MSL1, but by the MSL2 protein, more precisely via its CXC domain. The reporter gene assay already suggested that this domain is also required for recruitment *in vivo*, at least to a single reporter gene. Whether also the targeting to the entire X chromosome involves the DNA binding function of MSL2 was addressed next. MSL2- Δ CXC and MSL2-C544A/C547A derivatives, which show reduced affinity to DNA, as well as wildtype MSL2 were fused to GFP and stably expressed in *Drosophila melanogaster* SL2 cell lines. Western blot analysis showed that the stable cell lines express roughly the same amount of endogenous MSL2 and GFP-MSL2 (Figure 2-11A). Moreover, the level of expression was similar among all the transgenes. After several weeks of selection, the stable cell lines were analyzed for the localization of their MSL2-GFP protein (Figure 2-11B, C). The wildtype MSL2-GFP

colocalized perfectly with endogenous MSL1 to form a defined X-chromosomal territory in the nucleus. By contrast, when the DNA binding function of MSL2 was impaired also the X chromosomal targeting was disturbed: In around 75% of cells expressing the transgenes, the MSL2- Δ CXC or MSL2-C544A/C547A-GFP fusion proteins did not form well restricted X chromosome territories, but were localized to many, dispersed foci in the nucleus (Figure 2-11B). Interestingly, endogenous MSL1 perfectly colocalized with the mutated MSL2-GFP fusion proteins, showing that the interaction between MSL1 and the mutated MSL2 proteins is not disturbed (Figure 2-11C). More importantly, it clearly demonstrated that MSL1 cannot be maintained at its X chromosomal location without MSL2 and in turn followed MSL2 to ectopic sites. In summary, we conclude that for the proper targeting of the DCC *in vivo*, the functional DNA binding domain of MSL2 – the CXC domain – is required.

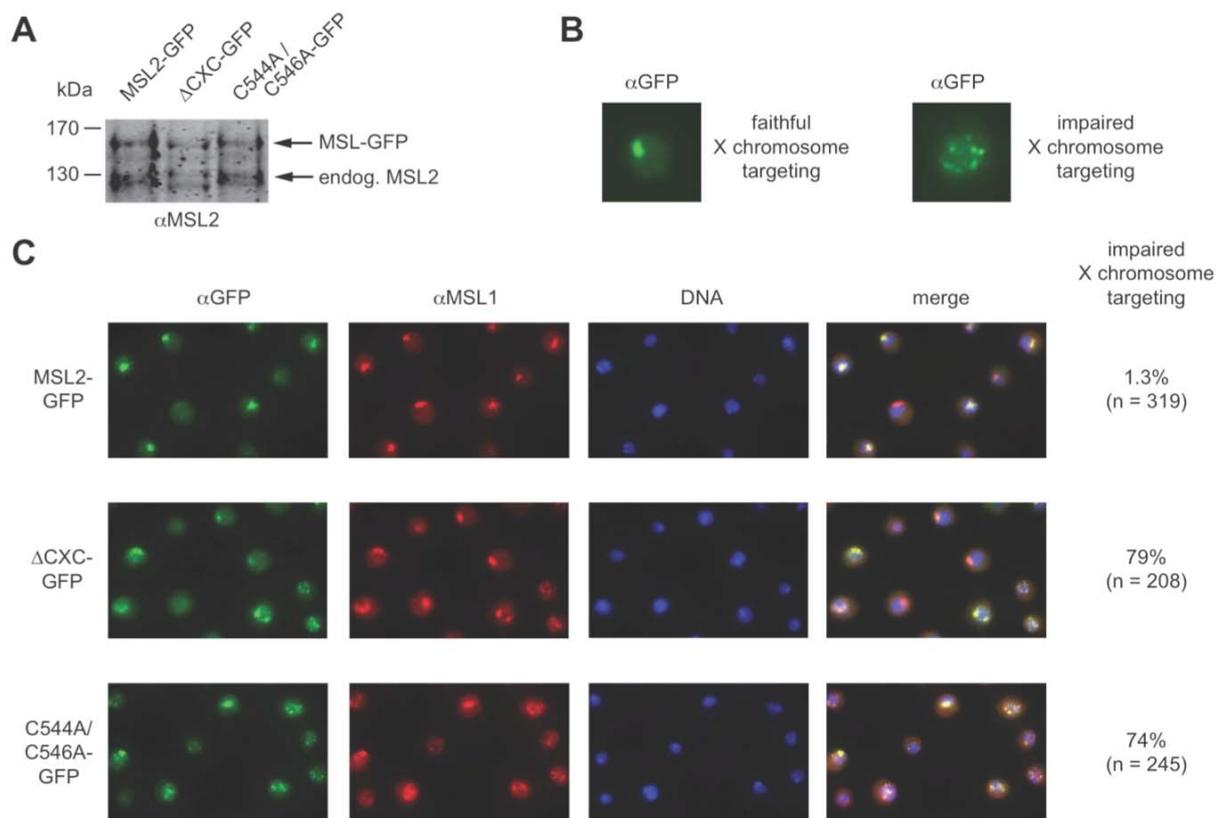


Figure 2-11: *In vivo* localization of different MSL2 constructs. Stable *Drosophila* SL2 cell lines, which express different MSL2 versions fused to GFP, were analyzed by Western blot analysis (A) and by immunofluorescence staining (B) and (C). (A) Cells were harvested, lysed and roughly 0.05×10^6 cells were subjected to Western blot analysis. A rabbit anti-MSL2 antibody was used to detect the endogenous MSL2 and the different MSL2-GFP fusion proteins. (B) Examples showing single nuclei with proper targeting of MSL2 to the X chromosomal territory (left) or dispersed, non-physiological distribution of the GFP fusion proteins. (C) Representative fields of cells. Localization of MSL2-GFP constructs was visualized by using anti-GFP antibodies. Endogenous MSL1 was detected with anti-MSL1 antibodies. DNA was counterstained with DAPI. Cells that express the MSL2-GFP transgene were analyzed for proper X territory staining. The percentage of cells, which show mislocalized, dispersed GFP signals and the number of total cells counted are displayed.

2.10 Binding of MSL2 to RNA-DNA triplex structures

As shown above, the binding specificity of the DCC *in vivo* is not brought about by different affinities of MSL2 to different DNA sequences *in vitro*. It might well be that a consensus sequence is indeed important, but that the representation as DNA sequence may be inappropriate. Examples are known, where DNA sequence has to be transcribed in order to serve as targeting elements [91, 147]. Consequently, a RNA-DNA triple helix could be an alternative target. Such triplexes are formed between homopyrimidine RNA and homopurine / homopyrimidine DNA sequences [148]. In fact, the proposed GA-rich consensus DNA sequence, as it is represented for example in the DBF12-L15 DNA element, contains a perfect homopurine / homopyrimidine recognition sequence and would be suited for forming a triple helix with a single stranded homopyrimidine RNA sequence, which might be represented in one of the roX RNAs. We therefore asked whether such a triplex would be a better-suited binding site compared to a DNA duplex.

In a first step different RNA-DNA triplexes were formed and subsequently tested as binding sites for MSL2 in EMSA. We initially used well-described RNA and DNA sequences (triplex forming oligonucleotides, TFO, Figure 2-12B and [149]) to test whether triplex formation works under our conditions. In fact, by increasing the concentration of the single stranded RNA it was possible to convert a DNA duplex into a slower migrating RNA-DNA triple helix. A published 27 bp (21TFO-S) as well as our designed longer 40 bp DNA duplex (40TFO) was suited for triplex formation (Figure 2-12B, C and [149]). Unfortunately, the 40 bp DBF12-L15 DNA did not form a triple helix with the corresponding 16 nt RNA (Figure 2-12D).

Nevertheless, EMSAs were performed, but not with DBF12-L15 but with the 40TFO DNA duplex and its corresponding triplex (Figure 2-12E, F). When the binding reaction was performed under conditions that ensure the formation of a stable triple helix (Figure 2-12C), MSL2 could not bind to the triple helix and also not to the 40TFO DNA duplex (data not shown). Consequently, triple helix formation was first carried out under the known triplex forming conditions, and then the binding reactions were carried out under the known conditions that allow MSL2 to bind DNA. Indeed, MSL2 formed a protein-DNA complex with the 40TFO DNA and also with its corresponding triplex, strikingly with roughly similar binding affinity (Figure 2-12E, binding curve not shown). However, the resolution of the native agarose gel does not allow a distinction between a duplex or triplex structure, making it impossible to clarify whether indeed a triplex structure was bound by MSL2, or whether the

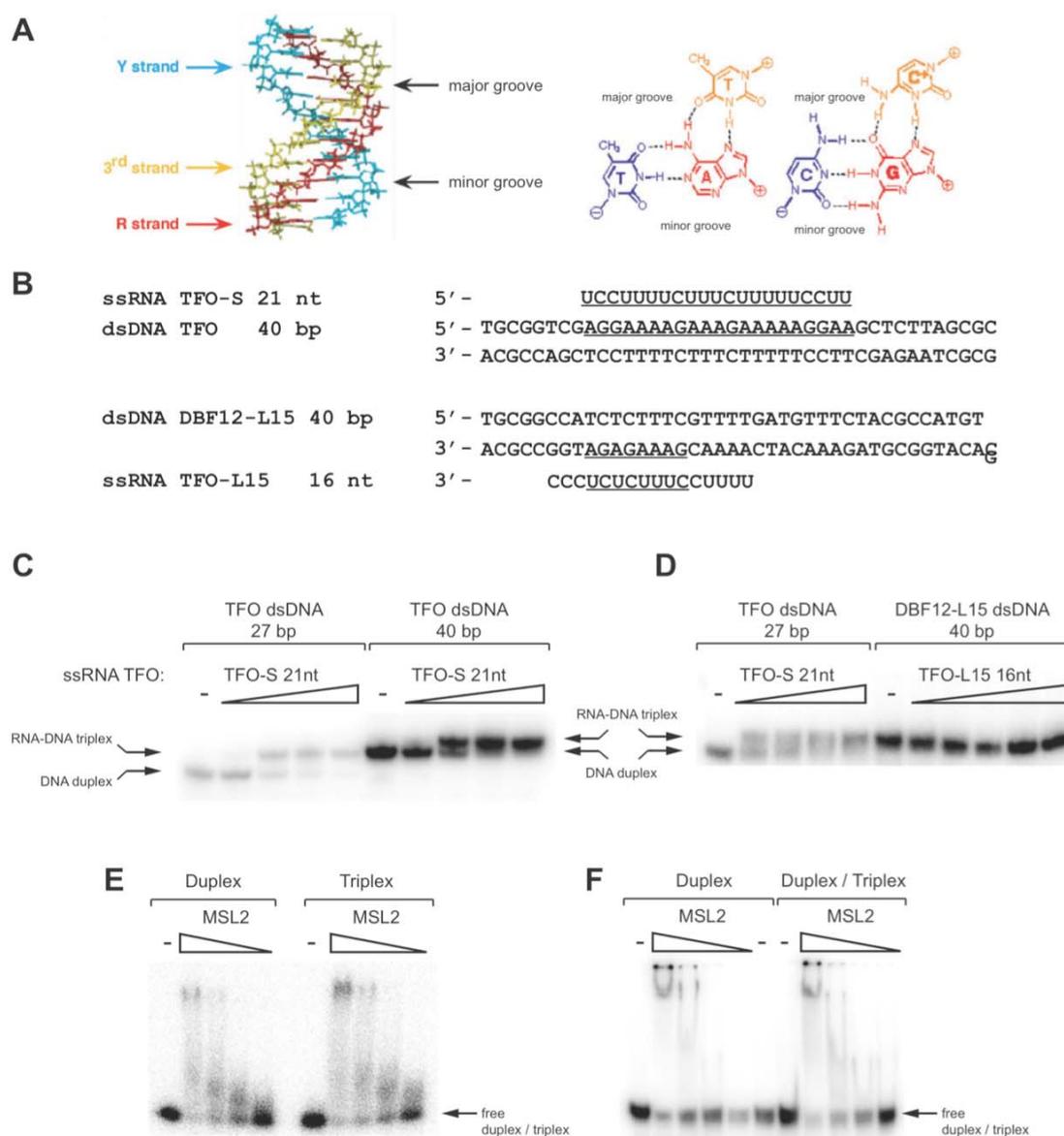


Figure 2-12: Binding of recombinant MSL2 to duplex and triplex structures *in vitro*. (A) DNA triplex structure formed by an oligopyrimidine strand (yellow) and an oligopurine (red) / oligopyrimidine (blue) duplex. In the picture on the right, details of the base interactions are shown for a single DNA triplet motif (modified, [148]). (B) ssRNA and dsDNA sequences that were used for triplex formation. Triplex recognition motifs are underlined. (C)-(F) Electrophoretic mobility shift assays. Increasing concentrations (1, 2, 4 and 8 μ M) of different TFO RNAs were incubated with radiolabeled double stranded TFO (C) or DBF12-L15 (D) DNA duplexes. Duplex and triplex structures were then separated in non-denaturing agarose gels. Increasing concentrations (from 5 nM to 250 nM) of MSL2 were incubated with radiolabeled duplex and triplex structures (E) or with duplex and duplex:triple ratio 1:1 (F) formed in (C). Protein-bound structures were separated from unbound duplex or triplex in native 1.2% agarose (E) or 12% PAA (F) gels.

triplex has simply fallen apart. This could either happen because of unfavorable binding or EMSA conditions or by the action of MSL2. MSL2 could simply bind the DNA duplex within the triplex structure and lead to the dissociation of the triplex. To gain a better resolution for small molecules, native PAA gels were used and EMSAs with MSL2 and the

40TFO DNA duplex or a 1:1 mixture of triplex and duplex structure were performed (Figure 2-12F). Again, a stable MSL2-DNA duplex complex was detected in native PAA gels, but it was again not possible to resolve duplex and triplex structures. Consequently, it remains unsolved whether a triplex is stable under the applied conditions or whether MSL2 simply promotes the dissociation of the RNA strand from the triplex by recognizing and binding to the duplex DNA within the triplex structure. Currently experiments are performed to test the formation of a triplex not under the known triplex forming conditions (as used in Figure 2-12C, D) but under conditions, which are used for the subsequent MSL2 binding reaction.

2.11 MSL2 is not sufficient for reporter gene activation in human cells

Since the *in vitro* binding system works with purified and consequently known factors, it might well be that additional, yet unknown factors are required for specific targeting of the DCC. Therefore the requirements of additional factors, other than MSL2 and MSL1, for activation of a reporter gene were assessed, using a modification of the reporter gene system. The basic idea was to test whether MSL2 is able act autonomously in activation of a reporter gene in a human cell line or whether MSL2 needs additional factors that are exclusively present in *Drosophila melanogaster* and are not present in humans.

The MSL-VP16 activation constructs and an inducible *luciferase* reporter construct, which contains one of the strongest DCC binding site – the roX1-DHS [83] – were transiently cotransfected into human HEK-293 cells and reporter gene activation, as an indirect measure for recruitment of MSL2 to the roX1-DHS was measured. The same experiment performed in *Drosophila* SL2 cells gives a huge activation of the reporter gene by MSL2 [92]. In HEK-293, MSL2 expression was confirmed by Western blot (Figure 2-13A), however, no MSL2-dependent activation of the reporter gene was observed (Figure 2-13B). The analysis was extended and dose-response experiments were performed (Figure 2-13C). By increasing the amount of transfected MSL2- or MSL1-VP16 activator constructs no activation, but rather a repression of reporter gene activity was measured. Yet, this effect was very mild and in the case of MSL1 the repression was also observed on a control *luciferase* reporter gene construct, which contained no binding site. Since MSL2 and MSL1 are thought to act together *in vivo*, we cotransfected MSL2 and MSL1, which again slightly reduced the reporter gene activity (fold activation from 1.3 to 0.8). In summary our results suggest that MSL2 is not

able to bind autonomously to HAS. Rather, it seems that one or more additional factors (other than MSL1) are required and that those factors are only present in *Drosophila melanogaster*.

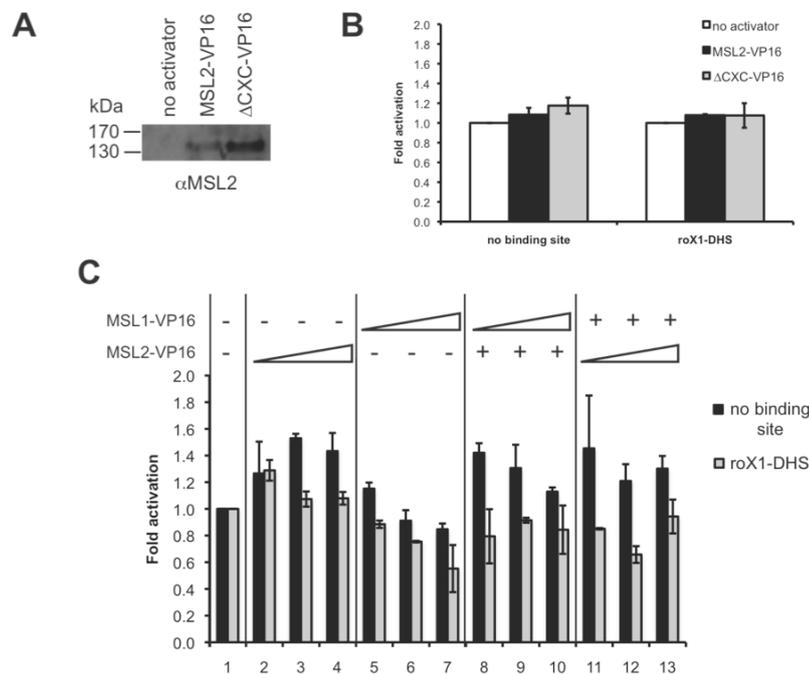


Figure 2-13: Reporter gene assay to measure the transactivation potential of MSL2 and MSL1 in a heterologous cell system. Human HEK-293 cells were transiently cotransfected with different MSL-VP16 constructs together with reporter gene constructs carrying the roX1-DHS binding site or no binding site. (A) Western Blot of roughly 0.1×10^6 cells, which were harvested, lysed and proteins precipitated using TCA. A rabbit anti-MSL2 antibody was used to detect the different MSL2 constructs. (B) Transactivation potential of MSL2 and MSL2-ΔCXC construct is displayed as fold activation. 112 ng of each MSL2 construct was transfected. (C) Increasing amounts (100, 300 and 600 ng) of MSL2 or MSL1 was transfected. '+' indicates an unchanged amount (100 ng), while '-' means 0 ng of the individual transfected MSL constructs.

2.12 Enrichment of a native DCC-containing fraction

Since there is evidence for additional factors that might be involved in targeting the DCC (see chapter above) and which might have been missed so far in the *in vitro* binding assays, a native fraction containing the DCC was enriched by affinity chromatography and it was looked for associated factors. For that reason, *Drosophila melanogaster* SL2 cell lines, which stably express either FLAG-tagged MOF or FLAG-tagged MOF containing an additional Gal4 DNA binding domain (Gal4DBD, construct created by M. Prestel, P. Becker lab) from the endogenous MOF promoter were created. MOF was immunoprecipitated from nuclear extracts of those cell lines using FLAG-affinity chromatography and finally eluted after washing with different salt concentrations. The enriched native MOF-containing fraction was analyzed for its composition and associated factors by Western blot and mass spectrometry.

