Aus dem
Biomedizinischen Centrum (BMC)
Lehrstuhl für Molekularbiologie
Institut der Ludwig-Maximilians-Universität München
Vorstand: Prof. Dr. Peter B. Becker

RNA modulation of structure and function of the *Drosophila* MSL complex *in vitro*

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität München

vorgelegt von
Anna Elisabeth Kiss

aus
Marburg

2024

Mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuer: Prof. Dr. Peter B. Becker

Zweitgutachter: Prof. Dr. Stefan Stricker

Dekan: Prof. Dr. med. Thomas Gudermann

Tag der mündlichen Prüfung: 27. September 2024

Erklärung zur Übereinstimmung der gebundenen Ausgabe der Dissertation mit der elektronischen Fassung

Promovierende Person:	Anna Elisabeth Kiss
Adresse:	Gollierstraße 25, 80339 München

Hiermit erkläre ich, dass die elektronische Version der eingereichten Dissertation mit dem Titel

RNA modulation of structure and function of the Drosophila MSL complex in vitro

in Inhalt und Formatierung mit den gedruckten und gebundenen Exemplaren übereinstimmt.

München, 27.09.2024   Anna Elisabeth Kiss

_____   _____
Ort, Datum   Unterschrift Doktorandin/Doktorand

Affidavit

Eidesstattliche Versicherung

Kiss, Anna Elisabeth

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

RNA modulation of structure and function of the *Drosophila* MSL complex *in vitro*

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 27. September 2024   Anna Elisabeth Kiss

_____   _____
Ort, Datum   Unterschrift Doktorandin bzw. Doktorand

Table of contents

Table of contents

Table of contents

Table of contents

Preface

In this thesis, I present and discuss my work on the two aspects of the *Drosophila melanogaster* dosage compensation complex, the structural and the functional aspect. The introduction covers both aspects, but I present the results and discussion separately as they can be perceived as two separate projects.

Parts of the work presented in this thesis have been published:

1. Jagtap, P. K. A., Müller, M., **Kiss, A. E.**, Thomae, A. W., Lapouge, K., Beck, M., Becker, P. B., & Hennig, J. (2023). Structural basis of RNA-induced autoregulation of the DExH-type RNA helicase maleless. *Molecular Cell*, 83(23), 4318–4333.e10. https://doi.org/10.1016/j.molcel.2023.10.026

2. **Anna E. Kiss**, Anuroop V. Venkatasubramani, Dilan Pathirana, Silke Krause, Aline Campos Sparr, Jan Hasenauer, Axel Imhof, Marisa Müller, Peter B. Becker: Processivity and specificity of histone acetylation by the male-specific lethal complex. *Nucleic Acids Research*, 2024, gkae123, https://doi.org/10.1093/nar/gkae123

1 Abstract

Gene expression is highly regulated in eukaryotes by histone modifications and corresponding reader-writer proteins. In the fruit fly *Drosophila melanogaster* the gene expression from the single male X chromosome is doubled, matching the transcriptional output from the two female X chromosomes. Absence of this gene dosage compensation results in male fly lethality. The molecular machinery involved is called dosage compensation complex (DCC), which consists of five proteins and a long non-coding RNA. The proteins of the DCC are the histone acetyltransferase MOF, the male-specific lethal proteins (MSL1, MSL2, MSL3) and the helicase MLE. At least one of two long non-coding RNAs roX1 or roX2 are critical for male fly viability *in vivo*. How the lncRNA impacts the function of the DCC remains hitherto unknown.

A hypothesis states that roX incorporation leads to structural changes of the DCC and potentially enhances one of its enzymatic activities, particularly its acetylation activity. To date, only limited information on the structure of the complete complex is available. Thus, I applied crosslinking mass spectrometry (XL-MS) to analyze structural interfaces within the DCC in absence or presence of roX RNA. Novel MSL-MSL protein interactions were found and previously reported contact sites were confirmed. Addition of roX2 RNA subtly changed XL patterns in MLE and the MSL1-MSL2 submodule of the MSL complex, which could hint towards a conformational change upon roX2 integration.

Transcriptional upregulation of the male X chromosome is molecularly linked to H4K16 acetylation by the HAT MOF. Hypothetically the activity and the specificity of MOF could be regulated by roX2 RNA. To address this hypothesis, I reconstituted nucleosome arrays *in vitro* and used purified MSL complexes for acetylation assays. These assays were evaluated by mass spectrometry, which allows for accurate site-specific acetylation identification and quantification. In absence of RNA MOF within the MSL complex is active, however not selective for H4K16ac. At longer incubation times the complete H4 tail is acetylated, starting at H4K16 and progressing outwards to H4K12, K8 and K5. This zipper-like processive behavior is supported by mathematical modelling. Upon addition of roX2 RNA or unrelated long RNA the oligo-acetylation of the H4 tail is suppressed, even at prolonged incubation times. If this effect can be linked to roX RNA *in vivo* or can be ensured as well by heterogeneous nuclear RNA (hnRNA) remains to be elucidated. Finally, dTIP60, which is a HAT complex of the same enzymatic family, does not show the processive oligo-acetylation mechanism and is not impacted by RNA.

In conclusion, the lncRNA roX2 induces subtle conformational changes in MLE and the MSL1-MSL2 submodule of the MSL complex. Moreover, it increases the specificity of the HAT MOF towards H4K16ac in nucleosome arrays. The possible connection between these roles via a shared allosteric mechanism awaits further investigation.

2 Zusammenfassung

Die Regulation der Genexpression ist lebenswichtig in allen Eukaryonten und wird durch Histonmodifikationen und dazugehöriger Leser- und Schreiber-Proteine vermittelt. In der Fruchtfliege *Drosophila melanogaster* ist die Verdopplung der Genexpression des einzelnen männlichen X-Chromosoms notwendig, um der Gendosis der zwei weiblichen X-Chromosomen zu entsprechen. Ein Defekt dieser Dosiskompensation ist für die Männchen letal. Der Dosiskompensationskomplex (DCC) besteht aus fünf Proteinkomponenten und einer langen nicht-kodierenden RNA (lncRNA). Die involvierten Proteine sind die Histonacetyltransferase MOF, die „male-specific lethal" Proteine (MSL1, MSL2, MSL3), sowie die Helikase MLE. Außerdem wird eine der beiden lncRNAs roX1 oder roX2 benötigt, um die Letalität der männlichen Fliegen *in vivo* zu verhindern. Wie genau die lncRNAs diese Funktion erfüllen, ist bis heute nicht verstanden.