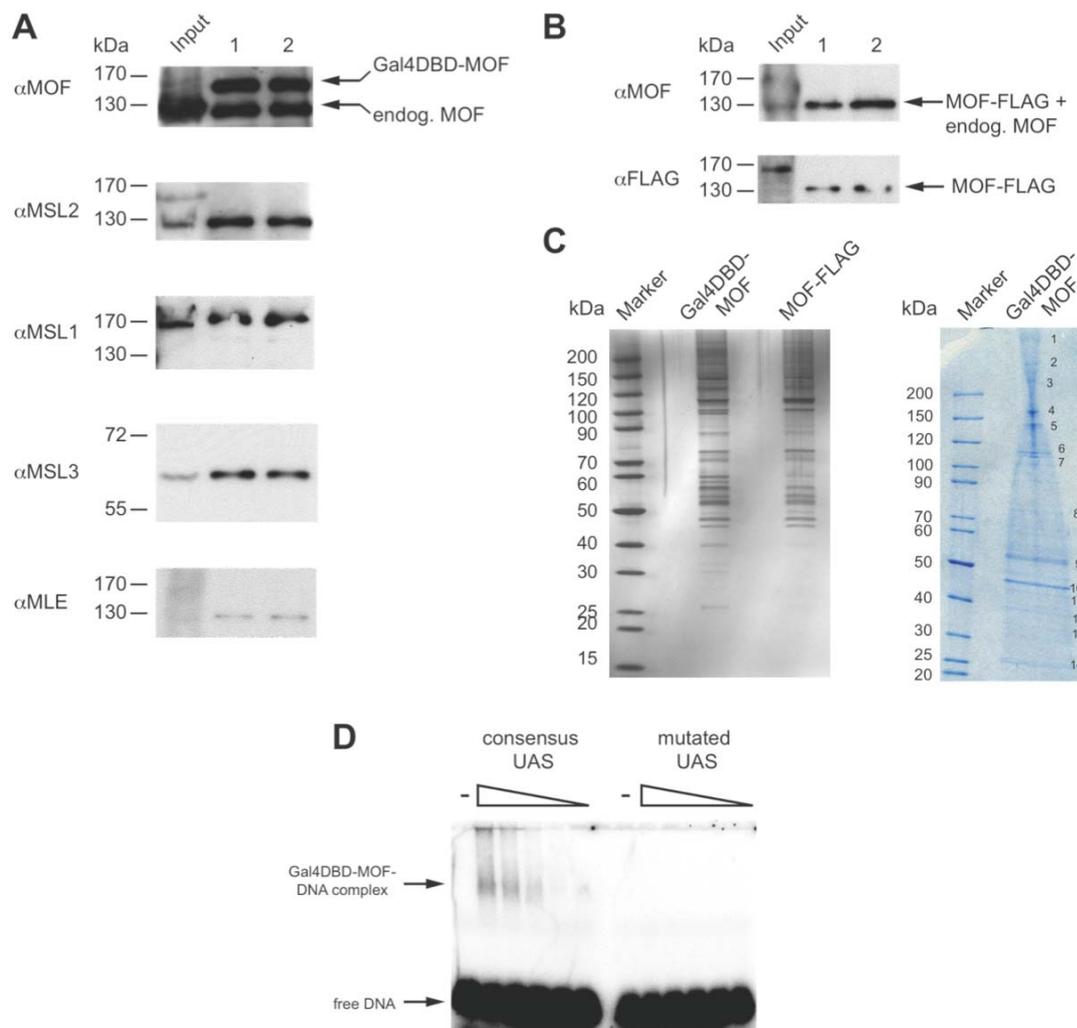


Figure 2-14: Enrichment of native DCC-containing fractions. Western blots with different antibodies on enriched Gal4DBD-MOF-FLAG (A) or MOF-FLAG (B) containing fractions after washing with 200 mM (1) and 400 mM (2) salt. (C) Silver stained SDS-PAGE gels of material used in (A) and (B) after 400 mM washing and Coomassie stained SDS-PAGE gels of (A). Cut gel slices that are subjected to mass spectrometry are numbered. (D) Binding of the enriched native Gal4DBD-MOF fraction (analyzed in (A) and (C)) to UAS sites *in vitro*. Electrophoretic mobility shift assays. Increasing concentrations of the Gal4DBD-MOF fraction were incubated with radiolabeled dsDNA of the consensus and the mutated UAS. Protein-DNA complexes were separated from unbound DNA in non-denaturing 0.7% agarose gels.

Western blot analysis confirmed that indeed all known subunits of the DCC were associated with tagged MOF (Figure 2-14A), including the MLE helicase, showing indirectly that roX RNAs are also associated with the enriched fraction. Interestingly, the MOF transgene and endogenous MOF were purified in the same stoichiometry, indicating that an even number of MOF molecules exist within the complex (dimer, tetramer, etc.). The enriched fractions were subjected to mass spectrometry analysis (performed by the Zentrallabor für Proteinanalytik, Adolf-Butenandt-Institute, LMU) and identified proteins are listed below (Table 2 and Figure 2-14C). Besides the already known interactors, like MSL1,

MBD-R2 and components of the exosome complex, no other striking proteins were identified. A functional relevance of the exosome complex for dosage compensation is rather unlikely, despite the already described copurification with the DCC by other labs [63], since control purifications showed that the association of exosome components is mediated non-specifically by the FLAG-agarose material (unpublished results, M. Prestel, P. Becker lab).

Table 2: Identified proteins in the enriched native Gal4DBD-MOF fraction using mass spectrometry. Numbers correspond to the gel slices shown in Figure 2-14. Identified proteins are listed with a short description of their functions according to the shown references. ‘-’ indicates that no protein was identified in the gel slice.

No.	Identified protein	Description	Reference
1	karst (beta Heavy-spectrin)	cytoskeleton of epithelial cells; rRNA processing	[150]
2	alpha spectrin	cytoskeleton of epithelial cells	[150]
3	-	-	-
4	MSL1 + MBD-R2	dosage compensation	[63]
5	MOF	dosage compensation	[63]
6	Dis3	3'-5'-exoribonuclease activity; RNA binding	[151]
7	Rrp6	3'-5'-exoribonuclease activity; RNA binding	[151]
8	Hsp70	chaperone	[152]
9	beta tubulin + eukaryotic initiation factor 4B	cytoskeleton microtubules translation initiation	[153] [154]
10	eukaryotic initiation factor 4B	translation initiation	[154]
11	actin 5C + eukaryotic initiation factor 4B	actin cytoskeleton translation initiation	[155] [154]
12	Mtr3	3'-5'-exoribonuclease activity; RNA binding	[151]
13	-	-	-
14	CG31938	3'-5'-exoribonuclease activity; RNA binding	[151]

2.13 DNA binding of an enriched native DCC fraction

In order to test whether the native fraction enriched in DCC is suited for *in vitro* studies, initial DNA binding experiments were performed. Indeed proteins of the native fraction were specifically targeted to a known consensus upstream activator sequence (UAS) via the Gal4DBD of MOF (Figure 2-14D). A control UAS, which contains two mutations described to disrupt binding of the Gal4DBD [156], also disrupted the interaction between Gal4DBD-MOF and DNA. These initial experiments indicate that in general sequence-specific binding can be detected using the *in vitro* system and that non-specific DNA binding activities are not detectable in the enriched fractions. Therefore this approach represents a starting point to study DNA binding *in vitro*, not only with recombinant but also with native fractions.

3 Discussion

3.1 Multimerization of the MSL2-MSL1 complex

The key subunits of the DCC – MSL2 and MSL1 – were expressed and purified as individual proteins and in form of a heteromeric complex. The analysis of the stoichiometry of the purified proteins revealed that MSL2 does not exist as a monomer but as a heterogeneous mixture of complexes with variable size – at least under the buffer conditions used in this study. The same is true for the MSL2-MSL1 subcomplex, which forms a complex larger than a heterotetramer of equal stoichiometry. Since no defined complex is formed, it is possible that during the gel filtration the MSL2 or MSL2-MSL1 complex is unstable and falls apart into different complexes of variable sizes. The MSL2-MSL1 subcomplex has not been analyzed so far and thus comparison with findings from other studies is not possible. However, a MSL2-MSL1-MSL3 complex could be purified from the male *Drosophila melanogaster* SL2 cell line and was analyzed by gel filtration [79]. A similar elution pattern over a wide molecular weight range was observed, despite the fact that a different source (recombinant versus native proteins) and different buffer conditions were used in their study [79]. The observed heterogeneity of complexes may be due to non-specific aggregation in the absence of a complete DCC. However, it is also possible that there might be a functional relevance for a large MSL2-MSL1 complex. For example, such an oligomerization may be responsible and even required for the very stable interaction of MSL2 *in vivo* as shown by FRAP analysis [100].

In addition, formation of large multimeric protein complexes is not unusual for complexes that are involved in modifying chromatin architecture, as shown for the GAGA factor (GAF). It is thought that forming a multimer enables GAF to bring together different regulatory regions, which are otherwise separated by long distances, in order to create a single regulatory structure [157]. One interesting aspect about GAF are its self-association properties, mediated by two domains, the BTP/POZ and the polyglutamine-rich Q domain [158]. The Q domain by itself is capable of binding single stranded DNA, whereas the entire GAF protein is even capable of binding triple stranded DNA containing GA-repeats [159]. Given that there might be triple helix structures involved in dosage compensation, GAGA factor could be in principle a candidate binding protein (see below). The multimerization of GAF leads to complexes of 150 - 300 Å in diameter [160]. The formation of GAF multimers brings several DNA binding domains together and hence allows the interaction with several

DNA binding sites, which in turn enhances DNA binding. It was even shown by using electron microscopy that GAF is able to connect two separate DNA molecules *in vitro* [160]. A similar observation was also found *in vivo*, where GAGA enhancer sequences could activate transcription from a GAGA promoter by acting *in trans*, which means that both type of sequences were located on different plasmids [161]. This suggests that GAF acts in bridging enhancer and promoter sequence to bring those distantly located regulatory elements together. A similar mode of action could be proposed for the MSL2-MSL1 complex. First, in this study the MSL2-MSL1 complex was also shown to form multimers and second, it was recently shown that the MSL2-MSL1 complex sets up a male-specific organization of the X chromosome [96]. Measuring the 3D distances of HAS in male and female nuclei revealed a closer distance in males. Since reducing the levels of MSL2-MSL1 (but not of MSL3) by RNAi abolished the distance differences between male and female nuclei it is clear that this long-range interaction is dependent on the MSL2-MSL1 subcomplex. Oligomerization may therefore be required for MSL2-mediated clustering of HAS *in vivo* (Figure 3-1E and [96]).

3.2 The DNA binding factor of the MSL2-MSL1 complex

The current model of HAS implies that within the MSL2-MSL1 core unit, which is the minimal unit required for initial recognition of HAS, a DNA binding function must exist. This has never been shown – until now. By combining *in vitro* binding assays (EMSA) and recombinant MSL proteins, a DNA binding function within the MSL2-MSL1 complex was mapped. It is the MSL2 protein that is capable of DNA binding *in vitro*, whereas its interaction partner MSL1 is not. The fact that MSL1 is dispensable for DNA binding is rather surprising, since Scott and colleagues proposed that within MSL1 a short N-terminal region might be required for binding to DNA [98]. This was based on the observation in transgenic flies that a MSL1 truncation, which lacked the first 26 amino acids, was not associated with the polytene X chromosome anymore. However, the mechanism was not addressed. The deletion could have affected the targeting, e.g. by impairing DNA binding or simply by disrupting the nuclear localization. In an earlier study the same group showed that MSL2 binding to the polytene X chromosome was reduced, but not abolished when MSL1 lacked the first 76 amino acids [97]. MSL2 therefore does not need the N-terminal region of MSL1 for X chromosome binding. Although its ability may be reduced in the absence of MSL1, MSL2 clearly has the ability to bind the X chromosome by itself. Those *in vivo* findings support our *in vitro* data, which suggests MSL2 as the factor that binds to DNA.

Despite the fact that MSL1 itself is not capable of DNA binding it could indirectly influence DNA binding of its partner MSL2. Such heterodimer formations are quite common for many DNA binding proteins, for example the very large family of steroid hormone receptors. Those ligand-activated transcription factors share a well-conserved DNA-binding domain, which upon dimerization allows the binding to hexameric hormone response elements, either oriented as direct or inverted repeats [125]. Some members of the nuclear hormone receptors, for example the retinoic X receptor RXR, are able to bind as a homodimer, but forming a heterodimer with different other hormone receptor monomers, such as the vitamin D₃ or thyroid hormone receptor, is known to alter the affinity and selectivity for the DNA binding site [126]. Judged from the results presented in this study such a mode of action is not the case for MSL2 and MSL1, since the MSL2-MSL1 complex has the same affinity for DNA as the MSL2 protein alone, showing that MSL1 does not increase the affinity of the heteromer. It can be concluded that MSL1 is not needed for DNA binding. This finding is supported by results from the *in vivo* reporter gene assay: When the domain that is required for interaction with MSL1 – the MSL2 RING finger – was deleted, the reporter gene activation could still be stimulated. This demonstrates that MSL2 is able to act without MSL1 in cells. The summary of our *in vitro* and *in vivo* binding data leads to the hypothesis that it is mainly MSL2, which endows the DCC with DNA binding capacity.

3.3 CXC domains as DNA binding modules

3.3.1 DNA binding by the CXC domain of MSL2

In this study, the DNA binding domain within MSL2 was mapped to the CXC domain and it was found to be required but not sufficient for DNA binding. In addition it was shown that the impaired DNA binding function also correlated with impaired targeting of the DCC to the X chromosome *in vivo*. Surprisingly, previous studies did not highlight the CXC domain as needed for X chromosome targeting. A series of MSL2 truncations had been analyzed in transgenic flies for their ability to bind to the polytene X chromosome [99]. Deleting a large region that included the CXC domain and the Pro/Bas patch led to a displacement from the usual binding sites to the pericentromeric heterochromatin. A construct that included again the CXC domain did not change this mislocalization, which led the authors to speculate that the CXC domain was not needed for targeting. However, this conclusion cannot be drawn, since all MSL2 truncations tested in their assay perturbed MSL2 function. Only when all

domains were present, which is in fact almost the full-length protein, MSL2 could localize normally to the thousands of sites on the X chromosome [99].

In order to further characterize the role of the CXC domain for DNA binding, two point mutations of conserved cysteines within the CXC domain were tested in addition. As for the deletion of the complete CXC domain also for the point mutations it was confirmed that impaired DNA binding is correlated with impaired *in vivo* targeting in cell culture. The *in vivo* relevance of the point mutations is not only based on our results in cells, but is further supported by early work from the Kuroda lab with transgenic flies [80]. The very same point mutations did not cause lethality for male flies, however, fly development was delayed and their viability reduced [80]. The *in vitro* studies presented here would therefore offer an explanation for this *in vivo* defect, namely that the mutation of the conserved cysteines disturbs, but does not abolish the DNA binding function of the CXC domain, which consequently leads to reduced affinity of MSL2 to DNA. Since the affinity is only reduced and DNA binding is not completely destroyed the effects on male viability are also not that extreme.

It is possible that in MSL2 the CXC domain cooperates with other structures, such as the Pro/Bas patch. The deletion of this patch led to a slightly lower affinity of MSL2 for nucleic acids and to a modest reduction of MSL2 targeting to a HAS in cells. Further evaluation of these hypotheses will require knowledge of the MSL domain structures at atomic resolution.

3.3.2 CXC domains and their function in DNA binding

The CXC domain is not only found in the MSL2 protein, but also other proteins, from plants to humans, contain CXC domains, with the exceptions of yeast and other fungi [146]. According to Marin these CXC domains can be placed into different groups, which vary in their N-terminal sequence: The Enhancer of Zeste E(Z), the TSO-1 and the MSL2 CXC domains [146]. Common to all groups is the presence of one to three CXC motifs at the N-terminal part and the general C-X₄-CXC-X₆-C-X₄₋₅-C-X₂-C formula at the C-terminal end of the domain. Often two CXC domains occur separated by a variable linker length within the same protein. [146]. Not only in MSL2, but also in other *Drosophila* proteins, as well as in plant proteins several mutations of the conserved cysteines have been described [80]. For example mutations in the genes of *Drosophila enhancer of zeste* [162] and of *Arabidopsis thaliana TSO-1* [163, 164] are known to cause developmental defects in the animals and in the plant. But it remained unclear how those mutations affected the function of the protein.

Recently, the tesmin/TSO1-like protein from *Arabidopsis thaliana*, which is composed of two CXC domains separated by a conserved linker region, was analyzed *in vitro* for its ability to bind different metal ions [165]. Zinc was the preferred metal ion that was bound, probably at least four zinc ions per two CXC domains [165]. Therefore replacing cysteines could impair chelating of zinc ions and subsequently could affect protein function. Two recent studies on novel *Drosophila* testis-specific protein complexes (tMAC) identified proteins of the tesmin/TSO1-family (dMip120 and Tombola), which share one or two CXC domains, respectively [166, 167]. *Drosophila* Mip120 was already found in the dREAM complex, which functions in DNA replication [168] and repression of a subset of genes with developmental- or sex-specific expression patterns [169, 170]. Similar to polycomb complexes, dREAM associates with transcriptionally silent chromatin, however, the binding sites of polycomb and dREAM complexes do not overlap [169]. Direct binding to DNA is likely achieved by the action of Mip120. In the testis-specific tMAC complex, Mip120 was not found, but a different family member, Tombola, which in contrast to Mip120 possesses only one CXC domain [166, 167]. Interestingly, two proteins of the tMAC complex (Topi and Achi/Vis) are thought to have inherent sequence-specific DNA-binding activity. This would mean that several DNA binding functions are contained together in the same complex, offering a way to increase the repertoire of potential DNA binding sites [167]. It is also speculated that two of the proteins (Topi and Achi/Vis) could bind with low affinity to the target promoter. The presence of the CXC domain containing protein Tombola could allow then the interaction with the two others, which in turn would lead to cooperative DNA binding [167]. It should be mentioned that cooperativity would not necessarily require direct binding of Tombola to DNA; a DNA-independent interaction with the two other, already DNA bound, proteins could be sufficient. Whether the CXC domain of Tombola is indeed functioning as a DNA binding structure has to be shown.

3.3.3 The CXC domain – a conserved DNA binding module

So far DNA binding was only demonstrated for one related member of the tesmin/TSO1-family: The CPP1 protein of the soybean is involved in the regulation of the expression of leghemoglobins, which are needed in the symbiotic root nodule to transport oxygen to the respiring bacteria [171]. It could be shown that the two clustered CXC domains of CPP1, which are separated by approximately 60 amino acids, are sufficient to bind DNA *in vitro*. In addition, the two CXC domains were able to specifically bind to a promoter fragment of a

leghemoglobin gene, showing that the CXC domains alone are sufficient to provide binding specificity [171]. Since both CXC domains have the same general CXC formula (see above) like MSL2 and others, it is clear that other N-, C-terminal and linker regions must contribute to the recognition in order to ensure binding to the correct target sequence. The affinity of the CPP1 CXC domains to DNA, however, was not determined precluding a comparison with results obtained in this study. When just the CXC domain of MSL2 by itself was used, DNA binding could be detected, too. Nevertheless, the affinity was much weaker compared to the full-length MSL2 protein, indicating that – in contrast to CPP1 – additional regions of MSL2 were required for high affinity binding.