Eine Hypothese besagt, dass sich durch den Einbau von roX RNA die Struktur des DCC ändert, was wiederum eine der enzymatischen Aktivitäten des DCC aktivieren könnte, insbesondere die Acetylierungsfunktion. Die Kenntnisse über die vollständige DCC-Struktur sind bisher lückenhaft. Daher habe ich den DCC und einige Subkomplexe in An- und Abwesenheit von roX2 RNA mittels Crosslinking-Massenspektrometrie (XL-MS) untersucht. Dabei wurden neue MSL-MSL Interaktionen gefunden und bereits bekannte Kontaktflächen bestätigt. Die Zugabe von roX2 RNA ergab kleine Änderungen der XLs in MLE und auch im Subkomplex von MSL1 und MSL2, was auf Konformationsänderungen hindeutet.

Die Transkriptionsaktivierung des X-Chromosoms ist mit der H4K16 Acetylierung durch MOF verknüpft. Daher ist eine weitere Hypothese, dass roX RNA für die Aktivierung von MOF und die Selektivität für H4K16 benötigt wird. Um diese Hypothese zu testen, habe ich Nukleosomen-Arrays *in vitro* rekonstituiert und die gereinigten MSL-Komplexe für Acetylierungsstudien verwendet. Die exakte Acetylierungsposition im Amino-Terminus der Histone kann mittels Massenspektrometrie bestimmt und gleichzeitig quantifiziert werden. MOF ist in Abwesenheit von RNA aktiv, jedoch nicht selektiv für H4K16. Insbesondere bei längeren Inkubationszeiten werden alle Lysine im H4 N-Terminus acetyliert. Mithilfe von mathematischer Modellierung konnte darüber hinaus gezeigt werden, dass MOF die Acetylierung am K16 beginnt und fortschreitend von innen nach außen K12, K8 und K5 modifiziert. Durch Zugabe von roX2, aber auch durch andere lange RNA, werden diese Oligo-Acetylierungsmuster unterdrückt. Ob diese erhöhte Spezifität in Anwesenheit von RNA *in vivo* roX2 spezifisch sein könnte, oder ob allgemein unspezifische heterogene nukleäre RNA (hnRNA) denselben Effekt hat, muss weiter untersucht werden. Abschließend konnten bei der Histonacetyltransferase Tip60, die mit MOF verwandt ist, weder Prozessivität noch ein Einfluss von RNA auf die enzymatische Aktivität gezeigt werden.

Zusammengefasst induziert die Zugabe von roX2 RNA kleine strukturelle Veränderungen in der Helikase MLE und im MSL1-MSL2 Subkomplex. Darüber hinaus, verhindern roX2 und andere lange RNAs die Oligoacetylierung des H4 N-Terminus durch MOF. Ob diese beiden Aspekte mittels eines allosterischen Mechanismus der roX RNA zusammenhängen, benötigt weitere strukturelle Aufklärung.

3 Introduction

3.1 An old hat

The origins of chromatin research can be traced back to 1882 when Walther Flemming first introduced the term 'chromatin' [1]. He coined this term after discovering readily dyeable substances within cell nuclei [1, 2]. Flemming began his work with an analogy that likened his exploration to that of a settler clearing the first tree in a wild, untouched forest [1]. This analogy, both interesting and modest, prompts us to consider how it might be updated in today's context. The forest, metaphorically representing our understanding of chromatin, has changed significantly. The colossal trees have been felled, symbolizing the advancements made in the field. The settler no longer wields a primitive axe but operates fast, efficient, and intricate machinery, surrounded not by solitude but by a network of collaborators and competitors.

From the initially 'colorfully' stainable substance in the cell nucleus known as chromatin, it took many decades to unravel that nucleic acids held the role of the genetic information's code [3]. However, despite deciphering the DNA structure, the regulation governing how this information could be extracted and utilized remained mysterious [4-6]. Another decade passed until a connection emerged between the regulation of transcription and modifications to the DNA-organizing histone proteins [7]. This breakthrough opened the door to the realm of epigenetics, acknowledging that proteins play a vital role in conveying regulatory information, a significance *on par* with that of the genetic material itself, with profound implications for development and health [8].

3.2 Chromatin

Chromatin is a complex of DNA and proteins found in the nucleus of eukaryotic cells [9, 10]. Chromatin allows for compacting DNA into a denser, compacted, yet organized structure. During different stages of the cell cycle, chromatin undergoes various structural changes, allowing for processes like DNA replication, transcription, and repair. Accessibility of the genetic information within chromatin plays a crucial role in regulating gene expression and ultimately determining the function of a cell. These chromatin basics are nowadays textbook-knowledge [8-10]. However, the epigenetic landscape undergoes time-dependent regulation during development, cell cycle, environmental changes and circadian rhythm, which still remain elusive. The crosstalk between different histone modifications and the intricate interplay between metabolism and the genome via the epigenome, remain at the forefront of current investigation.

3.3 Histone Modifications

Histones are the protein building blocks that, together with DNA, compose the nucleosome (Figure 1A, [11]). A DNA helix of 147 bp is wrapped around the histone octamer (Figure 1A, [11]). The nucleosome core consists of two copies of each H2A, H2B, H3 and H4. The histones are highly conserved proteins in eukaryotes, which display a specific 'histone fold' structure (Figure 1A, [11]). From the nucleosome core the histone tails are protruding, which leaves them accessible for chromatin binding proteins to recognize (Figure 1A, [11]). In addition to nucleosome binding, chromatin modifying enzymes can place histone modifications on the different tails (Figure 1B, [8, 12-14]). Histone modifications include various biochemical changes such as methylation, acetylation, phosphorylation, ubiquitination, and sumoylation, among others. Further modifications are still being discovered, and the crosstalk between two or multiple modifications is a field of current interest [15-17]. Histone modifications play key roles in regulating gene expression, DNA repair, chromosome condensation, and other chromatin-related processes. They can affect how 'tightly' DNA is packaged and influence the accessibility of genes to the transcription machinery (Figure 1B, [18, 19]). Histone modifications thus have profound effects on cellular functions and are integral to the complex mechanisms governing gene regulation.

In addition to their role in regulating gene expression and chromatin structure, histone modifications are also involved in various biological processes such as cell differentiation, development, and disease progression. Dysregulation of histone modifications has been implicated in numerous human diseases, including cancer, neurological disorders, and autoimmune conditions. Therefore, understanding the mechanisms and functions of histone modifications is of great importance in both basic research and clinical applications, offering potential targets for therapeutic interventions and diagnostic strategies.

3.4 Histone acetylation

One of the first histone modifications identified was acetylation [7]. Gain in accessibility and activation of transcription are the most common consequences of histone acetylation [20]. Particularly, the unmodified H4 tail is involved in chromatin compaction, because it allows the basic patch of H4 to interact with the acidic patch of the H2A-H2B dimer of a neighboring nucleosome (Figure 1; [18-20]). Upon acetylation of the H4 tail nucleosome stacking will be interrupted (Figure 1, [21]). Moreover, according to a log-held view, histone tail acetylation excludes repressive modifications at the very same lysine [12, 22]. Recently, the double modification of acetylation and methylation on the same lysine residue was discovered, undermining the prevailing paradigm of only one of the two modifications could be placed on a certain residue at a time [17]. The modification named 'Kacme' is most likely a specific mark for epigenetic reader proteins [17].

Introduction
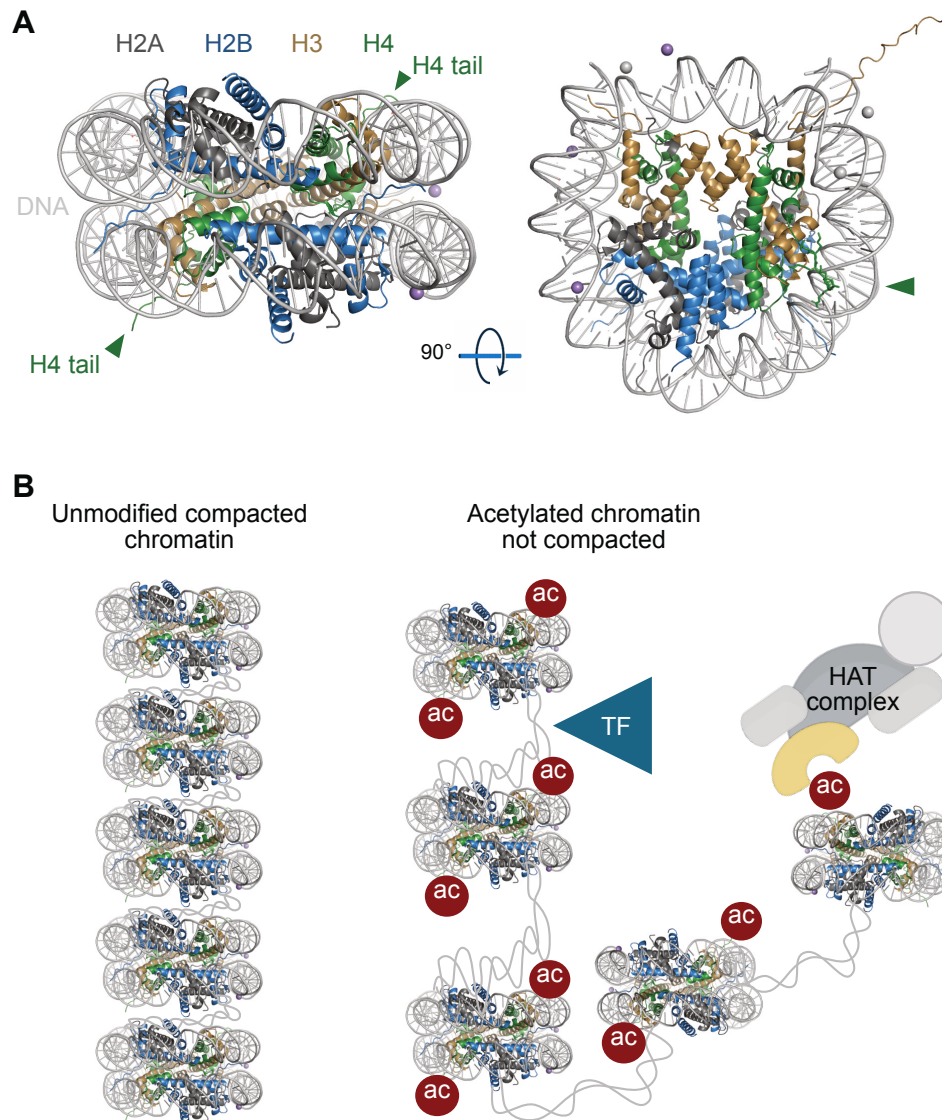
Figure 1: The nucleosome is the packaging unit of chromatin. **A** Structure of the nucleosome with 147 bp of DNA wrapped around [11]; **B** Without any histone modifications the nucleosomes are able to stack on top of each other leading to a compacted chromatin, which is transcriptionally inactive [18, 20]. If the nucleosomes are modified, e.g. by acetylation, the nucleosomes are repelling each other, such that the spacing between them will increase [21]. This allows transcription factors (TF) to bind to their target sequences and recruit the transcription machinery. Histone acetylations are placed by the enzyme class of histone acetyltransferases (HATs), which are often integrated in larger protein complexes. Figure from Luger *et al. [11]*, adapted with permission by Springer Nature.

3.5 Balancing sex chromosomal gene expression – basics of dosage compensation

Sex chromosomes, such as the X and Y chromosomes, determine an individual's genetic sex. In many species females typically possess two X chromosomes (XX) and males have one X and one Y chromosome (XY), leading to a gene dosage imbalance [23]. The Y chromosome, containing limited genetic information, is usually not dosage compensated, emphasizing the significance of the X chromosome in gene dosage regulation [24]. Historically, biologist Theodor Boveri's observations in sea urchin embryos underscored the importance of balanced chromosomal distribution for organismal

viability [25], while Thomas Hunt Morgan's discoveries in *Drosophila melanogaster* laid the groundwork for understanding X chromosomal genetics [26]. Dosage compensation mechanisms aim to equalize gene expression between sexes and between autosomes and sex chromosomes [23, 27-29]. In mammals, X-chromosome inactivation (XCI) in females ensures gene expression balance, whereas other organisms employ different strategies. For example, *Caenorhabditis elegans* halves the expression of both X chromosomes in hermaphrodites [24, 30, 31]. Moreover, the balance of the gene dosage between X chromosomes and autosomes involves an upregulation of transcription of genes on the active X in mammals [27, 32, 33]. Mechanistically this transcriptional activation of X-linked genes involves as well H4K16ac and the activity of the histone acetyltransferase MOF in mammals [33].

Research into the mechanism of dosage compensation in *D. melanogaster* has uncovered a notable divergence from mammals [23, 29]. Unlike mammals, where one X chromosome is suppressed in females, male fruit flies upregulate their single X chromosome approximately twofold [23, 29]. This upregulation, essential for male viability, correlates with increased H4K16ac presence on the X chromosome [34]. Additionally, the male X chromosome forms a distinct territory visible under standard fluorescent microscopes [35-39].