Because of the increased insight into the roles of CXC domain containing proteins it is tempting to suggest that all members of this protein family use their CXC domain to directly contact DNA [166]. However, such a function has to be proven experimentally for every single protein. But such a common role of those domains is consistent with our finding that a related CXC domain from a *Homo sapiens* protein can functionally replace the original *Drosophila melanogaster* CXC domain of MSL2 in DNA binding *in vitro*. About the human KIAA1585 protein not much information exists. It was discovered in a cDNA screen of unidentified human genes [172]. Homology database search classified the protein as being similar to *Drosophila melanogaster* MSL2 [146]. KIAA1585 and MSL2 were the two only proteins found in the databases that share the unique combination of a RING finger and a CXC domain, suggesting that they may form a functional module [146]. The data presented here do not support this speculation, since deleting the RING finger does neither affect DNA binding nor reporter gene activation in cells. Obviously for the complete function of MSL2, both domains are needed and maybe might even form intramolecular interactions. The CXC domains of MSL2 and its human orthologue share one N-terminal CXC motif, which is linked to the C-terminal C-X₄-CXC-X₆-C-X_{4,5}-C-X₂-C motif by a spacer of the same length (11 amino acids). Several amino acids between the conserved cysteines (depicted as X in the formula) are also identical [146]. The similarity of the conserved cysteines and their spacing are indeed sufficient to enable DNA binding. On the other hand, the chimeric MSL2 protein could not activate the reporter gene *in vivo*, showing that for MSL2's transactivation ability, in contrast to its DNA binding ability, also the non-conserved amino acids are needed.

3.4 MSL2 – an additional RNA binding protein within the DCC

The *in vivo* crosslinking analyses had pointed to sequence motifs that are enriched in HAS [93, 94], but they could not distinguish whether these sequences attract the DCC as double stranded or single stranded DNA or as RNA. The relative affinities of MSL2 for the DBF12-L15 sequence presented in the form of ss and dsDNA or RNA was determined in a series of competition assays. Single stranded nucleic acids could not be bound by MSL2, showing that the structure of a double helix is an essential feature that is used by the protein for recognition of the target site. However, we also observed binding of MSL2 to dsRNA (with similar affinity), which is quite surprising, since so far only the DCC components MSL3 and MOF were shown to be able to bind RNA, whereas MSL2-chromosome interactions *in vivo* are relatively RNase-insensitive [173, 174]. The RNA binding is not brought about by the CXC domain and also the RING finger and the Pro/Bas patch are dispensable for RNA binding, showing that the RNA binding function within MSL2 cannot be located. It was speculated by Scott and colleagues that the Pro/Bas patch, which is positively charged because of a cluster of basic amino acids, might be responsible for roX RNA binding. Such a simple positively charged stretch is found frequently within RNA binding domains and is a way for proteins to form contacts with RNA. The deletion of the Pro/Bas patch modestly reduced MSL2-mediated activation of the reporter gene, which points to a role for this patch *in vivo*. On the other hand, RNA binding *in vitro* was almost not affected, but it might well be that after deleting the Pro/Bas region, the RNA is bound by another, secondary binding domain (the CXC domain or the RING finger). The observed slight sigmoid shape of the competition curve would support this idea of the existence of more than one RNA binding site within MSL2. The presence of an additional novel RNA binding protein within the DCC would offer new exciting possibilities of regulating dosage compensation.

3.5 Recognition of high affinity sites

This study documents the first quantitative measurements of the binding of MSL proteins to a potential DNA HAS. The affinity constant K_D of MSL2 to the short 40 bp DBF12-L15 DNA element is around 33 nM. Such low nanomolar affinities are not unusual and in fact are typical for other DNA binding proteins, such as hormone receptors, which bind specifically to their consensus hormone response elements. For example recombinant glucocorticoid receptor (GR) was tested in gel shift assays for its ability to bind to different

hormone response elements. GR bound *in vitro* a 33 bp long DNA sequence, which contained a consensus hormone response element, with a K_D of around 21 nM [175]. In contrast to this high affinity binding, the same receptor bound a related response element of the androgen receptor with only low affinity (165 nM) [175]. The difference between 21 versus 168 nM does not seem to be large, but an eight times reduced affinity can have a strong impact on the *in vivo* function: The reduced affinity of GR to the target sequence led to a drastic reduction of GR-induced transcriptional activation (from 7 to 1.6 fold activation) in cell culture [175]. Moreover, interfering with the DNA binding function does not only show effects in cell culture systems, but can also have serious consequences for the health of an entire organism. The androgen insensitivity syndrome in humans is characterized by the inability to respond to the hormone. One cause for this syndrome can be mutations within the androgen receptor and its DNA binding domain. This example clearly illustrates that also a six times reduced affinity (from 33 to 189 nM) of MSL2- Δ CXC in comparison to the wild type protein can suffice to affect the protein function also *in vivo*. Indeed a similar reduction of MSL2-induced reporter gene activation (from 6 to 1 fold) and a mislocalization of MSL2 in cell culture were observed.

Currently it is thought that the hierarchy of DCC binding sites observed *in vivo* is based on their different affinities to a DCC subcomplex. However, recombinant MSL2 bound all DNA sequences with similar affinity. Since clustering of several sequence motifs was shown *in vivo* to create an even stronger HAS [92], a trimerized DNA element and in addition the well-characterized HAS within the *roX1* gene [83] was also tested. The affinity of MSL2 to those HAS was indeed increased (K_D around 1 nM), but again, a control DNA with similar length was bound equally well. In summary, MSL2 is capable of DNA binding, however, binding is sequence-independent and it seems that affinity depends on the length of the DNA, simply because more binding sites are available. The simple model by which the MSL2-MSL1 subcomplex distinguishes target sequences just because of different affinities cannot be confirmed by the *in vitro* findings presented in this study.

3.6 Alternative binding mechanisms

So which other principles may contribute to binding selectivity? It is obvious that our purified *in vitro* system so far includes a few components. It may well be that additional factors are required for achieving sequence-specific binding. Previously, the role of roX RNAs was proposed to alter the binding specificity of the MSL2-MSL1 complex by an

unknown principle, but needing the Pro/Bas patch [99]. However, it has to be noted that in this *in vivo* study not only the deletion of the Pro/Bas patch, but also other MSL2 deletions disturbed the targeting of MSL2 to the polytene X chromosome. Only when all domains of MSL2 were present roX RNA could efficiently be incorporated (by the known RNA binding proteins of the DCC) and consequently the regular targeting was achieved [99]. Consequently, the role of the Pro/Bas patch as a RNA binding domain is not convincing yet and needs further examination. The data presented in this study does not exclude a role of the Pro/Bas patch in HAS binding via roX RNA incorporation: RNA binding of MSL2 *in vitro* is slightly altered when the Pro/Basic patch is deleted, however, the binding affinity is decreased by less than 1.5-fold. The same MSL2- Δ Pro/Basic protein shows a reduced activation of the reporter gene, yet it is still a robust two-fold activation. In summary, the Pro/Bas patch and roX RNA might play a role for binding to HAS *in vivo*, but how this could be achieved is still not understood.

3.6.1 The nature of HAS: Involvement of triple helix structures?

So far it is not clear in which conformation the sequence that characterize HAS have to be presented. A non-B form DNA, a triple helix structure, or the context of chromatin might be required. The nucleosomal organization of the candidate sequence was not explored since HAS tend to reside in nucleosome-free regions [93, 94]. Instead, the involvement of triple helix structures was addressed: roX RNA might play a role in sequence-specific recognition, but not by interaction with the MSL2-MSL1 complex but rather by specific interaction with the DNA high affinity site itself. Such a RNA-DNA triplex could then be a more specific target for MSL2 (Figure 3-1B). The main question is, whether a HAS would be suited for triplex formation. The DNA major groove possesses several acceptor and donor groups that would allow hydrogen bonding with a third strand. This also means that not every sequence is suited for forming a triplex [148]. For example the homopurine strand of the duplex can contact a third strand (homopyrimidine) by forming T-A*T and C-G*C⁺ base triplets; however, a requirement is that the imino group of cytosine is protonated. But since its pKa is well below 7, the protonation only works under acidic pH conditions [176], which dramatically reduces the applicability of triplex formation *in vivo*. A screening under neutral pH conditions, revealed that a homopurine-homopyrimidine DNA duplex could be bound by RNA strands containing only pyrimidines, yet, the affinities were much weaker at neutral pH by comparison to pH 6.0 (57 versus 400 nM) [149]. It is striking that HAS are characterized

by GA repeats (homopurines), which would in theory be suited for formation of a triple helical structure with a third homopyrimidine RNA molecule. Our initial experiments showed that triplex formation with a perfect recognition motif is possible, but the short homopurine stretch in the 40 bp long DBF12-L15 HAS was not sufficient to bind the homopyrimidine RNA strand. Triplex formation was discussed to be involved in several biological systems. For example triplex formation at AG-rich satellite repeats of *Drosophila* could be possible *in vivo*, since the physiological copper concentration allows triplex formation even at neutral pH [177]. In the telomerase enzyme, which copies a short fragment of its integral telomerase RNA onto the ends of eukaryotic chromosomes, a triple helix formed only by RNA molecules was found to be essential for its function [178]. Also the GAGA factor (GAF) was shown to bind triple stranded DNA *in vitro* [159], which is discussed in the section below. In the context of dosage compensation it has been already tested, whether roX RNA could form an RNA-DNA duplex or triplex with roX1 or roX2 DHS *in vivo*. The deletion of roX RNA fragments that were complementary to the DNA binding site did not affect DCC binding to the roX binding sites on the polytene X chromosome, ruling out an involvement of at least those triplex structures formed between roX RNA and the DHS, in target site recognition [84].

3.6.2 Factor X required for sequence-specific binding?

Alternatively, an unknown protein factor X could be involved in the sequence-specific DNA binding (Figure 3-1C). Two scenarios are plausible: Factor X could have an intrinsic binding specificity for the consensus sequence and after its binding to DNA may recruit the MSL2-MSL1 complex whose sequence-independent DNA binding may then stabilize the interaction at this site. Alternatively, factor X could interact with MSL2-MSL1 to alter the complex conformation, which then could cause cooperative, that is sequence-specific binding to the consensus sequence. It should be highlighted that the proposed mechanisms require the DNA binding ability of MSL2, in order to provide a first unspecific contact with DNA or to stabilize the interaction with DNA. The question was addressed whether additional factors besides MSL2-MSL1 are needed for recognizing a HAS *in vivo*. In human cells that should not contain a HAS co-factor, MSL2 and MSL1 affected transcription of a reporter gene only very weakly. Moreover, this mild effect was independent of the binding site: Independent of whether the reporter gene construct carried the roX1 DHS high affinity site or not, this mild decrease of activation was observed. By contrast, in *Drosophila* SL2 cells the roX1 DHS

caused a strong, 20-fold MSL2-dependent activation [92]. These results favor the idea that MSL2-MSL1 are not sufficient to recognize HAS and that an additional factor exists, which is specific to *Drosophila*. It has to be noted that MSL2 and MSL1 could in fact work in another system than *Drosophila*. In a yeast two-hybrid assay, MSL2 and MSL1 were attached to a Gal4DBD and both fusion proteins led to activation of transcription upon binding to a UAS binding site [79]. One can conclude from those finding that an activation potential, either intrinsic or brought about by interaction with other transcriptional activators, is functional in yeast, however, it is the DNA binding function that is affected in other systems. Of course this could simply be because of wrong folding or missing posttranslational modifications of the protein, but also because a factor X, which brings about specificity, is not present in the artificial *in vitro* and also not in other *in vivo* systems.

The occurrence and clustering of GA repeats in HAS led to speculations about an involvement of the GAGA factor (GAF) in dosage compensation. GAF is known to act in cooperation with activating trxB (trithorax group) as well as with silencing PcG (polycomb group) complexes to activate or repress expression of homeotic genes, which control the proper realization of the body plan during animal development [157]. Homeotic genes as well as other developmental genes contain GAGA motifs to which the GAF is recruited [179, 180]. Its DNA binding domain, which is composed of a classical Cys₂His₂ zinc finger and three short basic regions [181, 182], was by itself able to specifically bind to GAG trinucleotide sequence in EMSAs [183]. Moreover, GAF's Q domain by itself is capable of binding single stranded DNA, whereas the entire GAF protein is even capable of binding triple stranded DNA containing GA-repeats [159]. Given that there might be triple helix structures involved in dosage compensation (see above), GAGA factor could be in principle a candidate binding protein. Since GA repeats were found in the proposed DCC consensus sequence (MRE), it was speculated that GAF could be factor X that binds to the target sequence and subsequently could recruit MSL2-MSL1. To investigate this idea GAF protein levels were reduced in SL2 cells by using RNAi, however, the activation of the reporter gene was not affected (C. König, P. Becker lab, unpublished results). Interfering with GAGA function in flies showed on the one hand increased male lethality but on the other hand MSL2 was still normally localized to its binding sites on the polytene X chromosome [184], excluding a role of GAF in dosage compensation.

GAGA motifs, although they have to be presented as longer stretches, can also be bound by another developmental regulator, the Pipsqueak protein, using a conserved helix-turn-helix motif [185]. Pipsqueak and GAF colocalize to hundreds of sites on the polytene

chromosomes, and protein-protein contacts are mediated via their BTB/POZ domains, suggesting that both proteins have similar functions [186]. Hence it is possible that disrupting GAF function by RNAi or by mutations in the coding gene as done in the experiments mentioned above will be without effect because Pipsqueak could carry out GAF's function.

3.6.3 High affinity sites as allosteric effectors?

In this study sequence-specific binding of MSL2 was not shown, however, affinity differences may not be necessary for the targeting to HAS (Figure 3-1A, D). The following scenarios may be considered. The NF- κ B transcription factors bind to a relatively loose consensus sequence and it is not clear how these sequences define specific binding sites. It was found that the binding site does rather not determine the effectivity of binding, but it determines which coactivators get recruited to the NF- κ B transcription factors [187]. This indicates that the DNA sequence is not just a docking site but can influence the conformation of the DNA-binding protein and subsequently change its function. A similar mode of action was proposed for the glucocorticoid receptor (GR). It was often found that despite the high affinity of the GR to a binding site the transcriptional output was low or that even a very weak affinity (> 1000 nM) led to extremely high transcriptional activation [188]. This suggests that selective activation is not only determined by the affinity to a certain DNA sequence. The DNA sequence was shown to trigger a conformational change within the DNA binding domain of the GR and as a consequence affected the regulatory activity of the protein [188]. Moreover, selectivity for target sequences do not need to be based on binding affinity, but rather on conformational changes subsequent to binding, as shown for DNA topoisomerase II. In the context of the DCC the following scenario may be envisioned: MSL2 would scan DNA using the sequence-independent binding ability of its CXC domain. Only when a particular sequence (e.g. HAS) acts as an allosteric effector it would trigger a conformational change within MSL2 which subsequently would allow effective recruitment of MSL1 or assembly of the entire DCC (Figure 3-1D).

3.7 Outlook

In this study, the direct interaction of recombinant DCC proteins with candidate binding sites was analyzed. Indeed, employing a biochemical approach allowed assigning a DNA binding function within the DCC, however, with no preference for certain DNA sequences *in*

vitro. The simple mechanism by which the DCC recognizes target sites based on their affinity cannot be confirmed by this study suggesting that other targeting principles must be involved (Figure 3-1). Within this work, indications for an unknown specificity factor, which is only present in *Drosophila*, emerged. Its identification and biochemical characterization would be crucial for understanding X chromosome targeting. Alternatively, sequence-specificity does not necessarily have to be based on affinity differences, but rather on triggering a conformational change by the target sequence. However, obtaining proof for such a mechanism is only possible by extending the biochemical analysis and ideally combining it with structural data of the MSL2-MSL1 complex.

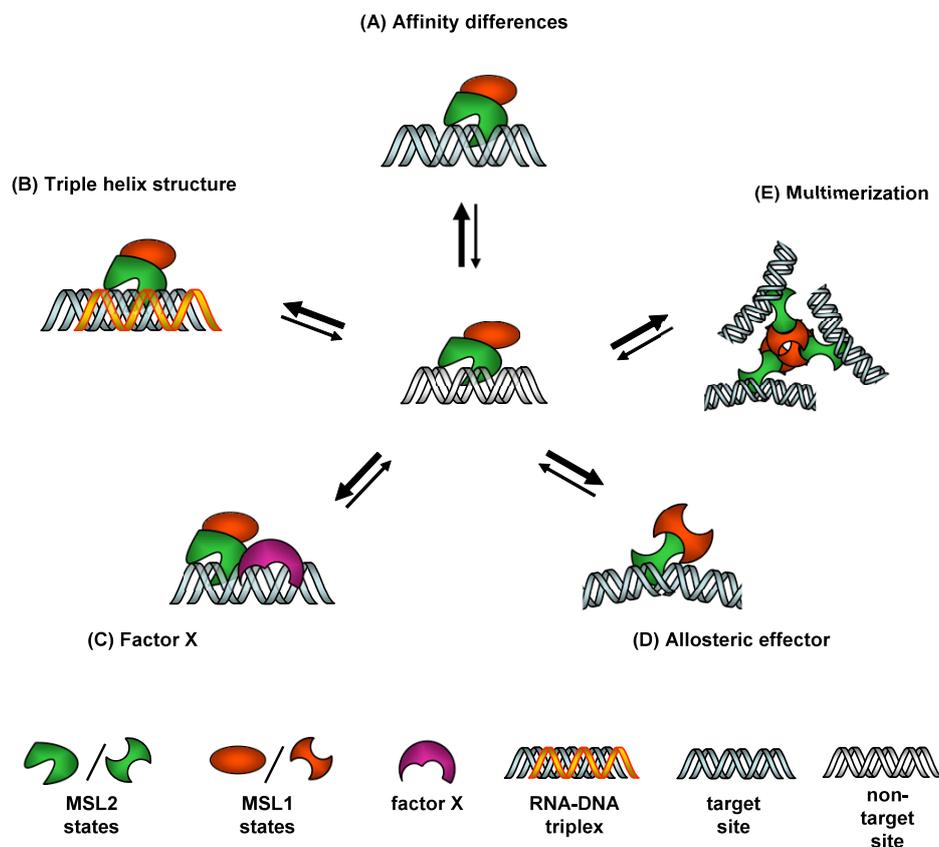


Figure 3-1: Alternative targeting principles. In all cases, the CXC domain of MSL2 is required for establishing protein-DNA contacts, whereas MSL1 is dispensable for DNA recognition, but might be needed for association with chromatin (not illustrated here). (A) Recruitment is based on different affinities of MSL2 to target and non-target sequences. However, biochemical data presented in this study does not support this model. (B) The recognition of target sites by MSL2 requires triple helix formation, e.g. between dsDNA of HAS and a single stranded region of roX RNA. (C) Sequence-specific DNA binding is mediated by binding of an unknown factor X, which is specific for *Drosophila*. Nevertheless, binding of MSL2 via its CXC domain is needed for stabilization. (D) The target site DNA (e.g. HAS) triggers a conformational change in MSL2, allowing efficient recruitment of MSL1 or other DCC subunits. (E) Multimerization of MSL2-MSL1 is needed to bring HAS together to form a male-specific nuclear compartment. The models proposed here are not mutually exclusive and can be combined. For example factor X (C) might be required to bind the triplex structures (B) and multimerization (E) does not necessarily require a conformational change of MSL2 (D) but could occur for all other MSL2 states as well (A-C) (pictures were created using the motifool toolkit).