3.6 The composition of the *Drosophila melanogaster* dosage compensation complex

The male-specific lethal (MSL) complex or dosage compensation complex (DCC) is a ribonucleoprotein complex. It consists of five protein components MSL1, MSL2, MSL3, MOF and MLE, and additionally long non-coding RNAs roX (Figure 2). According to the current model, the MSL complex first binds to PionX (Pioneering on the X) sites, then spreads to high-affinity (HAS) sites and subsequently to active genes, which are recognized by the presence of H3K36me3 (Figure 2, [29, 40]). This process results in the acetylation of chromatin at active genes by MOF, leading to elevated levels of H4K16ac and increased transcriptional activity on the male X chromosome (Figure 2).

The first gene associated with dosage compensation in flies was *'maleless'* (*mle*) [41, 42], followed by the identification of *male-specific lethal* genes (*msl-1, msl-2,* and *msl-3*) [43, 44]. It took over 15 years to identify *mof* (*males absent on the first*) due to its location on the X chromosome [45]. Subsequently, the discovery of the long, non-coding RNA roX (RNA on the X) revealed its role in dosage compensation, though the exact molecular mechanisms remain unclear [38, 39, 46-48].
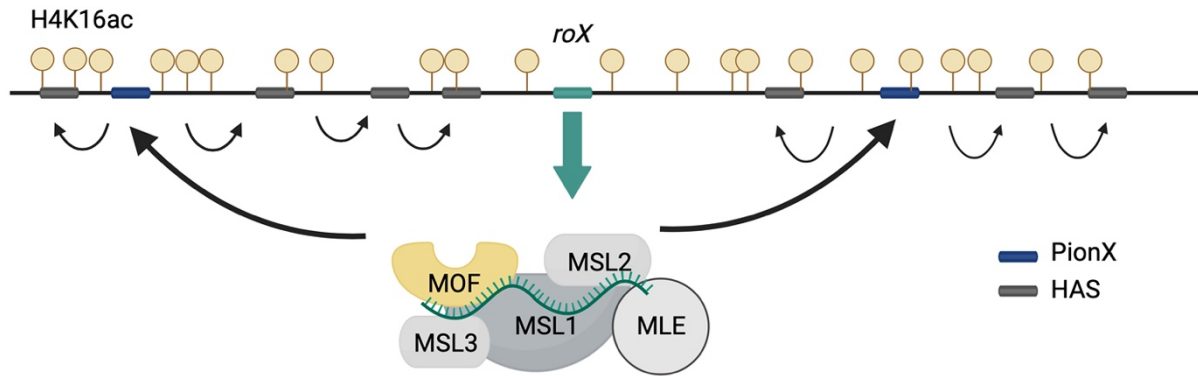
Figure 2: The dosage compensation complex binds to the *Drosophila melanogaster* male X chromosome and up-regulates transcription of X chromosomal genes through H4K16 acetylation by MOF [29]. The current model envisions a first step of binding to PionX (Pioneering on the X) sites [40], followed by "spreading" to high affinity sites (HAS) and to active genes, marked by H3K36me3. The chromatin at active genes is acetylated by MOF and thus the male X chromosome shows elevated H4K16ac and higher transcriptional output. Figure created with Biorender after Samata and Akhtar, 2018, with permission by Annual Reviews [29].

3.6.1 The 'scaffold' MSL1

MSL1 of *Drosophila melanogaster* is a large protein with 1039 amino acids (aa), containing a coiled coil region in the N-terminus and a PEHE domain in the C-terminus (Figure 3). The coiled coil region is conserved in the mammalian homologous protein, in which the interaction with MSL2 and dimerization are structurally characterized (Figure 4, [49]). A long IDR domain dominates the MSL1 central part in *D. melanogaster*, which is not conserved in mammals (Figure 3). The PEHE domain at the MSL1 C-terminus, where interaction with MOF and MSL3 are described, is conserved across species (Figure 4, [50]).
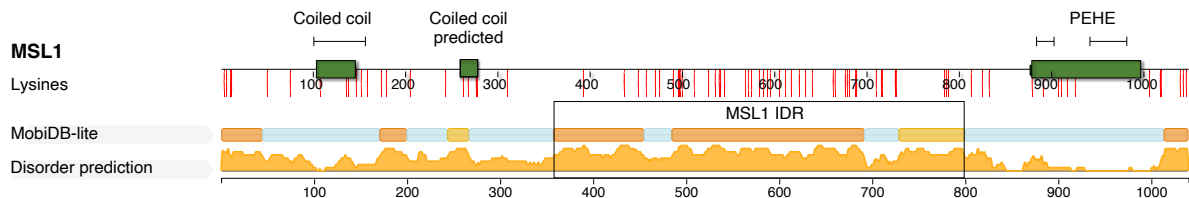
Figure 3: Domain scheme of MSL1 of *D. melanogaster*. The 1039 aa protein contains an N-terminal coiled coil, a predicted coiled coil and a C-terminal PEHE domain. Crystallized regions are marked by a black line above the cartoon [49, 50]. Lysine residues represented by red lines, are distributed throughout the protein, however many of them are found in the central IDR. Mobi-DB lite disorder prediction score presented below the domain scheme [51].
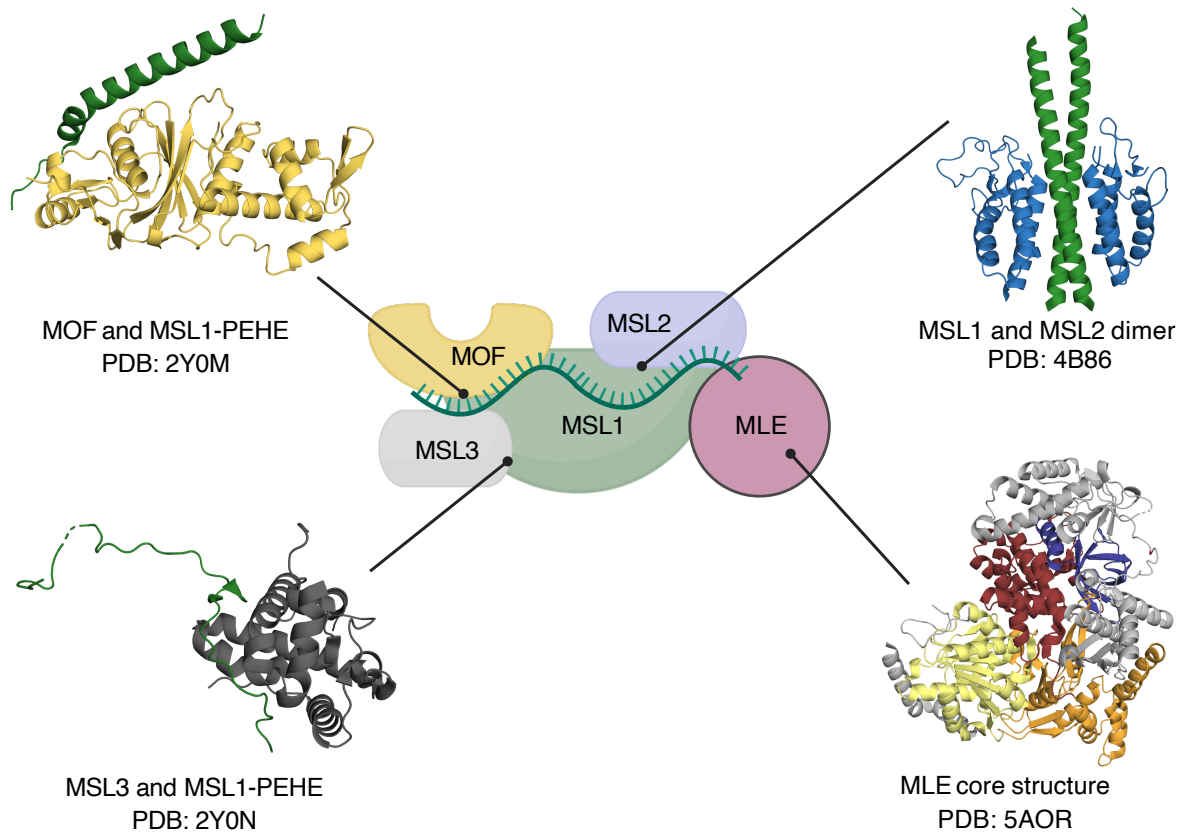
Figure 4: Structurally described interaction surfaces of protein subunits of the DCC. MSL1 interacts at its C-terminal PEHE domain both with MOF and MSL3 [50]. MSL1 and MSL2 dimerize through the coiled coil of MSL1 and the RING domain of MSL2 [49]. The MLE core is crystallized with ssRNA [52]. Structure visualization with pymol (Schrödinger).