4 Materials and Methods

4.1 Materials

4.1.1 General chemicals

All common chemicals were ordered from Merck (Darmstadt), Roth (Karlsruhe) or Sigma (Taufkirchen), exceptions and special chemicals are listed below.

Acrylamide (Rotiphorese Gel 30)	Roth, Karlsruhe
Agarose (LE GP)	Biozym, Hessisch Oldenburg
[γ - ³² P]- ATP	Perkin Elmer, Waltham, MA, USA
BSA (Bovine serum albumin), 98% pure	Sigma, Taufkirchen
BSA, purified	NEB, Frankfurt/Main
Coomassie G250	Serva, Heidelberg
[α - ³² P]-dCTP	GE Healthcare, Munich / Hartmann Analytic, Braunschweig
dNTP (single)	Roche, Mannheim
FLAG peptide	Sigma, Taufkirchen
Formaldehyde	Polysciences, Inc., Warrington, PA
Glutathione	Sigma, Taufkirchen
IPTG	Roth, Karlsruhe
NP-40 (Igepal CA-630)	Sigma, Taufkirchen
PMSF (Phenylmethanesulfonyl fluoride)	Sigma, Taufkirchen
Protease inhibitors: Aprotinin, Leupeptin, Pepstatin	Sigma, Taufkirchen
Spermidine trihydrochloride	Sigma, Taufkirchen
Triton X-100	Sigma, Taufkirchen
Tween 20	Sigma, Taufkirchen

4.1.2 Chemicals for tissue culture

Blasticidin	A.G. Scientific, San Diego, CA, USA
Gentamycin	PAA, Cölbe
Penicillin / Streptomycin	Sigma, Taufkirchen
Trypan Blue	Sigma, Taufkirchen

Effectene Transfection Reagent	Qiagen, Hilden
FuGENE HD Transfection Reagent	Roche, Mannheim
FBS (fetal bovine serum)	Sigma, Taufkirchen
DMEM High Glucose 4.5 g/L (+ L-Glutamine, + Sodium Pyruvate)	PAA, Cölbe
Schneider's <i>Drosophila</i> medium (+ L-Glutamine)	Invitrogen, Karlsruhe
Sf-900 II SFM Medium	Invitrogen, Karlsruhe

4.1.3 Antibodies

Primary antibodies

α MSL2	polyclonal rabbit serum (SABC), used 1:1000 in WB
α MSL1	polyclonal rabbit serum, used 1:1000 in WB and IF, kind gift from E. Schulze
α MSL3	affinity purified monoclonal mouse antibody (clone 1C9-5), used 1:50 in WB
α MOF	polyclonal rabbit serum (SA4897), used 1:1000 in WB
α MLE	monoclonal rat antibody, hybridoma supernatant (clone 6E11), used 1:1000 in WB
α GFP	monoclonal mouse antibody (clone 3E6), used 1:400 in IF, purchased from Molecular Probes (Invitrogen)
α FLAG	monoclonal mouse antibody, used 1:2000 in WB, purchased from Sigma
α H4K16ac	polyclonal rabbit antibody, used 1:1000 in IF, purchased from Active Motif

Secondary antibodies

α -rabbit HRP-conjugated	Amersham, GE Healthcare, Munich
α -mouse HRP-conjugated	Amersham, GE Healthcare, Munich
α -rat HRP-conjugated	Amersham, GE Healthcare, Munich
donkey α -rabbit IRDye 800-conjugated	Biomol, Hamburg
donkey α -mouse Alexa 488-conjugated	Invitrogen, Karlsruhe
donkey α -rabbit Cy3-conjugated	Invitrogen, Karlsruhe

Secondary antibodies were diluted 1:10000 when used for WB and diluted 1:500 when used for IF.

4.1.4 Bacteria and cell lines

<i>E.coli</i> BL21-CodonPlus (DE3)-RIL	Stratagene, La Jolla, CA, USA
<i>E.coli</i> DH10Bac	Invitrogen, Karlsruhe
<i>E.coli</i> XL1-Blue	Stratagene, La Jolla, CA, USA
HEK-293 cells (human embryonic kidney)	provided by M. Eckey, P. Becker lab
Sf21 cells (<i>Spodoptera frugiperda</i>)	Invitrogen
SL2 cells (<i>Drosophila melanogaster</i>)	provided by T. Straub, P. Becker lab

4.1.5 Dialysis, chromatographic and filtration material

Amicon Ultra-4 Centrifugal Filter Devices	Millipore, MA, USA
Anti-FLAG M2 Agarose from mouse	Sigma, Taufkirchen
Chromatography systems (ÄKTA, FPLC)	GE Healthcare, Munich
Gelfiltration column (Superose 6)	GE Healthcare, Munich
Glutathione Sepharose 4B	GE Healthcare, Munich
Micro Bio-Spin Chromatography Columns	Biozym, Hessisch Oldendorf
SpectraPor dialysis membrane	Roth, Karlsruhe

4.1.6 Kits and enzymes

All enzymes were purchased from New England Biolabs (NEB, Frankfurt/Main). All Kits were purchased from Qiagen (Hilden), with exception of the Dual Luciferase Reporter Assay System from Promega (Mannheim)

4.1.7 Oligonucleotides

For EMSA, DNA oligonucleotides were synthesized by Biomers (Ulm), whereas RNA oligonucleotides were synthesized by metabion (Munich). For cloning, DNA oligonucleotides were synthesized by MWG (Munich). Crucial consensus sequences (for details see ‘results’ section), restriction sites or mutated bases are underlined.

4.1.7.1 Oligonucleotides for EMSA

DBF12-L15 DNA fw	5'- TGCGGCCATCTCTTTTCGTTTTGATGTTTCTACGCCATGTG
DBF12-L15 DNA rv	5'- CACATGGCGTAGAAACATCAA <u>AAACGAAAGAGATGG</u>

DBF12-L18 DNA fw	5'- TCGGCCAAAAAATTCGTTTTGATGTTTCTACGCCATGTG
DBF12-L18 DNA rv	5'- CACATGGCGTAGAAACATCAAACGAATTTTTTGG
DBF12-L15 RNA fw	5'- UGCGGCCAUCUCUUUCGUUUUGAUGUUUCUACGCCAUGUG
DBF12-L15 RNA rv	5'- CACAUGGCGUAGAAACAUCAAAACGAAAGAGAUGGCCGCA
40TFO DNA fw	5'- TCGGGTCGAGGAAAAGAAAAGAAAAGGAAGCTCTTAGCGC
40TFO DNA rv	5'- GCGCTAAGAGCTTCCTTTTTCTTTCTTTCTTCCTCGA
27TFO DNA fw	5'- TCGAGGAAAAGAAAAGAAAAGGAAGC
27TFO DNA rv	5'- GCTTCCTTTTTCTTTCTTTCTTCCT
21TFO-S RNA	5'- UCCUUUUCUUUCUUUUUCCUU
16TFO-L15 RNA	5'- UUUUCCUUUCUCUCCC
UAS consensus DNA fw	5'- AGGTCGGAGGACTGTCCTCCGAGGT
UAS consensus DNA rv	5'- ACCTCGGAGGACAGTCCTCCG
UAS mutated DNA fw	5'- AGGTCCGAGGACTGTCCTCCGAGGT
UAS mutated DNA rv	5'- ACCTCCGAGGACAGTCCTCCG

4.1.7.2 Oligonucleotides for cloning and site-directed mutagenesis

Oligonucleotide name	Sequence from 5' to 3'	Description
Δ RING SalI fw	ACGCGTCGACATGCCTC GCATCAAGCTG	forward primer with SalI restriction site to clone MSL2- Δ RING into pFastBac1
MSL2 SalI fw	ACGCGTCGACATGGCCC AGACGGCATAAC	forward primer with SalI restriction site to clone MSL2- Δ CXC and MSL2- Δ Pro/Bas into pFastBac1
MSL2 FLAG NotI rv	GATTGCGGCCGCTTACT TATCGTCGTCATCCTTG TAATCCAAGTCATCCGA GCCCGA	reverse primer with NotI restriction site and cds for FLAG peptide to clone MSL2- Δ RING and Δ CXC into pFastBac1
Δ Pro/Bas FLAG NotI rv	GATTGCGGCCGCTTACT TATCGTCGTCATCCTTG TAATCGAAGATGATATT GAAGCCCTG	reverse primer with NotI restriction site and cds for FLAG peptide to clone MSL2- Δ Pro/Bas into pFastBac1
MSL2 KpnI fw	CGGGGTACCATGGCCCA GACGGCATAAC	forward primer with KpnI restriction site to clone MSL2- Δ Pro/Bas into pVP16
Δ Pro/Bas AgeI rv	CGCGACCGGTGGGAAGA TATTGAAGCCCTG	reverse primer with AgeI restriction site to clone MSL2- Δ Pro/Bas into pVP16
HsCXC PacI fw	CCTTAATTAAGCCGAAG TGTAATGTGGGCGTGC	forward primer with PacI restriction site to clone HsCXC domain into

	TA	pFastBac1 and pVP16
HsCXC <i>NheI</i> rv	CTAGCTAGCCGGATTCT TGCAGCCACGACATATA CAA	reverse primer with <i>NheI</i> restriction site to clone HsCXC domain into pFastBac1 and pVP16
MSL2 vect elong <i>PacI</i> fw	GGTTAATTAACGGAGGC TTCACCTTTTC	forward primer to exclude CXC domain and to introduce <i>PacI</i> restriction site into pFastBac1-MSL2 and pVP16-MSL2 vectors
MSL2 vect elong <i>NheI</i> rv	CTAGCTAGCCACAAGGA GGACTACGTTG	reverse primer to exclude CXC domain and to introduce <i>NheI</i> restriction site into pFastBac1-MSL2 and pVP16-MSL2 vectors
mut CC to AA fw	CCTGCCGGAATTCCCGA GCTCCTGCCTACAAGAG TTACAACAG	forward primer to change aa C544A and C546A by site-directed mutagenesis
mut CC to AA rv	CTGTTGTA ACTCTTGTA GGCAGGAGCTCGGGAAT TCCGGCAGG	reverse primer to change aa C544A and C546A by site-directed mutagenesis
mut Y to A fw	AATTC CC GATGTCCTTG CGCCAAGAGTTACAACA GCTGT	forward primer to change aa Y547A by site-directed mutagenesis
mut Y to A rv	ACAGCTGTTGTA ACTCT TGGCGCAAGGACATCGG GAATT	reverse primer to change aa Y547A by site-directed mutagenesis

4.1.8 Plasmids

Details of the cloning procedure for the individual plasmids can be found in the ‘cloning of expression constructs’ section.

Plasmid name	Description	Generated by
pFastBac1-MSL2-FLAG	used to create recombinant Baculovirus for expression of MSL2-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL2- Δ RING-FLAG	used to create recombinant Baculovirus for expression of MSL2- Δ RING-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL2- Δ CXC-FLAG	used to create recombinant Baculovirus for expression of MSL2- Δ CXC-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL2- Δ Pro/Bas-FLAG	used to create recombinant Baculovirus for expression of MSL2- Δ Pro/Bas-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL2-HsCXC-FLAG	used to create recombinant Baculovirus for expression of MSL2-HsCXC-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL2-C544A/C546A-FLAG	used to create recombinant Baculovirus for expression of MSL2-C544A/C546A-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL2-Y547A-FLAG	used to create recombinant Baculovirus for expression of MSL2-Y547A-FLAG in Sf21 cells	T. Fauth
pFastBac1-CXC-FLAG	used to create recombinant Baculovirus for expression of CXC-FLAG in Sf21 cells	T. Fauth
pFastBac1-MSL1-	used to create recombinant Baculovirus for	T. Fauth

FLAG	expression of MSL1-FLAG in Sf21 cells	
pFastBac1-MSL1	used to create recombinant Baculovirus for expression of untagged MSL1 in Sf21 cells	V. Morales
pcDNA3.1-MSL2-VP16	expression of MSL2-VP16 in mammalian cells	T. Fauth
pcDNA3.1-MSL2- Δ CXC-VP16	expression of MSL2- Δ CXC-VP16 in mammalian cells	T. Fauth
pcDNA3.1-MSL1-VP16	expression of MSL1-VP16 in mammalian cells	T. Fauth
pEGFP-N2	expression of GFP in mammalian cells	BD Biosciences
pVP16	expression of VP16 activation domain in SL2 cells	Clontech
pMSL2-VP16	expression of MSL2-VP16 in SL2 cells	C. König
pMSL2- Δ RING-VP16	expression of MSL2- Δ RING-VP16 in SL2 cells	C. König
pMSL2- Δ CXC-VP16	expression of MSL2- Δ CXC-VP16 in SL2 cells	C. König
pMSL2- Δ Pro/Bas-VP16	expression of MSL2- Δ Pro/Bas-VP16 in SL2 cells	T. Fauth
pMSL2-HsCXC-VP16	expression of MSL2-HsCXC-VP16 in SL2 cells	T. Fauth
pMSL2-C544A/C546A-VP16	expression of MSL2-C544A/C547A-VP16 in SL2 cells	T. Fauth
pMSL2-Y547A-VP16	expression of MSL2-Y547A-VP16 in SL2 cells	T. Fauth
pGL3-TK-Luc_mod	inducible firefly <i>luciferase</i> reporter	C. König
pRL-TK 'Renilla'	constitutive <i>Renilla luciferase</i> reporter	Promega
pGL3-DBF12-L15	DBF12-L15 binding site upstream of firefly <i>luciferase</i> reporter	C. König
pGL3-DBF12-L18	DBF12-L18 binding site upstream of firefly <i>luciferase</i> reporter	C. König
pGL3-3xDBF12-L15	Trimerized DBF12-L15 binding site upstream of firefly <i>luciferase</i> reporter	C. König
pGL3-roX1DHS	DHS of roX1 upstream of firefly <i>luciferase</i> reporter	C. König
pBluescript KS+	mcs of varying length was used in EMSA	Stratagene
pP12eL(15)3	pCASPER vector with 3 copies of the DBF12-L15 binding site. 167 bp fragment was used in EMSA	I. Dahlsveen
pUCA-roX1-DHS short	pUC vector with roX1-DHS binding site. 226 bp fragment was used in EMSA	G. Mengus
pGEX-2KG	inducible expression of GST in <i>E.coli</i>	Amersham
pGEX-2KG-CXC	inducible expression of GST-CXC in <i>E.coli</i>	T. Fauth
pEGFP-hsp-MSL2	used for expression of GFP-MSL2 in a stable SL2 cell line	T. Straub
pEGFP-hsp-MSL2- Δ CXC	used for expression of GFP- Δ CXC in a stable SL2 cell line	T. Straub
pEGFP-MSL2-	used for expression of GFP-C544A/C547A in a	T. Fauth

hsp-C544A/C547A	stable SL2 cell line	
pY3-Gal4-MOF-FLAG	used for expression of MOF-Gal4DBD-FLAG from its endogenous promoter in a stable SL2 cell line	M. Prestel
pY3-w/oGal4-MOF-FLAG	used for expression of MOF-FLAG from its endogenous promoter in a stable SL2 cell line	T. Fauth
pCoBlast	selection vector carrying Blasticidin resistance to create stable SL2 cell lines	Invitrogen

4.1.9 Baculoviruses

Recombinant baculoviruses were created by using pFastBac1 vectors (see section above) and the Bac-to-Bac expression system (Invitrogen). Amplification and maintenance of viruses was performed according to the manufacturer's instructions (Invitrogen). All baculoviruses were created within this study, with the exception of the MSL1 (untagged) baculovirus, which was created by V. Morales [75].

4.1.10 Other materials

DE81 anion exchanger chromatography paper	Whatman, Dassel
ECL detection system	GE Healthcare, Munich
Immobilon-P PVDF membrane	Millipore, MA, USA
Nitrocellulose membrane	Whatman, Dassel
Polystyrene tubes, 5 ml, 75 x 12 mm	Sarstedt, Nümbrecht
Protein marker	Peqlab Biotechnologie, Erlangen
Siliconised reaction tubes, 1.5 ml	Biozym, Hessisch Oldendorf
Super RX Fuji medical X-ray film	Fuji, Düsseldorf
Whatman 3MM paper	Whatman, Dassel

4.1.11 Standard buffers

Standard buffers are listed below, whereas additional buffers are described in the individual 'methods' sections.

LB media / agar plates:	1% (w/v)	Peptone
	0.5% (w/v)	Yeast extract
	1% (w/v)	NaCl
	for LB agar plates 2% (w/v) agar was added	

SDS-PAGE Loading Buffer (4x):	0.5 M	Tris / HCl pH 6.8
	4% (v/v)	SDS
	0.08% (w/v)	Bromphenol Blue
	40% (v/v)	Glycerol
	1.144 M	β -Mercaptoethanol
SDS-PAGE Running Buffer:	25 mM	Tris
	192 mM	Glycine
	0.1% (w/v)	SDS
PBS:	137 mM	NaCl
	2.7 mM	KCl
	10 mM	$\text{Na}_2\text{HPO}_4 \times 12 \text{ H}_2\text{O}$
	2.0 mM	KH_2PO_4
TBE:	90 mM	Tris base
	90 mM	Boric acid
	2 mM	EDTA

4.2 Mammalian and insect tissue culture

4.2.1 General cell culture conditions

Tissue culture work was performed under sterile conditions. All solutions were preheated before use either to 37°C when used for human cells or to RT when used for insect cells. To determine the cell number, cells were first mixed with the same volume of Trypan Blue to stain dead cells and then counted in a hemacytometer.

4.2.2 Cultivation of human HEK-293 cells

The human embryonic kidney cell line HEK-293 was kindly provided by M. Eckey (P. Becker lab). HEK-293 cells were kept in DMEM Medium (see section ‘chemicals for tissue culture’) supplemented with 9% (v/v) FBS in 25 cm²-flasks at 37°C and 5% CO₂. When cells reached confluency, cells were detached by carefully pipetting up and down and diluted again in 7 ml fresh medium to 0.5 x 10⁶ cells.

4.2.3 Cultivation of *Spodoptera frugiperda* Sf21 cells

Sf21 cells were cultured in Sf-900 II SFM that was supplemented with 9 mg/ml gentamycin and 9% (v/v) FBS. Cells were grown in 1L-shaker flasks in 100 - 300 ml medium at 27°C and 75 rpm. The cell density was kept between 0.5×10^6 and 4.0×10^6 cells/ml, thus cells stay in the exponential growth phase.

4.2.4 Cultivation of *Drosophila melanogaster* SL2 cells

SL2 cells were cultured at 26°C in Schneider's *Drosophila* medium (Invitrogen) containing L-glutamine and additionally supplemented with penicillin, streptomycin and 9% (v/v) FBS. Cells were kept in 75 cm²-flasks in 20 ml medium between 0.5×10^6 and 4.0×10^6 cells/ml, in order to keep cells exponentially growing.

4.2.5 Freezing and thawing of SL2 cells

The SL2 cells must be diluted to around 1.5×10^6 cells/ml in a 75 cm²-flask two days prior freezing. Then cells were centrifuged at 500 g for 4 min at RT and carefully resuspended in 3 ml Freezing Medium (60% (v/v) FBS, 10% (v/v) DMSO, 30% (v/v) medium). 0.5 ml aliquots were made, stored first for 30 min at -20°C then stored overnight at -80°C. For long-term storage, cells were kept in liquid nitrogen. A frozen stock was thawed rapidly in hand and quickly transferred in 10 ml fresh medium to dilute the toxic DMSO. Cells were centrifuged as above, resuspended in 7 ml medium and transferred to a small flask (25 cm²). Further cultivation was performed as described above.

4.2.6 Generation of stable SL2 cell lines

Stable SL2 cell lines were generated by cotransfection of the pCoBlast selection vector and the expression vector. In SL2 cells vectors are generally integrated as multicopies that form arrays of more than 500 - 1000 copies in a head to tail fashion (Invitrogen manual). 0.5×10^6 cells were seeded in a 6 well plate containing 2 ml fresh medium and incubated as described above. The next day 20 ng of pCoBlast and 400 ng of the expression vector were transfected into SL2 cells using the Effectene reagent (Qiagen). First the vectors were mixed in 100 µl EC Buffer (Qiagen), then 3.2 µl of Enhancer Solution (Qiagen) was added and vortexed for 1 sec. After incubation for 5 min at RT 10 µl Effectene Reagent was added, vortexed for 10 sec and again incubated for 10 min at RT. Finally 0.6 ml fresh medium was

mixed with the DNA-transfection mix and slowly added to 1×10^6 cells in 1.6 ml fresh medium per 6-well. After 24 h the transfection mix was removed and 3 ml blasticidin containing medium was added (25 ng/ μ l). The procedure was repeated after further 2, 5 and 10 days, when the stable cells grow out. Within the next one or two weeks resistant cells were transferred together with the conditioned and additional fresh blasticidin containing medium to small flasks (25 cm², 7 ml total volume). Before transferring to 75cm²-flasks cells were centrifuged as above to remove dead cells. Then the stable cell line was routinely kept in 75 cm²-flasks, under constant selection (20 ng/ μ l blasticidin).

4.3 Methods for manipulation of DNA

4.3.1 General molecular biology methods

Molecular biology methods that deal with the analysis and manipulation of nucleic acids were performed according to standard (Laboratory Manuals from Sambrook and Russell) and manufacturer's protocols (New England Biolabs). This includes concentration determination of nucleic acids, restriction enzyme digestion and ligation of DNA fragments, analysis of DNA on agarose gels and amplification of DNA fragments by polymerase chain reaction (PCR). In addition preparation of chemically competent and transformation of *E.coli* cells were performed as described in [189]. Plasmid DNA was prepared by using the Qiagen Midi or Maxi Kit. For isolation of DNA fragments from agarose gels and for clean-up of enzymatic reactions the QIAquick Gel Extraction Kit and the QIAquick PCR Purification Kit (both Qiagen) were used respectively.

4.3.2 Cloning of expression constructs

For expression in Sf21 cells, the different MSL2-FLAG constructs were cloned into the pFastBac1 vector by using the SalI/NotI restriction sites, either by subcloning or by introducing the SalI/NotI restriction sites via PCR with primers that contain the corresponding restriction sites, whereby the primer with the NotI restriction site contained in addition the cds for the FLAG-tag.

For cloning MSL1 into pFastBac1 vector, the MSL1-FLAG cds was first cut out from the existing pGSC-URA vector using NotI/PacI (blunt) restriction enzymes and then subcloned into the NotI/XbaI (blunt) sites.

The chimeric MSL2 construct was created by first simultaneously removing the DmCXC domain from the wildtype MSL2 cds (within the pFastBac1 and pVP16 vector) and introducing PacI/NheI restriction sites via inverse PCR. These restriction sites were then used to insert the HsCXC domain, which was amplified from cDNA of HeLa cells via PCR using primers that include PacI/NheI restriction sites.

For expression in HEK-293 cells, the different MSL constructs were subcloned into the pcDNA3.1 vector from already existing pMSL-VP16 vectors using KpnI/NotI restriction sites.

For expression in SL2 cells, all MSL constructs had been previously cloned into pVP16 plasmids (C. König, P. Becker lab), except the MSL2- Δ Pro/Bas construct, which was subcloned by using KpnI/AgeI restriction sites.

For immunofluorescence stainings on SL2 cells, all MSL constructs had been previously cloned into the pEGFP vector (T. Straub, P. Becker lab), except the MSL2 construct containing double point mutations (MSL2-C544A/C546A), which was created by subcloning a small fragment containing the point mutations using PflMI/BstEII restriction enzymes.

For bacterial expression as a GST fusion protein, the sequence of the CXC domain was amplified by PCR using primers that contain XmaI/SacI restriction sites and subsequently cloned into XmaI/SacI restriction sites of the pGEX-2KG vector.

For creating the MOF-FLAG construct, the already existing MOF-Gal4DBD construct (M. Prestel, P. Becker lab) was digested with RsrII to remove the Gal4DBD and then was religated again.

Details of the composition of the individual primers can be found above in the ‘oligonucleotides’ section. The correctness of all constructs was verified by sequencing.

4.3.3 Site-directed mutagenesis

Specific mutations were introduced into the DNA sequence by site-directed mutagenesis following the manufacturer’s instructions (QuickChange Kit, Stratagene). The designed primers carrying the changed base triplets are listed above in the ‘oligonucleotides’ section.

4.3.4 Preparation of long DNA fragments

The trimerized DBF12-L15 DNA element (167 bp) and a control DNA fragment of similar length (170 bp) was obtained by restriction digestion of pP12eL(15)3 vector with BamHI/XhoI or of pBluescript KS+ vector with PvuII/XmaI respectively. The roX1-DHS DNA fragment (226 bp) and a control DNA fragment of similar length (225 bp) was obtained by restriction digestion of pUCA-roX1DHS-short plasmid with AvaI or of pBluescript KS+ plasmid with AvaII respectively. The DNA fragments were then purified from Agarose-gels using the Qiagen Gel extraction Kit, further concentrated by EtOH precipitation and finally resolved in ddH₂O and stored at -20°C. Since restriction digestion of all plasmids created a 5'-overhang, the overhanging ends were filled-in with radioactive dCTP by the Klenow enzyme.

4.3.5 Annealing of oligonucleotides

Double stranded oligonucleotides were created by mixing equimolar concentrations (10 µM) of the sense and the anti-sense strands in Annealing Buffer (10 mM Tris/HCl pH 7.5, 50 mM NaCl, 0.1 mM EDTA). Both strands were annealed by heating for 5 min at 95°C followed by gradually cooling down to RT. Finally, the double stranded oligonucleotides were stored in small aliquots at -20°C.

4.3.6 Radioactive labeling of DNA fragments

DNA fragments were radioactively labeled by fill-in of overhanging ends using the Klenow enzyme. 5'-overhangs of DNA fragments were created either by annealing of the sense and a shorter anti-sense strand or by restriction digestion of plasmid DNA (see sections above). The 5'-overhangs were then filled-in by the Klenow fragment of DNA-Polymerase I in the presence of radioactive [α -³²P]-dCTP and non-radioactive dNTPs. 5 pmol of DNA and 5 U of the Klenow fragment were incubated for 15 min at 25°C with non-limiting amounts of [α -³²P]-dCTP (10 mCi/ml, 3000 Ci/mmol), non-radioactive dATP, dGTP, dTTP (100 µM each) in NEB Buffer 2 in a final volume of 50 µl.

In contrast to DNA, RNA oligonucleotides were labeled at their 5'-ends using T4-Polynucleotide Kinase (T4-PNK). Since the RNA oligonucleotides were synthesized without a phosphate group at their 5'-ends, it was not necessary to dephosphorylate the ends prior to the kinase reaction. 20 pmol of double stranded RNA and 20 U of T4-PNK were incubated

for 1 h at 37°C with 75 pmol, which is a non-limiting amount, of [γ - 32 P]-dATP (150 mCi/ml, 6000 Ci/mmol) in T4-PNK Buffer (New England Biolabs) in a final volume of 20 μ l.

After the labeling reactions, radiolabeled DNA and RNA were purified using the QIAquick Nucleotide Removal Kit (Qiagen) according to the manufacture's protocol. The purified radiolabeled DNA and RNA were finally eluted with 100 μ l buffer (10 mM Tris/HCl pH 8.0), measured in a scintillation counter (Beckman LS 1801) and stored in aliquots at -20°C.

4.3.7 Formation of RNA-DNA triplexes

The RNA-DNA triplexes were formed by incubating unlabeled single stranded RNA (varying amounts, see results section) and radiolabeled DNA duplex (0.1 - 0.4 nM, 6000 - 10000 cpm) in Triplex Formation Buffer (50 mM MOPS/NaOH pH 7.4, 200 nM spermidine trihydrochloride, 5% (v/v) glycerol) for 15 min at 25°C. The triplex formation was then controlled by separating duplex and triplex structures in native PAA gels (50 mM MOPS/NaOH pH 7.4, 5% (v/v) polyacrylamide solution (37.5:1), 0.05% (v/v) TEMED, 0.1% (v/v) APS, 6 mM MgCl₂) at 4°C in 6 mM MgCl₂ containing MOPS buffer (50 mM). The gels were pre-electrophoresed at 110 V for 30 min at RT and, subsequently, the triplex and duplex structures were separated at 120 V at 4°C. Gels were dried on anion exchanger chromatography and Whatman paper and radiolabeled nucleic acids were visualized by using a Phosphor-Imager (FujiFilm FLA-3000).

4.4 Methods for protein analysis

4.4.1 Precipitation of proteins with Trichloroacetic acid (TCA)

In order to obtain higher concentrated samples for SDS-PAGE protein samples were mixed with TCA (20% (v/v) final concentration) and incubated for at least 10 min on ice. After centrifugation at 13,000 rpm for 30 min at 4°C the protein precipitate was washed twice with ice-cold acetone. The protein pellet was air-dried, then dissolved in an appropriate volume of 8 M urea and 2 x SDS-PAGE Loading Buffer. Before subjected to SDS-PAGE, protein samples were boiled for 5 min at 95°C.

4.4.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Separation of proteins by SDS-PAGE was performed in the Novex Mini-Cell system (Invitrogen). Resolving and stacking gels were poured according to the manufacture's

instructions using 30% (v/v) polyacrylamide solution (Rotigel, 37.5:1 acrylamide to bisacrylamide) and resolving gel buffer (375 mM Tris/HCl pH 8.8) or stacking gel buffer (125 mM Tris/HCl pH 6.8) respectively. Prior to loading samples were mixed with SDS-PAGE Loading Buffer and heat-denatured for 5 min at 95°C. Various molecular weight protein standards were purchased from Bio-Rad. Electrophoresis was performed at 40 - 80 mA at RT for varying time. Afterwards SDS-PAA gels were either used for Coomassie staining, silver staining or Western blotting.

4.4.3 Silver staining of protein gels

All steps were performed in a clean, round glass container. The SDS-PAA gel was first briefly rinsed in ddH₂O and proteins fixed by incubation for 15 min in 50% MeOH, followed by a second incubation for 15 min in 5% MeOH. The gel was briefly rinsed in ddH₂O before gentle swirling for 15 min in Reducing Solution (32 μM DTT). It followed a two times rinsing step for 10 sec in ddH₂O and a short 10 sec preincubation followed by 15 min incubation (swirling) in Silver Stain Solution (0.1% (w/v) AgNO₃). The gel was again rinsed twice for 10 sec in ddH₂O before a short 10 sec preincubation in a small volume of Developer Solution (3% (w/v) Na₂CO₃, 0.05% (v/v) formaldehyde). Developer Solution was added again, the gel was swirled gently and the staining process was followed until the desired color intensity was reached. The reaction was stopped by adding solid Citric acid monohydrate until gas formation ceases. The gel was incubated for a few more minutes before it was swirled and stored in ddH₂O. Finally the gel was scanned.

4.4.4 Staining of protein gels by Coomassie Blue

After separation by SDS-PAGE, proteins were at the same time fixed and stained by incubating the polyacrylamide gel in Fixation-Staining Solution (40% (v/v) EtOH, 10% (v/v) acetic acid, 0.62 g (w/v) Coomassie G-250). After 1 h shaking at RT, the protein gel was transferred into Destain Solution (10% (v/v) acetic acid) and shaken at RT until protein bands became visible. The destained protein gel was either used for quantification or identification of protein bands (by mass spectrometry), before it was finally dried on a Whatman paper at 80°C for 1 h on a gel dryer (BioRad).

4.4.5 Determination of protein concentration

For quantitative experiments (e.g. measuring affinity constants with EMSAs) the concentration of the purified recombinant protein must be exactly determined. Therefore the recombinant protein was resolved from other eventually co-purified proteins by SDS-PAGE, then visualized by Coomassie staining and finally quantified by densitometry (Odyssey Infrared Imaging System, LI-COR Biosciences). The protein concentration was then determined by comparing the measured intensity of the protein band and the measured intensities of a standard curve, which was created by titration of known amounts of BSA (purchased from New England Biolabs).

4.4.6 Protein transfer to membranes (Western blot)

The transfer of proteins, which were separated by SDS-PAGE, from the polyacrylamide gel to a PVDF membrane, was achieved with a semidry transfer cell (Trans-Blot SD, Bio-Rad). Before the transfer, the polyacrylamide gel and the PVDF membrane (after its activation in MeOH and rinsing in ddH₂O) or nitrocellulose membrane were equilibrated in Cathode Buffer and Anode Buffer II respectively. The blot was then assembled the following way (from anode to cathode): 2 Whatman papers soaked in Anode Buffer I (300 mM Tris/HCl pH 7.5, 20% (v/v) MeOH), 1 Whatman paper soaked in Anode Buffer II (25 mM Tris/HCl pH 7.5, 20% (v/v) MeOH), membrane, polyacrylamide gel and finally 4 Whatman papers soaked in Cathode Buffer (70 mM CAPS pH 10.5, 10% (v/v) MeOH). The Transfer was performed at 0.8 mA/cm² membrane for either 1.5 h (7.5% polyacrylamide gel) or 2 h (10% polyacrylamide gel). Afterwards the nitrocellulose membrane was directly used for immunodetection, whereas the PVDF membrane was air-dried and stored at RT until further use for immunodetection.

4.4.7 Immunodetection of transferred proteins

The membranes were incubated in Blocking Buffer (1 x PBS, 0.05% (v/v) Tween 20, 3 - 5% (w/v) milk powder) for 1 h at RT in order to saturate free binding sites on the membrane and hence minimize the non-specific background. Since the PVDF membrane was dried after the Western blot, it has to be activated again in MeOH and rinsed in dH₂O before putting into the Blocking Buffer. After blocking the membranes were incubated either for 1 h at RT or o/n at 4°C with an appropriate dilution of the primary antibody. Before and after the incubation for an additional hour with horseradish peroxidase (HRP)- or infrared (IR) dye-

coupled secondary antibody, the membranes were extensively washed for several times in PBS-T Buffer (1 x PBS, 0.05% (v/v) Tween 20). Finally, the antigen-antibody complexes were detected by chemoluminescence using the ECL detection system (GE Healthcare) and autoradiography. After a variable exposure time the X ray films were developed and the autoradiograph was scanned. Alternatively, when secondary antibodies that are coupled to an IR dye were used proteins were detected by using the Odyssey Infrared Imaging System (LI-COR Biosciences).

4.4.8 Superose 6 gel filtration of purified MSL proteins

The stoichiometry and composition of the recombinant MSL2 protein and of the purified MSL2-MSL1 complex was analyzed by chromatographic gel filtration using the Äkta FPLC system (Amersham pharmacia). Chromatography runs were carried out at 4°C, all solutions were filtered (0.22 µm pore size) and chilled to 4°C. The Superose 6 column (20 ml CV) was washed with ddH₂O before it was equilibrated in Elution Buffer (100 mM KCl for MSL2 and 300 mM KCl for MSL2-MSL1). Purified protein solutions that were used for gel filtration did not contain NP-40, but were reduced by adding DTT (1 mM). Before loading onto the equilibrated Superose 6 column the protein solution was centrifuged at 13,000 g for 15 min at 4°C to remove protein aggregates. Around 60 µg of the protein solution in 250 µl was loaded onto the column and 500 µl fractions were collected at a flow-rate of 0.5 ml/min (pressure limit 1.5 MPa). Protein fractions were precipitated with TCA and analyzed by SDS-PAGE and Coomassie staining.

Elution Buffer: 50 mM Hepes/KOH pH 7.6
 100 mM or 300 mM KCl
 5% (v/v) Glycerol
 1 mM MgCl₂
 0.5 mM EDTA
 5 mM DTT

4.4.9 Mass spectrometry

Mass spectrometry analysis was performed by the ZfP ('Zentrum für Proteinanalytik', Adolf-Butenandt-Institute, LMU) according to the following protocol. Cut gel slices from SDS-PAGE were washed twice with water and twice with 40 mM ammoniumbicarbonate

each for 30 min. After two-times treatment with 50% acetonitrile for 5 min, trypsin (Sequencing Grade Modified, Promega) was added and proteins were digested overnight in 40 mM ammoniumbicarbonate at 37°C while shaking (600 rpm). For protein identification probes/peptides were first purified (desalted) using C18 ZipTips (Millipore) and then used for nano-ESI-LC-MS/MS. Each sample was first separated on a C18 reversed phase column (75 µm i.d. x 15 cm, packed with C18 PepMap, 3 µm, 100 Å) by LC Packings via a linear acetonitrile gradient. MS and MS/MS spectra were recorded on an Orbitrap mass spectrometer (Thermo Electron). The resulting spectra were then analyzed via the Mascot Software (Matrix Science) using the NCBI Inr Protein Database.

4.5 Expression and purification of proteins and protein complexes

4.5.1 Heterologous expression of MSL proteins in insect Sf21 cells

4.5.1.1 Infection of Sf21 cells with baculoviruses

MSL proteins were expressed in Sf21 cells using recombinant baculoviruses. Wildtype MSL2 and MSL1, as well as all truncated or mutated MSL2 versions contained C-terminal FLAG-tags. In the coexpressed MSL2-MSL1 complex only MSL2 was FLAG-tagged, MSL1 was untagged. Baculovirus infections were usually carried out in shaker flasks (1×10^6 cells/ml, 250 ml) with the optimal amount for each virus, which was determined before in test expressions. After 2 days of expression using the same conditions as described above cells were harvested, the cell pellet washed with ice-cold PBS, frozen in liquid nitrogen and stored at -80°C until preparation of cell extracts.