3.6.2 The male-specific nucleic acid binder MSL2

The MSL2 protein (773 aa) is the male-specific factor of the DCC as it is not expressed in female fly cells [53]. The gene expression of *msl-2* is regulated by sex-lethal (SXL) by suppressing translation of msl-2 mRNA in females while allowing for MSL2 translation in males [54-56].

MSL2 is a DNA binding protein, which recognizes the X chromosome in male flies at specific binding sites (Figure 2, [40, 57-59]). If MSL2 is expressed in female cells, it will recognize its binding sites on the X chromosome as well [53, 60]. The presence of MSL2 triggers the assembly of the other MSL proteins into the MSL complex and helps stabilization of roX RNA [39, 61, 62]. This leads to MSL complex binding to the X chromosome, X territory formation and H4K16 acetylation, a hallmark of dosage compensation [29, 34].

Additionally, MSL2 is an E3 ubiquitin ligase, which can ubiquitylate proteins [49]. The ubiquitylated proteins can be recognized for further cellular pathways, for example, for proteosomal degradation, however the detailed mechanism is poorly understood [49]. Among the target proteins of MSL2 are the other MSL proteins, such as itself, MSL1, MSL3, MOF and MLE (*unpublished*) [59, 63].

MSL2 contains an N-terminal 'RING' domain, which harbors the E3 ligase activity and interacts with MSL1 (Figure 5, [49, 59]). The CXC domain, which harbors zinc finger motifs, is involved in nucleic acid binding and targeting to the X chromosome [64-67]. The C-terminal domain of MSL2 harbors a

dominantly basic and proline-rich ('Pro-rich') region, which is thought to be important for RNA binding, as well [64, 67].
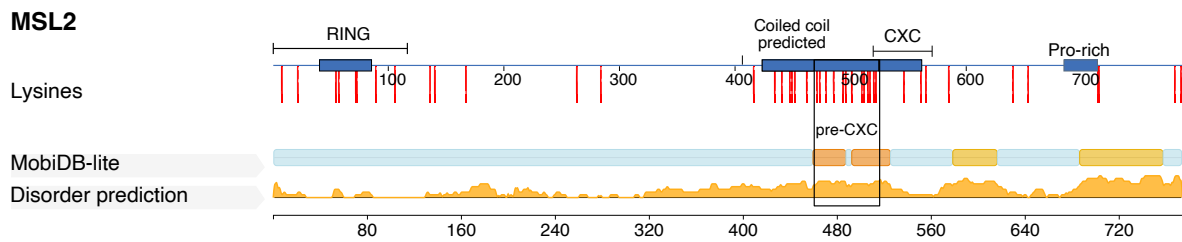
Figure 5: Domain scheme of MSL2 of *Drosophila melanogaster*. The N-terminal RING domain was co-crystallized with MSL1 (black line above represents the crystallized domain) [49]. The CXC domain was analyzed by NMR (black line above) [66]. Lysine residues (red lines) are present throughout the amino acid sequence, however accumulate in the 'pre-CXC' region between a predicted coiled coil and the CXC domain. The 'pre-CXC' region and the C-terminus have a high disorder score according to MobiDB-lite disorder prediction (below) [51].

MSL2's preference for selective DNA binding to Pioneering on the X (PionX) sites involves a consensus motif called the MSL recognition element (MRE) and specific DNA shape characteristics [40, 68]. MSL2 cooperates with CLAMP to enhance targeting efficiency and 'off-target' binding is counteracted by the GAGA-factor (GAF) [68, 69]. Additionally, fluorescence recovery after photobleaching (FRAP) experiments demonstrate MSL2's low-diffusion binding to the X chromosome territory [70]. This could be interpreted as a polymer phase-separated cellular compartment that could be induced by roX RNA within the nucleus [70].