4.5.1.2 Preparation of Sf21 cell extracts

Sf1 cell pellets, which were stored at -80°C, were rapidly thawed in a water bath at RT and then put on ice immediately. Per 5 ml PCV 15 ml ice-cold Lysis Buffer was added and cells completely resuspended. After 10 min incubation on ice and occasional mixing, cells were lysed by sonication (4 x 10 sec pulses with 20% Amplitude = 6 Watt and 20 sec pause, Branson Digital Sonifier 250D) under constant cooling in ice water. Cell debris and insoluble proteins were removed by centrifugating the cell lysate twice – one time for 30 min and a second time for 15 min at 30,000 g and 4°C. The supernatant represents the soluble protein fraction, which is directly used for further protein purification.

Lysis Buffer: 50 mM Hepes/KOH pH 7.6
 300 mM KCl
 5% (v/v) Glycerol
 0.05% (v/v) NP-40
 1 mM MgCl₂
 0.5 mM EDTA
 Protease inhibitors (Aprotinin 1 µg/ml, Leupeptin 1 µg/ml and
 Pepstatin 0.7 µg/ml) were added just before use

4.5.1.3 Purification of FLAG-tagged MSL proteins from Sf21 cell extracts

The soluble protein fraction was incubated with in chilled Lysis Buffer equilibrated FLAG beads (Anti-FLAG M2 Agarose, Sigma) for 2.5 h at 4°C on a rotating wheel, whereby 250 µl beads (which equals 500 µl original slurry) were used per 5 ml PCV. All subsequent washing steps were carried out with ice-cold and at least 10 ml of buffer. The beads were washed briefly with Lysis Buffer, followed by a second wash step on the rotating wheel for 5 min. In order to reduce non-specific interactions with the beads and with the FLAG-tagged proteins, another washing step for 10 min on a rotating wheel with high-salt (1 M) was added. It follows a quick wash with lower salt 300 mM KCl and two quick washes in Elution Buffer. The FLAG-tagged MSL proteins were then eluted for 2.5 h at 4°C on a rotating wheel in the presence of 0.5 mg/ml FLAG-Peptide (Sigma) in a total volume of 900 µl per 250 µl beads. The Elution Buffer contained either 100 mM KCl for MSL2 or 300 mM KCl for MSL1 and the MSL2-MSL1 complex. The eluted proteins were separated from FLAG-beads by short centrifugation through Micro Bio-Spin columns (Bio Rad) and were further concentrated by centrifugation at 3000 rpm using Amicon Ultra-4 centrifugal filter devices (50 kDa or 3 kDa exclusion limit, Millipore). Purified proteins were then rapidly frozen as small aliquots (10 - 20 µl) in liquid nitrogen and finally stored at -80°C.

Elution Buffer: composition as Lysis Buffer
 but containing either 100 mM or 300 mM KCl

4.5.2 Bacterial expression of the GST-tagged CXC domain

4.5.2.1 Induction of protein expression

E.coli BL21-CodonPlus cells were transformed with the pGEX-2KG-GST-CXC vector, which allows upon induction by IPTG the expression of the isolated CXC domain fused to a GST-tag. A 200 ml preculture, which grew overnight at 22°C and 200 rpm, was used to inoculate a 2 L expression culture ($OD_{600} = 0.1$). Cells were grown at 20°C and 200 rpm under constant selection pressure in LB-Amp-Cam medium. When cell density reached $OD_{600} = 0.7 - 0.8$ expression was induced with 0.3 mM IPTG and lasted for 2 h. Cells were then harvested by centrifugation for 15 min at 4000 g at 4°C and washed once with PBS. The bacterial cell pellet was frozen in liquid nitrogen and stored at -80°C until extract preparation.

4.5.2.2 Preparation of bacterial cell extract

The bacterial cell pellet was rapidly thawed and resuspended in 30 ml ice-cold Extraction Buffer per 5 ml PCV. Cells were lysed by using the French Press (2 x 1500 psi) and by sonication (3 x 10 sec pulses with 25% Amplitude = 7 Watts and 30 sec pause, Branson Digital Sonifier 250D) under constant cooling in ice water. Cell debris and insoluble proteins were removed by centrifugating the cell lysate for 30 min at 30,000 g and 4°C. The supernatant represents the soluble protein fraction, which is directly used for further protein purification.

Extraction Buffer: 1 x PBS
 0.36 M NaCl
 0.05% (v/v) NP-40
 1 mM DTT
 50 μ M ZnCl₂
 1 mM PMSF
 Protease inhibitors (Aprotinin 1 μ g/ml, Leupeptin 1 μ g/ml, Pepstatin
 0.7 μ g/ml), DTT and PMSF were added prior to usage

4.5.2.3 Purification of GST-tagged CXC domain from bacterial cell extracts

The bacterial protein extract was allowed to enter the packed material of a custom-made affinity column consisting of around 300 µl of loosely packed Glutathione Sepharose (GE Healthcare), which was equilibrated with ice-cold Extraction Buffer (without ZnCl₂ and PMSF), by gravity flow. After passing through the column the protein extract was applied several times again. Afterwards the column was washed with Extraction Buffer, followed by a high-salt wash in 1 M salt and another wash in Extraction Buffer. The column was then equilibrated with Elution Buffer and GST-tagged protein was eluted from the column by incubating the column for 2 h at 4°C in the presence of 40 mM glutathione with 1.3 x CV of Elution Buffer. Finally the eluted protein is collected from the column by adding another 1.3 x CV of Elution Buffer containing 40 mM glutathione. Aliquots of the eluted protein were quickly frozen in liquid nitrogen and stored at -80°C.

Elution Buffer: 200 mM Tris/HCl pH 8.0
 10% (v/v) Glycerol
 150 mM NaCl
 0.05% (v/v) NP-40
 1 mM DTT
 Protease inhibitors (Aprotinin 1 µg/ml, Leupeptin 1 µg/ml, Pepstatin
 0.7 µg/ml) and DTT were added prior to usage

4.5.3 Enrichment of native MOF containing fractions from SL2 cells

4.5.3.1 Preparation of nuclear extracts from SL2 cells

Nuclear extracts were prepared according to a protocol developed by Andreas Hochheimer (B.R.A.I.N. AG, Zwingenberg, Germany), which is based on two already published protocols [190, 191]. SL2 cell culture was expanded from a 75cm²-flask to several roller bottles (maximum 400 ml) and a final concentration of 2 - 3 x 10⁶ cells/ml. Cells were harvested at 1300 g for 15 min at RT and washed 2 - 3 times with ice-cold PBS containing 12.5 mM MgCl₂. Prior to lysis cells were resuspended in 3 x PCV of hypotonic Buffer A (chilled) and left swelling for 30 min on ice with occasional mixing. Complete lysis of cells but maintaining of intact nuclei was achieved by douncing 15 - 17 times (tight B pestle) on ice. After mixing with 1/10th volume of ice-cold Buffer B the nuclei were pelleted by centrifugating at 8000 rpm (SS-34 rotor) for 15 min at 4°C. 1.5 x PCV of ice-cold Buffer AB

was used to resuspend nuclei. Nuclei were lysed by adding 1/10th volume of 4 M (NH₄)₂SO₄ and the resulting extract was rotated for 20 min at 4°C. Ultracentrifugation at 38,000 rpm (Ti 60 rotor) for 2 h at 4°C cleared the extract, which was then precipitated by slowly adding small amounts of total 0.3 g of solid (NH₄)₂SO₄ per 1 ml extract. Precipitated proteins were centrifuged at 15,000 rpm (SS-34 rotor) for 15 min at 4°C and can be stored for up to one week at 4°C. The protein pellet was resuspended in 0.5 x volumes of ice-cold Buffer C and dialyzed 2 x 2 h against Buffer C. Insoluble material was removed by centrifugation at 16,000 g for 5 min. Protease inhibitors were added again and the extract was quickly frozen as small aliquots in liquid nitrogen before long-term storage at -80°C.

Buffer A:	10 mM Hepes/KOH pH 7.6	Buffer B: 50 mM Hepes/KOH pH 7.6
	10 mM KCl	1 M KCl
	2 mM MgCl ₂	30 mM MgCl ₂
	0.1 mM EDTA	0.1 mM EDTA
	1 mM DTT	0.2 mM DTT
	0.1 mM PMSF	40 μM PMSF
	1 mM Na ₂ S ₂ O ₅	0.2 mM Na ₂ S ₂ O ₅

Protease inhibitors (Aprotinin 1 μg/ml, Leupeptin 1 μg/ml, Pepstatin 0.7 μg/ml), DTT, PMSF and Na₂S₂O₅ were added to Buffer A and B prior to use

Buffer AB: dilute Buffer B in Buffer A 1:10

Buffer C: 25 mM Hepes/KOH pH 7.6
20% (v/v) Glycerol
150 mM KCl
12.5 mM MgCl₂
0.1 mM EDTA
1 mM DTT
0.2 mM PMSF
1 mM Na₂S₂O₅
Protease inhibitors (Aprotinin 1 μg/ml, Leupeptin 1 μg/ml, Pepstatin 0.7 μg/ml), DTT, PMSF and Na₂S₂O₅ were added prior to use

4.5.3.2 Purification of FLAG-tagged MOF from SL2 nuclear extracts

The nuclear extract was incubated with in chilled Buffer C equilibrated FLAG beads (Anti-FLAG M2 Agarose, Sigma) for 2 h at 4°C on a rotating wheel, whereby 40 µl beads (which equals 80 µl original slurry) were used per 300 µl nuclear extract. All subsequent washing steps were carried out with ice-cold and at least 1 ml of buffer. The beads were two times washed briefly with Buffer C, followed by a third washing step with higher salt (400 mM KCl) on the rotating wheel for 5 min at 4°C. In order to reduce non-specific interactions with the beads and with the FLAG-tagged proteins, another three washing steps for 10 - 15 min on a rotating wheel in high-salt Wash Buffer was added. It follows a quick wash with lower salt (400 mM KCl) and three quick washes in Elution Buffer. The FLAG-tagged MOF complexes were then eluted for 2 h at 4°C on a rotating wheel in the presence of 0.5 mg/ml FLAG-Peptide (Sigma) in a total volume of 65 µl per 25 µl beads. Eluted complexes were removed from FLAG-beads by short centrifugation through Micro Bio-Spin columns (Bio Rad) and then rapidly frozen as small aliquots (5 µl) in liquid nitrogen and finally stored at -80°C.

4.6 Reporter gene assays in insect and human cells

4.6.1 Reporter gene assay in SL2 cells

4.6.1.1 Transient transfection of reporter gene constructs

Reporter gene assays in *Drosophila melanogaster* SL2 cells were performed as described in [92]. 0.25×10^6 cells were seeded the day before transfection in 12-well plates and incubated as described above. The following plasmids were transfected using the Effectene Reagent (Qiagen): 15 ng *Renilla* reporter construct, 315 ng reporter gene construct carrying the DBF12 or other binding site attached to an inducible firefly *luciferase* reporter gene, and 160 ng of an MSL2-VP16 activator construct. The plasmids were first mixed with 75 µl EC Buffer (Qiagen) then 3.9 µl of Enhancer (Qiagen) per 490 ng of total DNA transfected was added and vortexed for 1 sec. After 5 min incubation at RT 9 µl Effectene Reagent was added and vortexed for 10 sec. The DNA-Effectene complexes are formed at RT within 10 min and were carefully mixed with 400 µl fresh medium before slowly added to 0.5×10^6 cells in 800 µl fresh medium per 12-well. After 2 days incubation at 26°C SL2 cells were lysed in order to analyze reporter gene activation.

4.6.1.2 Preparation of cell extracts from transiently transfected SL2 cells

After 2 days the cells were harvested from the 6-well plate by pipetting and were pelleted by centrifugation for 10 min at 2000 rpm at 4°C. Cells were once washed with PBS before they were lysed in 100 µl Passive Lysis Buffer (Qiagen) at 4°C with occasional mixing. Finally, cell debris were removed by centrifuging at 13,000 rpm for 2 min at 4°C and the supernatant was used on the one hand for measuring luciferase activities and on the other hand for detection of transfected MSL-VP16 proteins via Western blot. After usage, the extracts were frozen in liquid nitrogen and stored at -80°C.

4.6.1.3 Measuring luciferase activities

Luciferase activities were measured by using the Dual Luciferase Kit (Promega) according to the manufacture's protocol. 20 µl of cell extracts from transiently transfected SL2 cells were incubated for 2 min at RT in vials in order to reach RT. Then 95 µl of LARII substrate was added and after another 2 min incubation at RT (but light protected) light emission of the firefly luciferase was measured using a luminometer (Berthold) and 5 ml polystyrene tubes (Sarstedt). Each measurement took 10 sec and was repeated 3 times. After mixing with 95 µl of Stop and Glow Reagent the light emission of the *Renilla* luciferase was measured as above. Firefly luciferase activity was normalized to *Renilla* luciferase activity (normalized RLU). The ratio of the normalized RLU of MSL2-VP16 relative to the normalized RLU of the VP16 activation domain alone gives the fold induction, which can be seen as a measurement for the MSL2-VP16 recruitment to the reporter gene. Fold inductions are shown as the mean value plus and minus the standard deviation from several technical replicates and from at least two independent plasmid preparations. The expression of the MSL2-VP16 constructs was checked by Western blots using a rabbit anti-MSL2 antibody.

4.6.2 Reporter gene assay in human HEK-293 cells

4.6.2.1 Transient transfection of reporter gene constructs

The day before transfection 0.1×10^6 cells were seeded in a 24-well plate in 0.5 ml medium and incubated as described above. The following plasmids were transfected: *Renilla* reporter construct (4 ng), a reporter construct carrying a binding site attached to an inducible firefly *luciferase* reporter gene (75 ng), MSL-VP16 activator constructs (varying amounts, see 'results' section), pEGFP-N2 vector (25 ng) to estimate transfection efficiency and the

pBluescript KS+ cloning vector as ‘filler-DNA’. In total 1 μg of plasmid DNA was diluted in 50 μl serum- and antibiotic-free medium and mixed with 2.5 μl Fugene HD Reagent (Fugene:DNA ratio 5:2). After 20 min incubation at RT 25 μl of the transfection mix was added slowly to the cells in a 24-well plate. After 2 days cells were harvested by pipetting and centrifuged at 400 g for 5 min at 4°C. After washing with PBS cells were lysed and luciferase activities were measured as described above for SL2 cells.

4.7 Protein-DNA interactions

4.7.1 Electrophoretic mobility shift assays (EMSA)

Purified MSL proteins were incubated with sub-saturating concentrations of radiolabeled DNA or RNA fragments (< 0.2 nM, 5000 - 10000 cpm) in 50 mM HEPES/KOH pH 7.6, 100 mM KCl, 5% glycerol, 0.05% NP-40, 0.5 mM EDTA, 1 mM MgCl_2 and 0.1 $\mu\text{g}/\mu\text{l}$ BSA (New England Biolabs) in a total volume of 12 μl . The binding reactions were started by adding the MSL protein and were analyzed after 15 min incubation at 25°C on non-denaturing 1.2% agarose gels (12 x 7 cm) in 0.5 x TBE. Alternatively, native PAA gels (0.5 x TBE, 5% (v/v) PAA (37.5:1), 0.05% (v/v) TEMED, 0.1% (v/v) APS) in the Novex Mini-Cell system (Invitrogen) were used when the small CXC protein was analyzed for DNA binding. The EMSA gels were pre-electrophoresed at 90 V for 20 min and, subsequently, the protein-nucleic acid complexes were separated at 90 V for 50 - 60 min at 18 - 20°C. Gels were dried on anion exchanger chromatography and Whatman paper and radiolabeled nucleic acids were visualized by using a Phosphor-Imager (FujiFilm FLA-3000). For competition experiments MSL proteins were first incubated at a concentration close to their K_D value (50 nM) with sub-saturating concentrations of radiolabeled DNA for 15 min at 25°C. Then unlabeled competitor DNA or RNA was added. After additional 20 min incubation at 25°C the reactions were analyzed by non-denaturing EMSA gels as described above.

4.7.2 Calculation of affinity constants

The scanned EMSA gels were analyzed with the Aida Image Analyzer Software. The signal of nucleic acid-bound protein (AB) and the signal of total nucleic acid (A_{total}) were quantified and the fraction bound (AB/A_{total}) calculated. Binding curves were obtained by performing non-linear regression with KaleidaGraph software and a standard bimolecular

model: $y = \frac{AB_{\max} \cdot x}{K_D + x}$ where y is the fraction bound, x is the concentration of protein, AB_{\max} is the maximum of nucleic acid-bound protein and K_D is the affinity constant. For competition experiments, first AB was calculated from the difference between the signal of total nucleic acid (A_{total}) and the signal of free nucleic acid (A_{free}). Then the fraction bound was calculated (AB/A_{total}) and normalized to the fraction bound measured at 0 nM competitor. Competition curves were obtained with the following competition model: $y = \frac{AB_{\max}}{(1 + (x/IC_{50})^{\text{HillSlope}})}$ where y is the normalized fraction bound, x is the concentration of competitor, AB_{\max} is the maximum of nucleic acid-bound protein at 0 nM competitor, IC_{50} is the half maximal inhibitory concentration and *Hill Slope* describes the steepness of the curve. Affinity constants and IC_{50} values are shown as the mean value plus and minus the standard deviation from several technical replicates and from at least two independent protein preparations.

4.8 Immunofluorescence staining on SL2 cells

The protocol for immunofluorescence stainings is based on the protocol published by Morales et al. [75]. Microscope slides were washed in ddH₂O and coated with Poly L-Lysin according to the manufacture's protocol (Sigma). Per well (14 mm) 0.5 - 1.0 x 10⁶ SL2 cells were seeded and allowed to settle for 2 h. The cells were then washed briefly with ice-cold PBS and fixed in 2% (v/v) formaldehyde in PBS for 7 min on ice. Afterwards cells were permeabilized in 1% (v/v) formaldehyde and 0.25% (v/v) Triton X in PBS for 7.5 min on ice and then 2 x washed in ice-cold PBS. All subsequent steps were performed at RT. Unspecific binding sites were blocked with Blocking Solution (1 x PBS, 2% (w/v) BSA (Sigma), 5% (v/v) FBS) for 1 h at RT. After blocking the microscope slides were incubated for 1 h with an appropriate dilution of the primary antibody in blocking solution. Before and after the incubation in Blocking Solution for an additional hour with secondary antibodies, the cells were washed two times with PBS. DNA was counterstained with DAPI and mounted using Moviol. Pictures were taken with a Zeiss Axiovert microscope (67x oil immersion objective) coupled to a CCD Camera (AxioCamMR, Zeiss). Images were level-adjusted in Adobe Photoshop CS4.