3.6.3 The histone modification reader MSL3

MSL3 is often considered to be the histone modification 'reader' of the MSL complex. MSL3 is the smallest of the MSL proteins (512 aa). It has a conserved chromobarrel domain (chromo) at the N-terminus (aa 2-91), which is linked to the also conserved MRG domain by a long, flexible linker (Figure 6, [71-73]).
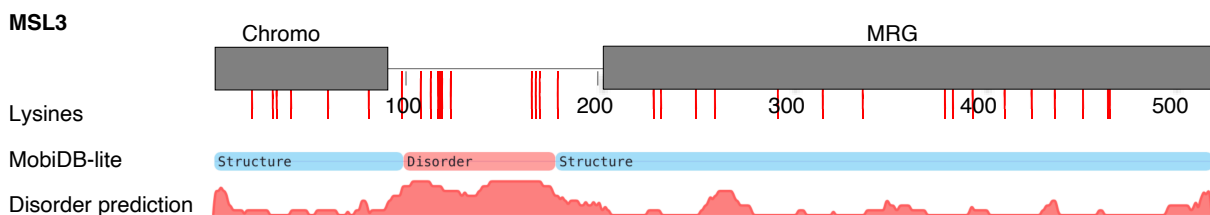
Figure 6: Domain scheme of MSL3 of *Drosophila melanogaster*. MSL3 harbors a chromodomain at the N-terminus ('Chromo') and an MRG domain at the C-terminus. The MRG domain has been structurally described in the mammalian conserved protein [50]. Lysine residues are marked as red lines below the scheme. The linker region between the chromo and the MRG domain has a high disorder score according to MobiDB-lite disorder prediction (below) [51].

The conserved chromodomain of MSL3 led to investigation on its potential interaction with various histone modifications, such as H4K20 and H3K36 methylation [74, 75]. However, evidence on affinity towards H3K36me3 and chromodomain mutants affecting dosage compensation shifted the focus towards this histone mark as the primary target for MSL3 [75, 76]. Nowadays, it is widely accepted that H3K36me3 is the most relevant target of MSL3 in the context of dosage compensation [29].

Introduction

H3K36me3 predominantly resides within actively transcribed regions of the genome, especially within gene bodies undergoing transcriptional activity [75, 77-80]. Although primarily associated with active transcription, H3K36me3 can also be found in certain types of heterochromatin, such as pericentric heterochromatin, albeit not as its main location [75, 78, 79, 81]. The trimethylation of lysine 36 on histone H3 (H3K36me3) is catalyzed by Set2, a histone methyltransferase, which travels with DNA-dependent RNA polymerase PolII [79, 80]. This methylation occurs predominantly within actively transcribed genes, particularly within intron-poor, highly transcribed housekeeping genes [75, 78].

Disruption of H3K36me3 trimethylation impacts dosage compensation and H4K16ac levels [75, 79, 80]. The prevailing model suggests that after initially binding to the X chromosome at PionX (or HAS) sites, the MSL complex spreads to active genes on the X chromosome by targeting H3K36me3 via MSL3 (Figure 2, [29]).

Early on, doubts were raised about the importance of MSL3's chromobarrel domain binding to the H3K36me3 mark, with evidence suggesting preferential binding to H4K20 monomethylated H4 peptides [74]. An alternative hypothesis proposed that the spreading mechanism relies on MSL3 acetylation by MOF and the subsequent loss of interaction with RNA roX2, by which the MSL complex would disassemble [82]. More recent, contentious findings have raised questions about the essential role of H3K36 methylation in MSL spreading on the X chromosome [83]. Additionally, MSL3 has been observed on pericentric heterochromatin, where H3K36me2 is prevalent [78], suggesting that MSL3 has the ability to interact with various substrates depending on molecular context, such as acetylation status [82]. This flexibility allows for tunable binding and targeting, enabling dissociation of the MSL complex and spreading from high-affinity sites to lower-affinity sites and gene bodies [29].

The C-terminal MRG domain of MSL3 is likewise conserved from *Drosophila* to mammals [50]. A crystallography study has shown that the MRG domain and the MSL1-PEHE domain interact (Figure 4, [50]).

Overall, the MSL proteins are conserved in humans [29, 84, 85]. However, they do not primarily regulate X chromosomal dosage compensation in mammalian species. Although their molecular role is still under current investigation, mutations in the *msl* genes, for instance in *msl-3* can cause developmental defects in humans [85, 86]. This syndrome, recently named the Basilicata-Akhtar syndrome, impairs the brain and facial development leading to mental retardation, growth defects and premature death of patients in their teenage years [86, 87]. The defect is related to aberrant H4K16ac due to loss of interactions between the human MSL proteins, when human MSL3 is defective [86].

3.6.4   The histone acetyltransferase MOF

MOF can form complexes not only with the MSL proteins, but also with the NSL (non-specific lethal) proteins, creating the NSL complex [88-91]. The NSL complex is present all flies independent of sex and is detected at promoter regions of chromatin. Different from the MSL complex, the NSL complex acetylates H4K5ac and K8ac [92]. In addition, the NSL complex is conserved in mammals, where MOF

executes its acetylation of chromatin [93, 94]. The mammalian NSL complex plays key roles in piRNA production from telomeres and ciliation in kidney cells, highlighting the importance of MOF in human physiology and disease [93, 94].

Moreover, MOF is indispensable in dosage compensation, because it serves as the effector within the MSL complex [29]. Its H4K16ac deposition on the male X chromosome enables the twofold upregulation of the transcription chromosome-wide [95, 96]. The enzymatic activity of MOF can be readily studied in biochemical assays. This is crucial because histone acetylation directly relates to dosage compensation, the main biological issue of interest of this project [34, 45, 95, 97].
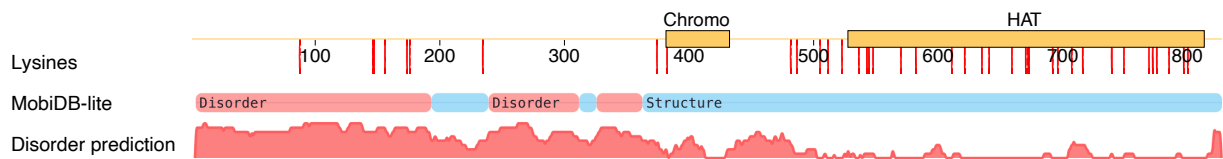
Figure 7: Domain scheme of MOF of *Drosophila melanogaster*. MOF harbors a chromobarrel domain ('Chromo') and a HAT domain at the C-terminus. The HAT domain has been structurally described [50]. Lysine residues are marked as red lines below the scheme. The N-terminus has a high disorder score according to MobiDB-lite disorder prediction (below) [51].