5 References

1. Payer, B. and J.T. Lee, *X chromosome dosage compensation: how mammals keep the balance*. *Annu Rev Genet*, 2008. **42**: p. 733-72.
2. Lucchesi, J.C., W.G. Kelly, and B. Panning, *Chromatin remodeling in dosage compensation*. *Annu Rev Genet*, 2005. **39**: p. 615-51.
3. Straub, T. and P.B. Becker, *Dosage compensation: the beginning and end of generalization*. *Nat Rev Genet*, 2007. **8**(1): p. 47-57.
4. Gupta, V., et al., *Global analysis of X-chromosome dosage compensation*. *J Biol*, 2006. **5**(1): p. 3.
5. Lin, H., et al., *Dosage compensation in the mouse balances up-regulation and silencing of X-linked genes*. *PLoS Biol*, 2007. **5**(12): p. e326.
6. Nguyen, D.K. and C.M. Disteche, *Dosage compensation of the active X chromosome in mammals*. *Nat Genet*, 2006. **38**(1): p. 47-53.
7. Allis, C.D., T. Jenuwein, and D. Reinberg, *Epigenetics*, ed. C.S. Harbor. 2007, New York: Cold Spring Harbor Laboratory Press.
8. Bonisch, C., et al., *Chromatin proteomics and epigenetic regulatory circuits*. *Expert Rev Proteomics*, 2008. **5**(1): p. 105-19.
9. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications*. *Nature*, 2000. **403**(6765): p. 41-5.
10. Sarma, K. and D. Reinberg, *Histone variants meet their match*. *Nat Rev Mol Cell Biol*, 2005. **6**(2): p. 139-49.
11. Becker, P.B. and W. Horz, *ATP-dependent nucleosome remodeling*. *Annu Rev Biochem*, 2002. **71**: p. 247-73.
12. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes*. *Annu Rev Biochem*, 2009. **78**: p. 273-304.
13. Hirano, T., *The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair*. *Genes Dev*, 2002. **16**(4): p. 399-414.
14. Hagstrom, K.A., et al., *C. elegans condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis*. *Genes Dev*, 2002. **16**(6): p. 729-42.
15. Lieb, J.D., et al., *MIX-1: an essential component of the C. elegans mitotic machinery executes X chromosome dosage compensation*. *Cell*, 1998. **92**(2): p. 265-77.
16. Chuang, P.T., J.D. Lieb, and B.J. Meyer, *Sex-specific assembly of a dosage compensation complex on the nematode X chromosome*. *Science*, 1996. **274**(5293): p. 1736-9.
17. Davis, T.L. and B.J. Meyer, *SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome*. *Development*, 1997. **124**(5): p. 1019-31.
18. Dawes, H.E., et al., *Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate*. *Science*, 1999. **284**(5421): p. 1800-4.
19. Csankovszki, G., P. McDonel, and B.J. Meyer, *Recruitment and spreading of the C. elegans dosage compensation complex along X chromosomes*. *Science*, 2004. **303**(5661): p. 1182-5.
20. McDonel, P., et al., *Clustered DNA motifs mark X chromosomes for repression by a dosage compensation complex*. *Nature*, 2006. **444**(7119): p. 614-8.
21. Ercan, S., et al., *X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation*. *Nat Genet*, 2007. **39**(3): p. 403-8.

22. Ercan, S. and J.D. Lieb, *C. elegans* dosage compensation: a window into mechanisms of domain-scale gene regulation. *Chromosome Res*, 2009. **17**(2): p. 215-27.
23. Jans, J., et al., A condensin-like dosage compensation complex acts at a distance to control expression throughout the genome. *Genes Dev*, 2009. **23**(5): p. 602-18.
24. Meyer, B.J., *X-Chromosome dosage compensation*. *WormBook*, 2005: p. 1-14.
25. Trent, C., et al., Sex-specific transcriptional regulation of the *C. elegans* sex-determining gene *her-1*. *Mech Dev*, 1991. **34**(1): p. 43-55.
26. Parvin, J.D. and P.A. Sharp, DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*, 1993. **73**(3): p. 533-40.
27. Kimura, K. and T. Hirano, ATP-dependent positive supercoiling of DNA by 13S condensin: a biochemical implication for chromosome condensation. *Cell*, 1997. **90**(4): p. 625-34.
28. Chow, J. and E. Heard, X inactivation and the complexities of silencing a sex chromosome. *Curr Opin Cell Biol*, 2009. **21**(3): p. 359-66.
29. Clemson, C.M., et al., XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol*, 1996. **132**(3): p. 259-75.
30. Lee, J.T., et al., A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell*, 1996. **86**(1): p. 83-94.
31. Penny, G.D., et al., Requirement for Xist in X chromosome inactivation. *Nature*, 1996. **379**(6561): p. 131-7.
32. Lee, J.T., L.S. Davidow, and D. Warshawsky, Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*, 1999. **21**(4): p. 400-4.
33. Sado, T., Y. Hoki, and H. Sasaki, Tsix silences Xist through modification of chromatin structure. *Dev Cell*, 2005. **9**(1): p. 159-65.
34. Navarro, P., et al., Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation. *Genes Dev*, 2005. **19**(12): p. 1474-84.
35. Wutz, A., T.P. Rasmussen, and R. Jaenisch, Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet*, 2002. **30**(2): p. 167-74.
36. Chow, J.C., et al., Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proc Natl Acad Sci U S A*, 2007. **104**(24): p. 10104-9.
37. Kalantry, S. and T. Magnuson, The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet*, 2006. **2**(5): p. e66.
38. Leeb, M. and A. Wutz, Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol*, 2007. **178**(2): p. 219-29.
39. Ng, K., et al., Xist and the order of silencing. *EMBO Rep*, 2007. **8**(1): p. 34-9.
40. Costanzi, C. and J.R. Pehrson, Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, 1998. **393**(6685): p. 599-601.
41. Chadwick, B.P. and H.F. Willard, A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol*, 2001. **152**(2): p. 375-84.
42. Sado, T., et al., X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev Biol*, 2000. **225**(2): p. 294-303.
43. Cai, S., C.C. Lee, and T. Kohwi-Shigematsu, SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*, 2006. **38**(11): p. 1278-88.
44. Agrelo, R., et al., SATB1 defines the developmental context for gene silencing by Xist in lymphoma and embryonic cells. *Dev Cell*, 2009. **16**(4): p. 507-16.

45. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females*. *Nature*, 2005. **434**(7031): p. 400-4.
46. Johnston, C.M., et al., *Large-scale population study of human cell lines indicates that dosage compensation is virtually complete*. *PLoS Genet*, 2008. **4**(1): p. e9.
47. Lyon, M.F., *Do LINEs Have a Role in X-Chromosome Inactivation?* *J Biomed Biotechnol*, 2006. **2006**(1): p. 59746.
48. Bailey, J.A., et al., *Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis*. *Proc Natl Acad Sci U S A*, 2000. **97**(12): p. 6634-9.
49. Carrel, L., et al., *Genomic environment predicts expression patterns on the human inactive X chromosome*. *PLoS Genet*, 2006. **2**(9): p. e151.
50. Avner, P. and E. Heard, *X-chromosome inactivation: counting, choice and initiation*. *Nat Rev Genet*, 2001. **2**(1): p. 59-67.
51. Reenan, R.A., C.J. Hanrahan, and B. Ganetzky, *The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na⁺ channel transcript in a region of RNA editing*. *Neuron*, 2000. **25**(1): p. 139-49.
52. Gelbart, M.E. and M.I. Kuroda, *Drosophila dosage compensation: a complex voyage to the X chromosome*. *Development*, 2009. **136**(9): p. 1399-410.
53. Wang, Y., et al., *The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in Drosophila*. *Cell*, 2001. **105**(4): p. 433-43.
54. Zhang, W., et al., *Genetic and phenotypic analysis of alleles of the Drosophila chromosomal JIL-1 kinase reveals a functional requirement at multiple developmental stages*. *Genetics*, 2003. **165**(3): p. 1341-54.
55. Furuhashi, H., M. Nakajima, and S. Hirose, *DNA supercoiling factor contributes to dosage compensation in Drosophila*. *Development*, 2006. **133**(22): p. 4475-83.
56. Bai, X., A.A. Alekseyenko, and M.I. Kuroda, *Sequence-specific targeting of MSL complex regulates transcription of the roX RNA genes*. *EMBO J*, 2004. **23**(14): p. 2853-61.
57. Meller, V.H. and B.P. Rattner, *The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex*. *EMBO J*, 2002. **21**(5): p. 1084-91.
58. Bashaw, G.J. and B.S. Baker, *The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control*. *Cell*, 1997. **89**(5): p. 789-98.
59. Kelley, R.L., et al., *Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila*. *Cell*, 1995. **81**(6): p. 867-77.
60. Smith, E.R., et al., *The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation*. *Mol Cell Biol*, 2000. **20**(1): p. 312-8.
61. Birchler, J.A., M. Pal-Bhadra, and U. Bhadra, *Dosage dependent gene regulation and the compensation of the X chromosome in Drosophila males*. *Genetica*, 2003. **117**(2-3): p. 179-90.
62. Hilfiker, A., et al., *mof, a putative acetyl transferase gene related to the Tip60 and MOZ human genes and to the SAS genes of yeast, is required for dosage compensation in Drosophila*. *EMBO J*, 1997. **16**(8): p. 2054-60.
63. Mendjan, S., et al., *Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila*. *Mol Cell*, 2006. **21**(6): p. 811-23.
64. Straub, T., I.K. Dahlsveen, and P.B. Becker, *Dosage compensation in flies: mechanism, models, mystery*. *FEBS Lett*, 2005. **579**(15): p. 3258-63.

65. Hamada, F.N., et al., *Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster*. *Genes Dev*, 2005. **19**(19): p. 2289-94.
66. Straub, T., et al., *The Drosophila MSL complex activates the transcription of target genes*. *Genes Dev*, 2005. **19**(19): p. 2284-8.
67. Robinson, P.J., et al., *30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction*. *J Mol Biol*, 2008. **381**(4): p. 816-25.
68. Corona, D.F., et al., *ISWI regulates higher-order chromatin structure and histone H1 assembly in vivo*. *PLoS Biol*, 2007. **5**(9): p. e232.
69. Siriaco, G., et al., *Drosophila ISWI regulates the association of histone H1 with interphase chromosomes in vivo*. *Genetics*, 2009. **182**(3): p. 661-9.
70. Corona, D.F., et al., *Modulation of ISWI function by site-specific histone acetylation*. *EMBO Rep*, 2002. **3**(3): p. 242-7.
71. Alekseyenko, A.A., et al., *High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome*. *Genes Dev*, 2006. **20**(7): p. 848-57.
72. Gilfillan, G.D., et al., *Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex*. *Genes Dev*, 2006. **20**(7): p. 858-70.
73. Smith, E.R., C.D. Allis, and J.C. Lucchesi, *Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in Drosophila males*. *J Biol Chem*, 2001. **276**(34): p. 31483-6.
74. Akhtar, A. and P.B. Becker, *Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila*. *Mol Cell*, 2000. **5**(2): p. 367-75.
75. Morales, V., et al., *Functional integration of the histone acetyltransferase MOF into the dosage compensation complex*. *EMBO J*, 2004. **23**(11): p. 2258-68.
76. Kind, J., et al., *Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila*. *Cell*, 2008. **133**(5): p. 813-28.
77. Bell, O., et al., *Transcription-coupled methylation of histone H3 at lysine 36 regulates dosage compensation by enhancing recruitment of the MSL complex in Drosophila melanogaster*. *Mol Cell Biol*, 2008. **28**(10): p. 3401-9.
78. Larschan, E., et al., *MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism*. *Mol Cell*, 2007. **28**(1): p. 121-33.
79. Copps, K., et al., *Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly*. *EMBO J*, 1998. **17**(18): p. 5409-17.
80. Lyman, L.M., et al., *Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association*. *Genetics*, 1997. **147**(4): p. 1743-53.
81. Meller, V.H., et al., *Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila*. *Curr Biol*, 2000. **10**(3): p. 136-43.
82. Dahlsveen, I.K., et al., *Targeting determinants of dosage compensation in Drosophila*. *PLoS Genet*, 2006. **2**(2): p. e5.
83. Kageyama, Y., et al., *Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site*. *EMBO J*, 2001. **20**(9): p. 2236-45.
84. Park, Y., et al., *Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex*. *Mol Cell*, 2003. **11**(4): p. 977-86.
85. Fagegaltier, D. and B.S. Baker, *X chromosome sites autonomously recruit the dosage compensation complex in Drosophila males*. *PLoS Biol*, 2004. **2**(11): p. e341.

86. Oh, H., J.R. Bone, and M.I. Kuroda, *Multiple classes of MSL binding sites target dosage compensation to the X chromosome of Drosophila*. *Curr Biol*, 2004. **14**(6): p. 481-7.
87. Demakova, O.V., et al., *The MSL complex levels are critical for its correct targeting to the chromosomes in Drosophila melanogaster*. *Chromosoma*, 2003. **112**(3): p. 103-15.
88. Bhadra, U., M. Pal-Bhadra, and J.A. Birchler, *Role of the male specific lethal (msl) genes in modifying the effects of sex chromosomal dosage in Drosophila*. *Genetics*, 1999. **152**(1): p. 249-68.
89. Oh, H., Y. Park, and M.I. Kuroda, *Local spreading of MSL complexes from roX genes on the Drosophila X chromosome*. *Genes Dev*, 2003. **17**(11): p. 1334-9.
90. Park, Y., et al., *Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins*. *Science*, 2002. **298**(5598): p. 1620-3.
91. Kind, J. and A. Akhtar, *Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes*. *Genes Dev*, 2007. **21**(16): p. 2030-40.
92. Gilfillan, G.D., et al., *Cumulative contributions of weak DNA determinants to targeting the Drosophila dosage compensation complex*. *Nucleic Acids Res*, 2007. **35**(11): p. 3561-72.
93. Alekseyenko, A.A., et al., *A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome*. *Cell*, 2008. **134**(4): p. 599-609.
94. Straub, T., et al., *The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex*. *PLoS Genet*, 2008. **4**(12): p. e1000302.
95. Rattner, B.P. and V.H. Meller, *Drosophila male-specific lethal 2 protein controls sex-specific expression of the roX genes*. *Genetics*, 2004. **166**(4): p. 1825-32.
96. Grimaud, C. and P.B. Becker, *The dosage compensation complex shapes the conformation of the X chromosome in Drosophila*. *Genes Dev*, 2009. **23**(21): p. 2490-5.
97. Scott, M.J., et al., *MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila*. *EMBO J*, 2000. **19**(1): p. 144-55.
98. Li, F., D.A. Parry, and M.J. Scott, *The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively*. *Mol Cell Biol*, 2005. **25**(20): p. 8913-24.
99. Li, F., A.H. Schiemann, and M.J. Scott, *Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex*. *Mol Cell Biol*, 2008. **28**(4): p. 1252-64.
100. Straub, T., et al., *Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment*. *Chromosoma*, 2005. **114**(5): p. 352-64.
101. Murphy, F.V.t. and M.E. Churchill, *Nonsequence-specific DNA recognition: a structural perspective*. *Structure*, 2000. **8**(4): p. R83-9.
102. Garvie, C.W. and C. Wolberger, *Recognition of specific DNA sequences*. *Mol Cell*, 2001. **8**(5): p. 937-46.
103. Pabo, C.O. and R.T. Sauer, *Transcription factors: structural families and principles of DNA recognition*. *Annu Rev Biochem*, 1992. **61**: p. 1053-95.
104. Ohlendorf, D.H., et al., *Comparison of the structures of cro and lambda repressor proteins from bacteriophage lambda*. *J Mol Biol*, 1983. **169**(3): p. 757-69.
105. Steitz, T.A., et al., *Structural similarity in the DNA-binding domains of catabolite gene activator and cro repressor proteins*. *Proc Natl Acad Sci U S A*, 1982. **79**(10): p. 3097-100.
106. Jordan, S.R. and C.O. Pabo, *Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions*. *Science*, 1988. **242**(4880): p. 893-9.

107. Hope, I.A. and K. Struhl, *Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast*. Cell, 1986. **46**(6): p. 885-94.
108. Landschulz, W.H., P.F. Johnson, and S.L. McKnight, *The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins*. Science, 1988. **240**(4860): p. 1759-64.
109. Murre, C., P.S. McCaw, and D. Baltimore, *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins*. Cell, 1989. **56**(5): p. 777-83.
110. Hai, T. and T. Curran, *Cross-family dimerization of transcription factors Fos/Jun and ATF/CREB alters DNA binding specificity*. Proc Natl Acad Sci U S A, 1991. **88**(9): p. 3720-4.
111. Agre, P., P.F. Johnson, and S.L. McKnight, *Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP*. Science, 1989. **246**(4932): p. 922-6.
112. Schumacher, M.A., et al., *Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices*. Science, 1994. **266**(5186): p. 763-70.
113. Kim, J.L., D.B. Nikolov, and S.K. Burley, *Co-crystal structure of TBP recognizing the minor groove of a TATA element*. Nature, 1993. **365**(6446): p. 520-7.
114. Kim, Y., et al., *Crystal structure of a yeast TBP/TATA-box complex*. Nature, 1993. **365**(6446): p. 512-20.
115. Pabo, C.O. and L. Nekludova, *Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition?* J Mol Biol, 2000. **301**(3): p. 597-624.
116. Luscombe, N.M., R.A. Laskowski, and J.M. Thornton, *Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level*. Nucleic Acids Res, 2001. **29**(13): p. 2860-74.
117. Murphy, F.V.t., R.M. Sweet, and M.E. Churchill, *The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition*. EMBO J, 1999. **18**(23): p. 6610-8.
118. Robinson, H., et al., *The hyperthermophile chromosomal protein Sac7d sharply kinks DNA*. Nature, 1998. **392**(6672): p. 202-5.
119. Krishna, S.S., I. Majumdar, and N.V. Grishin, *Structural classification of zinc fingers: survey and summary*. Nucleic Acids Res, 2003. **31**(2): p. 532-50.
120. Rubin, G.M., et al., *Comparative genomics of the eukaryotes*. Science, 2000. **287**(5461): p. 2204-15.
121. Brown, R.S., C. Sander, and P. Argos, *The primary structure of transcription factor TFIIIA has 12 consecutive repeats*. FEBS Lett, 1985. **186**(2): p. 271-4.
122. Miller, J., A.D. McLachlan, and A. Klug, *Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes*. EMBO J, 1985. **4**(6): p. 1609-14.
123. Pavletich, N.P. and C.O. Pabo, *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å*. Science, 1991. **252**(5007): p. 809-17.
124. Nolte, R.T., et al., *Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex*. Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2938-43.
125. Bain, D.L., et al., *Nuclear receptor structure: implications for function*. Annu Rev Physiol, 2007. **69**: p. 201-20.
126. Khorasanizadeh, S. and F. Rastinejad, *Nuclear-receptor interactions on DNA-response elements*. Trends Biochem Sci, 2001. **26**(6): p. 384-90.

127. Gamsjaeger, R., et al., *Sticky fingers: zinc-fingers as protein-recognition motifs*. Trends Biochem Sci, 2007. **32**(2): p. 63-70.
128. Mackay, J.P. and M. Crossley, *Zinc fingers are sticking together*. Trends Biochem Sci, 1998. **23**(1): p. 1-4.
129. Brown, R.S., *Zinc finger proteins: getting a grip on RNA*. Curr Opin Struct Biol, 2005. **15**(1): p. 94-8.
130. Hall, T.M., *Multiple modes of RNA recognition by zinc finger proteins*. Curr Opin Struct Biol, 2005. **15**(3): p. 367-73.
131. Pelham, H.R. and D.D. Brown, *A specific transcription factor that can bind either the 5S RNA gene or 5S RNA*. Proc Natl Acad Sci U S A, 1980. **77**(7): p. 4170-4.
132. Picard, B. and M. Wegnez, *Isolation of a 7S particle from Xenopus laevis oocytes: a 5S RNA-protein complex*. Proc Natl Acad Sci U S A, 1979. **76**(1): p. 241-5.
133. Pieler, T., J. Hamm, and R.G. Roeder, *The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains with variable spacing*. Cell, 1987. **48**(1): p. 91-100.
134. Lu, D., M.A. Searles, and A. Klug, *Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition*. Nature, 2003. **426**(6962): p. 96-100.
135. Kelley, C.M., et al., *Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors*. Curr Biol, 1998. **8**(9): p. 508-15.
136. McCarty, A.S., et al., *Selective dimerization of a C2H2 zinc finger subfamily*. Mol Cell, 2003. **11**(2): p. 459-70.
137. Liew, C.K., et al., *Zinc fingers as protein recognition motifs: structural basis for the GATA-1/friend of GATA interaction*. Proc Natl Acad Sci U S A, 2005. **102**(3): p. 583-8.
138. Kadrmas, J.L. and M.C. Beckerle, *The LIM domain: from the cytoskeleton to the nucleus*. Nat Rev Mol Cell Biol, 2004. **5**(11): p. 920-31.
139. Deane, J.E., et al., *Tandem LIM domains provide synergistic binding in the LMO4:Ldb1 complex*. EMBO J, 2004. **23**(18): p. 3589-98.
140. Deshaies, R.J. and C.A. Joazeiro, *RING domain E3 ubiquitin ligases*. Annu Rev Biochem, 2009. **78**: p. 399-434.
141. Kruse, J.P. and W. Gu, *MSL2 promotes Mdm2-independent cytoplasmic localization of p53*. J Biol Chem, 2009. **284**(5): p. 3250-63.
142. Bienz, M., *The PHD finger, a nuclear protein-interaction domain*. Trends Biochem Sci, 2006. **31**(1): p. 35-40.
143. Li, H., et al., *Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF*. Nature, 2006. **442**(7098): p. 91-5.
144. Pena, P.V., et al., *Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2*. Nature, 2006. **442**(7098): p. 100-3.
145. Shi, X., et al., *ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression*. Nature, 2006. **442**(7098): p. 96-9.
146. Marin, I., *Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes*. J Mol Evol, 2003. **56**(5): p. 527-39.
147. Sanchez-Elsner, T., et al., *Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax*. Science, 2006. **311**(5764): p. 1118-23.
148. Duca, M., et al., *The triple helix: 50 years later, the outcome*. Nucleic Acids Res, 2008. **36**(16): p. 5123-38.
149. Soukup, G.A., A.D. Ellington, and L.J. Maher, 3rd, *Selection of RNAs that bind to duplex DNA at neutral pH*. J Mol Biol, 1996. **259**(2): p. 216-28.
150. Campos, I., et al., *Genetic Screen in Drosophila melanogaster Uncovers a Novel Set of Genes Required for Embryonic Epithelial Repair*. Genetics, 2009.

151. Lebreton, A. and B. Seraphin, *Exosome-mediated quality control: substrate recruitment and molecular activity*. *Biochim Biophys Acta*, 2008. **1779**(9): p. 558-65.
152. Daugaard, M., M. Rohde, and M. Jaattela, *The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions*. *FEBS Lett*, 2007. **581**(19): p. 3702-10.
153. Dutcher, S.K., *The tubulin fraternity: alpha to eta*. *Curr Opin Cell Biol*, 2001. **13**(1): p. 49-54.
154. Pestova, T.V., et al., *Molecular mechanisms of translation initiation in eukaryotes*. *Proc Natl Acad Sci U S A*, 2001. **98**(13): p. 7029-36.
155. Doherty, G.J. and H.T. McMahon, *Mediation, modulation, and consequences of membrane-cytoskeleton interactions*. *Annu Rev Biophys*, 2008. **37**: p. 65-95.
156. Vashee, S., et al., *How do "Zn2 cys6" proteins distinguish between similar upstream activation sites? Comparison of the DNA-binding specificity of the GAL4 protein in vitro and in vivo*. *J Biol Chem*, 1993. **268**(33): p. 24699-706.
157. Lehmann, M., *Anything else but GAGA: a nonhistone protein complex reshapes chromatin structure*. *Trends Genet*, 2004. **20**(1): p. 15-22.
158. Adkins, N.L., T.A. Hagerman, and P. Georgel, *GAGA protein: a multi-faceted transcription factor*. *Biochem Cell Biol*, 2006. **84**(4): p. 559-67.
159. Jimenez-Garcia, E., et al., *The GAGA factor of Drosophila binds triple-stranded DNA*. *J Biol Chem*, 1998. **273**(38): p. 24640-8.
160. Katsani, K.R., M.A. Hajibagheri, and C.P. Verrijzer, *Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology*. *EMBO J*, 1999. **18**(3): p. 698-708.
161. Mahmoudi, T., K.R. Katsani, and C.P. Verrijzer, *GAGA can mediate enhancer function in trans by linking two separate DNA molecules*. *EMBO J*, 2002. **21**(7): p. 1775-81.
162. Carrington, E.A. and R.S. Jones, *The Drosophila Enhancer of zeste gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution*. *Development*, 1996. **122**(12): p. 4073-83.
163. Hauser, B.A., et al., *TSO1 is a novel protein that modulates cytokinesis and cell expansion in Arabidopsis*. *Development*, 2000. **127**(10): p. 2219-26.
164. Song, J.Y., et al., *Regulation of meristem organization and cell division by TSO1, an Arabidopsis gene with cysteine-rich repeats*. *Development*, 2000. **127**(10): p. 2207-17.
165. Andersen, S.U., et al., *The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc in vitro and TSO1 is required for both male and female fertility in Arabidopsis thaliana*. *J Exp Bot*, 2007. **58**(13): p. 3657-70.
166. Beall, E.L., et al., *Discovery of tMAC: a Drosophila testis-specific meiotic arrest complex paralogous to Myb-Muv B*. *Genes Dev*, 2007. **21**(8): p. 904-19.
167. Jiang, J., et al., *Tombola, a tesmin/TSO1-family protein, regulates transcriptional activation in the Drosophila male germline and physically interacts with always early*. *Development*, 2007. **134**(8): p. 1549-59.
168. Beall, E.L., et al., *Role for a Drosophila Myb-containing protein complex in site-specific DNA replication*. *Nature*, 2002. **420**(6917): p. 833-7.
169. Korenjak, M., et al., *Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes*. *Cell*, 2004. **119**(2): p. 181-93.
170. Lewis, P.W., et al., *Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex*. *Genes Dev*, 2004. **18**(23): p. 2929-40.
171. Cvitanich, C., et al., *CPPI, a DNA-binding protein involved in the expression of a soybean leghemoglobin c3 gene*. *Proc Natl Acad Sci U S A*, 2000. **97**(14): p. 8163-8.

172. Nagase, T., et al., *Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro*. DNA Res, 2000. **7**(4): p. 273-81.
173. Buscaino, A., et al., *MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex*. Mol Cell, 2003. **11**(5): p. 1265-77.
174. Richter, L., J.R. Bone, and M.I. Kuroda, *RNA-dependent association of the Drosophila maleless protein with the male X chromosome*. Genes Cells, 1996. **1**(3): p. 325-36.
175. Schoenmakers, E., et al., *Differential DNA binding by the androgen and glucocorticoid receptors involves the second Zn-finger and a C-terminal extension of the DNA-binding domains*. Biochem J, 1999. **341** (Pt 3): p. 515-21.
176. Manzini, G., et al., *Triple helix formation by oligopurine-oligopyrimidine DNA fragments. Electrophoretic and thermodynamic behavior*. J Mol Biol, 1990. **213**(4): p. 833-43.
177. Paris, C., et al., *Mechanism of copper mediated triple helix formation at neutral pH in Drosophila satellite repeats*. Biophys J, 2007. **92**(7): p. 2498-506.
178. Shefer, K., et al., *A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA*. Mol Cell Biol, 2007. **27**(6): p. 2130-43.
179. van Steensel, B., J. Delrow, and H.J. Bussemaker, *Genomewide analysis of Drosophila GAGA factor target genes reveals context-dependent DNA binding*. Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2580-5.
180. Wilkins, R.C. and J.T. Lis, *Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation*. Nucleic Acids Res, 1997. **25**(20): p. 3963-8.
181. Omichinski, J.G., et al., *The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode*. Nat Struct Biol, 1997. **4**(2): p. 122-32.
182. Pedone, P.V., et al., *The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding*. Proc Natl Acad Sci U S A, 1996. **93**(7): p. 2822-6.
183. Wilkins, R.C. and J.T. Lis, *GAGA factor binding to DNA via a single trinucleotide sequence element*. Nucleic Acids Res, 1998. **26**(11): p. 2672-8.
184. Greenberg, A.J., J.L. Yanowitz, and P. Schedl, *The Drosophila GAGA factor is required for dosage compensation in males and for the formation of the male-specific-lethal complex chromatin entry site at 12DE*. Genetics, 2004. **166**(1): p. 279-89.
185. Lehmann, M., et al., *The pipsqueak protein of Drosophila melanogaster binds to GAGA sequences through a novel DNA-binding domain*. J Biol Chem, 1998. **273**(43): p. 28504-9.
186. Schwendemann, A. and M. Lehmann, *Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci*. Proc Natl Acad Sci U S A, 2002. **99**(20): p. 12883-8.
187. Leung, T.H., A. Hoffmann, and D. Baltimore, *One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers*. Cell, 2004. **118**(4): p. 453-64.
188. Meijnsing, S.H., et al., *DNA binding site sequence directs glucocorticoid receptor structure and activity*. Science, 2009. **324**(5925): p. 407-10.
189. Inoue, H., H. Nojima, and H. Okayama, *High efficiency transformation of Escherichia coli with plasmids*. Gene, 1990. **96**(1): p. 23-8.
190. Dignam, J.D., R.M. Lebovitz, and R.G. Roeder, *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei*. Nucleic Acids Res, 1983. **11**(5): p. 1475-89.
191. Heberlein, U. and R. Tjian, *Temporal pattern of alcohol dehydrogenase gene transcription reproduced by Drosophila stage-specific embryonic extracts*. Nature, 1988. **331**(6155): p. 410-5.

6 Appendix

6.1 Abbreviations

Å	Angstrom
A	Adenine
A ₂₈₀	Absorption at 280 nm
ac	Acetylation
Amp	Ampicillin
APS	Ammonium persulfate
ATP	Adenosin triphosphate
bp	Basepairs
BSA	Bovine serum albumin
C	Cytosine
°C	Degree Celsius
Cam	Chloramphenicol
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
cDNA	Complementary DNA
cds	Coding sequence
CES	Chromatin entry sites
ChIP	Chromatin immunoprecipitation
Ci	Curie
cpm	Counts per minute
CV	Column volume
Da	Dalton
dATP	Desoxyadenosin triphosphate
dCTP	Desoxycytosin triphosphate
ddH ₂ O	Double distilled water
dGTP	Desoxyguanosin triphosphate
dH ₂ O	Distilled water
DAPI	4',6-diamidino-2-phenylindole
DBF	DCC binding fragment
DBD	DNA binding domain
DCC	Dosage compensation complex
DHS	DNase hypersensitive site
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
dNTP	Desoxyribonucleotide triphosphate
DTT	Dithiothreitol
dTTP	Desoxythymidin triphosphate
ds	Double stranded
EDTA	Ethylendiamintetraacetate
EMSA	Electrophoretic mobility shift assay
EtOH	Ethanol
FBS	Fetal bovine serum
FPLC	Fast protein liquid chromatography
FRAP	Fluorescence recovery after photobleaching
fw	Forward
g	Gravities
G	Guanine
GFP	Green fluorescent protein
GST	Glutathione-S-transferase
h	Hours
HAS	High affinity site
HAT	Histone acetyltransferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
HTH	Helix-turn-helix

IC_{50}	Half maximal inhibitory concentration
IF	Immunofluorescence
IPTG	1-isopropyl- β -D-1-thiogalacto-pyranoside
IR	Infrared
ISWI	Imitation switch
K	Lysine
kbp	Kilobasepairs
K_D	Affinity constant
kDa	Kilodalton
LB	Lauria-Bertani
M	Molar
mA	Milliampere
mAU	Milli absorbance units
me	Methylation
MeOH	Methanol
mcs	Multiple cloning site
min	Minutes
MOPS	3-(N-morpholino)propanesulfonic acid
MPa	Megapascals
MRE	MSL recognition element
MSL	Male-specific lethal
NP-40	Nonidet P-40
nt	Nucleotides
o/n	Overnight
OD_{600}	Optical density at 600 nm
^{32}P	Radioactive phosphorus isotope ^{32}P
PAA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCV	Packed cell volume
PMSF	Phenylmethanesulfonyl fluoride
psi	Pounds per square inch
PVDF	Polyvinylidene fluoride
RLU	Relative luciferase units
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions per minute
rRNA	ribosomal RNA
RT	Room temperature
rv	Reverse
SDS	Sodium dodecyl sulfate
sec	Seconds
ss	Single stranded
T	Thymine
TBE	Tris borate EDTA buffer
TCA	Trichloroacetic acid
TEMED	N,N,N',N'-Tetramethylethylenediamine
TFO	Triplex forming oligonucleotide
Tris	Tris(hydroxymethyl)aminomethane
U	Uracil
UAS	Upstream activator sequence
ub	Ubiquitylation
V	Volts
VP16-AD	Herpes simplex viron protein 16 activation domain
v/v	Volume per volume
WB	Western Blot
wt	Wildtype
w/v	Weight per volume
Xa	active X chromosome
Xi	inactive X chromosome

6.2 Acknowledgments

Ich möchte mich zuallererst bei meinem Doktorvater Prof. Dr. Peter B. Becker für die Vergabe des Themas, für sein immer offenes Ohr trotz vollem Terminkalender, sowohl für wissenschaftliche Diskussionen, als auch für die kleinen Sorgen, und vor allem für sein Vertrauen und all die mir gegebenen Freiheiten bedanken.

Bedanken möchte ich mich auch bei Prof. Dr. Dirk Eick für das Schreiben des Zweitgutachtens.

Ebenfalls danke ich meinem 'Thesis Advisory Committee' – Philipp Korber und Tobias Straub – für ihre Vorschläge und ihr Engagement.

Allen Kollegen aus dem Institut danke ich für ihre Hilfsbereitschaft und Anregungen, den großartigen Zusammenhalt und die einzigartigen Grill-, und Fußballfeste auf Großleinwand.

Eine ganze Reihe von Leuten haben das Leben im und außerhalb des Labors zu etwas ganz Besonderem gemacht: All die fleißigen GAP-gänger/innen, das legendäre 'Ibero-Bavarian consorcio' und die nachfolgenden Braumeister, die glorreichen Theresienwiese-Kicker, die Biergarten-Genießer und natürlich auch die unerschrockenen Mensa-Pilger (Ana, Verena, Clemens, Christian, Henrike, Caro, Theresa und Ignasi) und alle, die am 19. 5. 2007 dabei waren (Jochen)!

Ganz besonderer Dank geht zum einen an die 'Unholy Alliance' – Michi, Ralf und Andreas – für den 'Black Crusade', das 'Hellfest' und die unvergessenen Backstagenächte mit Tool und Co.! Und zum anderen an den uneinnehmbaren Laborblock inklusive Gäste: Matthias ('den Rehrücken mit Rotwein ablöschen ...'), Felix ('es ist schon kurz nach vier ...'), Maren ('Seufz, muss noch in die Zellkultur'), Angie ('Telefon'), Catherine ('ich hab noch Wein zu Hause'), Diana ('magst 'n Kaffee'), Sara ('Roadrunner'), Papa-Christian, Clemens ('bitte mit Kartoffelsalat'), Space-Sandra, Landsfrau-Simone und unvergessen Sonja!

Meiner deutschen und meiner neuen türkischen Familie danke ich für ihre Unterstützung und für ihre menschliche und kulturelle Bereicherung.

An dieser Stelle möchte ich einige Worte meinen Eltern widmen: Eure selbstlose Unterstützung, euer Mut, eure unerschöpfliche Kraft und eure nie schwindende Zuversicht – auch in den traurigsten und dunkelsten Stunden – habe ich immer bewundert, haben mich geprägt und haben all das erst möglich werden lassen!

Dem wichtigsten Menschen in meinem Leben gehört der größte Dank: Dafür, dass du immer an mich geglaubt hast; dafür, dass du allem erst einen Sinn gegeben hast; dafür, dass es eine wunderbare Zukunft geben wird, in der wir gemeinsam alles meistern und all unsre Träume wahr werden lassen können. Dafür, möchte ich mich bei dir – meine Işın – bedanken. Seni çok seviyorum!

6.3 Ehrenwörtliche Versicherung

Ich versichere hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den

.....

(Torsten Fauth)

6.4 Erklärung

Hiermit erkläre ich, dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

München, den

.....

(Torsten Fauth)

6.5 Curriculum vitae

Name	Torsten Fauth
Date of birth	18 th December 1978
Place of birth	Göppingen
08/2005 – present	PhD thesis at the Ludwig-Maximilians-University Munich, Adolf-Butenandt-Institute, Prof. Dr. Peter B. Becker Title: 'Interaction of the Dosage Compensation Complex with DNA'
11/2004 – 04/2005	Research scholar at the University of North Carolina at Greensboro, Prof. Dr. Vincent C. Henrich funded by a scholarship from the 'Landesstiftung Baden-Württemberg'
11/2003 – 09/2004	Diploma thesis at the University of Ulm, Department of General Zoology and Endocrinology, Prof. Dr. Margarethe Spindler-Barth Title: 'Functional characterization of the <i>Drosophila melanogaster</i> ecdysteroid receptor'
10/1999 – 10/2003	Studies of biology at the University of Ulm 'Diplom Biologe'
07/1998 – 04/1999	Military service, Memmingerberg
08/1989 – 06/1998	Freihof-Gymnasium, Göppingen 'Abitur'

Publications

Fauth, T., Müller-Planitz, F., König, C., Straub, T., and Becker, P.B. The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome, *Nucleic Acids Res* 2010; doi: 10.1093/nar/gkq026.

Gilfillan, G. D., König, C., Dahlsveen, I. K., Prakoura, N., Straub, T., Lamm, R., **Fauth, T.**, and Becker, P. B. (2007) Cumulative contributions of weak DNA determinants to targeting the *Drosophila* dosage compensation complex, *Nucleic Acids Res* 35, 3561-3572.

Beatty, J., **Fauth, T.**, Callender, J. L., Spindler-Barth, M., and Henrich, V. C. (2006) Analysis of transcriptional activity mediated by *Drosophila melanogaster* ecdysone receptor isoforms in a heterologous cell culture system, *Insect Mol Biol* 15, 785-795.

Greb-Markiewicz, B., **Fauth, T.**, and Spindler-Barth, M. (2005) Ligand binding is without effect on complex formation of the ligand binding domain of the ecdysone receptor (EcR), *Arch Insect Biochem Physiol* 59, 1-11.

Grebe, M., **Fauth, T.**, and Spindler-Barth, M. (2004) Dynamic of ligand binding to *Drosophila melanogaster* ecdysteroid receptor, *Insect Biochem Mol Biol* 34, 981-989.