The MOF protein (827 aa) contains two functionally relevant domains (Figure 7, Figure 8). The N-terminus of MOF does not contain any conserved domains or distinct structural features, but has been proposed to play an important role for the autoinhibition of the HAT [98]. In the central part of MOF resides the chromobarrel domain, which was described to bind directly to RNA and/or DNA (Figure 7, [46, 98, 99]). In contrast, other chromobarrel domains are predominantly known to bind histone tails and histone modifications [73, 75, 76, 100, 101].

Introduction
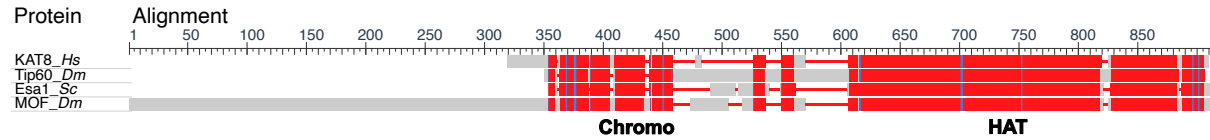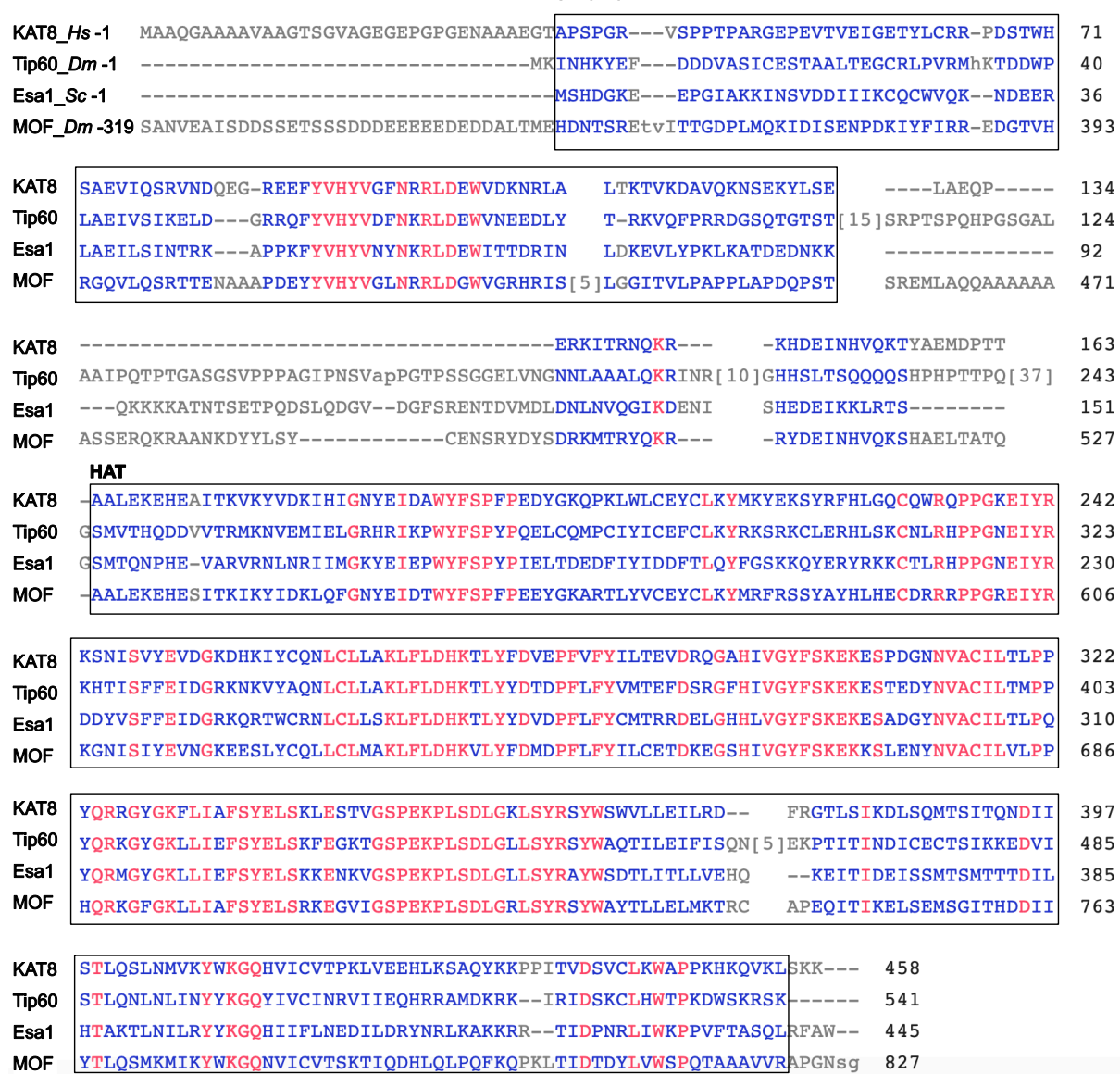
Figure 8: Alignment shows conserved domains of four MYST family HAT enzymes: KAT8 of *H. sapiens*, Tip60 of *D. melanogaster*, Esa1 of *S. cerevisiae* and MOF of *D. melanogaster*. **A** Overview the conserved domains are colored in red; deletions are symbolized by a horizontal line. **B** Detailed view of the conserved C-terminal domains of the HATs. Deletions are represented as a dash, non-conserved residues are in grey font, similar residues are in blue font and identical residues conserved are highlighted in red font. The chromo domain and the HAT domain of MOF (*D. melanogaster*) are boxed. Alignment performed with BLASTp [102, 103] and visualization by NCBI Multiple Sequence Alignment Viewer (https://www.ncbi.nlm.nih.gov/projects/msaviewer/, last access: 11.04.2024).

The acetyltransferase domain of MOF is found at its C-terminus and it belongs to the MYST family of acetyltransferases (Figure 8, [50]). In the MSL complex the MOF HAT domain interacts directly with the MSL1-PEHE domain (Figure 4, [50]). Examples of the MYST HAT family are widely conserved throughout evolution in eukaryotes (Figure 8, [50, 104, 105]).

Introduction

The main target on histones of the MOF acetyltransferase is the H4 tail. In the context of the MSL complex, the main substrate is H4K16, which is elevated on the single male X chromosome *in vivo [34, 95]*. Nevertheless, further substrates of acetylation are the autoacetylation of MOF and acetylation of other MSL members such as MSL3 [82, 106]. Moreover, MOF acetylates mitochondrial proteins [107, 108], is involved in metabolic pathways [109], can acetylate lamin [110], the histone demethylase LSD1 [111], the E3 ligase UHRF1 [112], the transcription factor YY1 [113], the immune response factor IRF3 [114], p53 [115], and the list is still to be completed.

3.6.5 The helicase MLE

MLE was the first factor associated with dosage compensation [41-43, 116]. It is the largest of the five dosage compensation proteins (1293 aa). MLE belongs to the DExH family of ATP-dependent helicases [117, 118]. This family of helicases, such as the mammalian orthologues DHX9 (also known as RHA), can ATP-dependently unwind double-stranded RNA, G-quadruplexes, DNA, triple-stranded DNA and DNA-RNA hybrids such as R-loops [119-123]. As a result of the multitude of substrates DExH helicases target, they are involved in many cellular functions, for example in transcriptional regulation, translation regulation and DNA damage repair [123-126]. Moreover, the mammalian helicases are described to play important roles in viral infections by positive strand RNA viruses such as the *Flaviviridae* Hepatitis C virus (HCV) and the bovine diarrhea virus (BVDV) [127-129]. DHX9 binds viral RNA, shields it from the innate immune system of the host cell, protects the viral genome from RNases and unwinds stem-loop structures which could impair translation [123, 130]. Its function in viral infections makes DHX9 a potential anti-viral drug target [123].

In flies, MLE is implicated in various functions beyond dosage compensation, including its involvement in RNA A-to-I editing and splicing [124, 131-134]. These additional roles underscore the versatility of MLE beyond its function in dosage compensation.
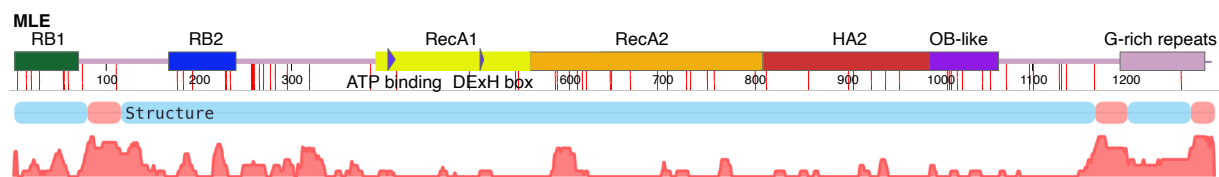
Figure 9: Domain scheme of MLE of *Drososphila melanogaster*. The MLE-core (aa 105–1158) has been crystallized bound to ssRNA [52]. Recently, the cryo-EM structure of the MLE-core was obtained bound to dsRNA [118]. Lysine residues are marked as red vertical lines below the scheme. The N-terminal linker and the G-rich repeats at the C-terminus have a high disorder score according to MobiDB-lite disorder prediction (below) [51]. Figure adapted from Prabu *et al.* with permission [52].

Two conserved dsRNA-binding domains (dsRB1 and dsRB2, Figure 9) reside in the MLE N-terminus, which are connected by a flexible linker [118, 135, 136]. Both of these domains have the characteristic dsRB fold to bind double stranded RNA [136], however only dsRB2 has been shown to be essential for the MLE helicase activity [52, 118, 137, 138]. The exact function of dsRB1 is unknown to date. The MLE helicase core can be further subdivided into the RecA1, RecA2 and HA domains, which bind ATP, contain the active site with the DExH-motif and exert the unwinding helicase function [52, 118]. C-

terminally of the HA domain resides the OB-fold or OB-like domain, and at the very C-terminus a G-rich domain is found. Due to the repetitiveness of the G-rich repeats this region was often excluded from structural studies to gain more uniform and compact molecular states [52, 118]. The crystal structure of the MLE helicase core with single-stranded poly-U RNA is known (Figure 4), as well as the recently published cryo-EM structure [118], which gives new insights into the unwinding mechanism of MLE [52, 118].

Contrary to DHX9, which has no obvious RNA sequence-preference and binds the nucleic acids primarily due to secondary structure features, such as G-quadruplexes, R-loops and triple-helical DNA [122, 139-142], MLE binds primarily roX RNA (roX1 or roX2) [143]. Nevertheless, additional substrates were identified in *in vitro* immunoprecipitation studies ('vitRIP'), which include U/A-rich sequences, with stem-loop secondary structures and are typically found in the fly brain [143]. The MSL 'core' complex (MSL1-MSL2-MSL3-MOF) plays a critical role in target selectivity of MLE. Through a series of experiments involving RNA binding assays and vitRIP, it was observed that MLE, in the presence of the MSL core, specifically enriches roX2 and a subset of other RNAs from the transcriptome [143].

The interactions between the DCC subunits and MLE remain poorly understood, representing an area of investigation in current research. The question of whether MLE functions as a transient or integral component of the DCC remains open. It is possible that MLE is predominantly transient within the DCC and dissociates following the transfer of roX RNA [135, 143, 144].

3.7   Long non-coding RNAs

3.7.1   Definition of long non-coding RNAs

Long non-coding RNAs (lncRNAs) are typically longer than 200 nucleotides [145]. These RNAs, even though they are not translated, can be 5'-capped, 3'-polyadenylated, spliced and post-transcriptionally modified [145]. Despite not having protein-coding potential, lncRNAs play crucial roles in various cellular processes and biological functions, including gene expression regulation, chromatin remodeling, epigenetic modifications, and cell differentiation [145, 146].

The lncRNAs can regulate gene expression at different levels, such as transcriptional, post-transcriptional, and epigenetic regulation. They can interact with DNA, RNA and proteins to modulate gene expression either by promoting or suppressing the activity of genes.

3.7.2   The nucleic acid components with elusive roles – roXes: roX1 and roX2

Among the best and most extensively studied examples of lncRNAs besides *Xist* are the roX RNAs of *Drosophila melanogaster [146]*. Two functionally redundant RNAs have been characterized in the past: roX1 and roX2 (Figure 10 A, [38, 47, 48]). These two lncRNAs differ greatly in size, roX1 contains approximately 3700 bases, while roX2 has a length of about 570 nt [144, 147]. Both RNAs are 3'-

Introduction

polyadenylated and can be spliced at various splice sites, resulting in multiple isoforms for both RNAs [148, 149]. roX1 and roX2 have little sequence conservation [150, 151], which is a common characteristic of lncRNAs and especially of roXes in the *Drosophila* genus [145, 146, 152-154]. Nevertheless, they share structural features of U/A base pair rich stem-loops (Figure 10 A, B). These U-rich sequence stretches are termed 'roX boxes' and can be bound by the MSL complex, mainly by MSL2. However, all subunits of the MSL complex possess RNA binding affinity and are hypothetically involved in the coordination of roX RNA.
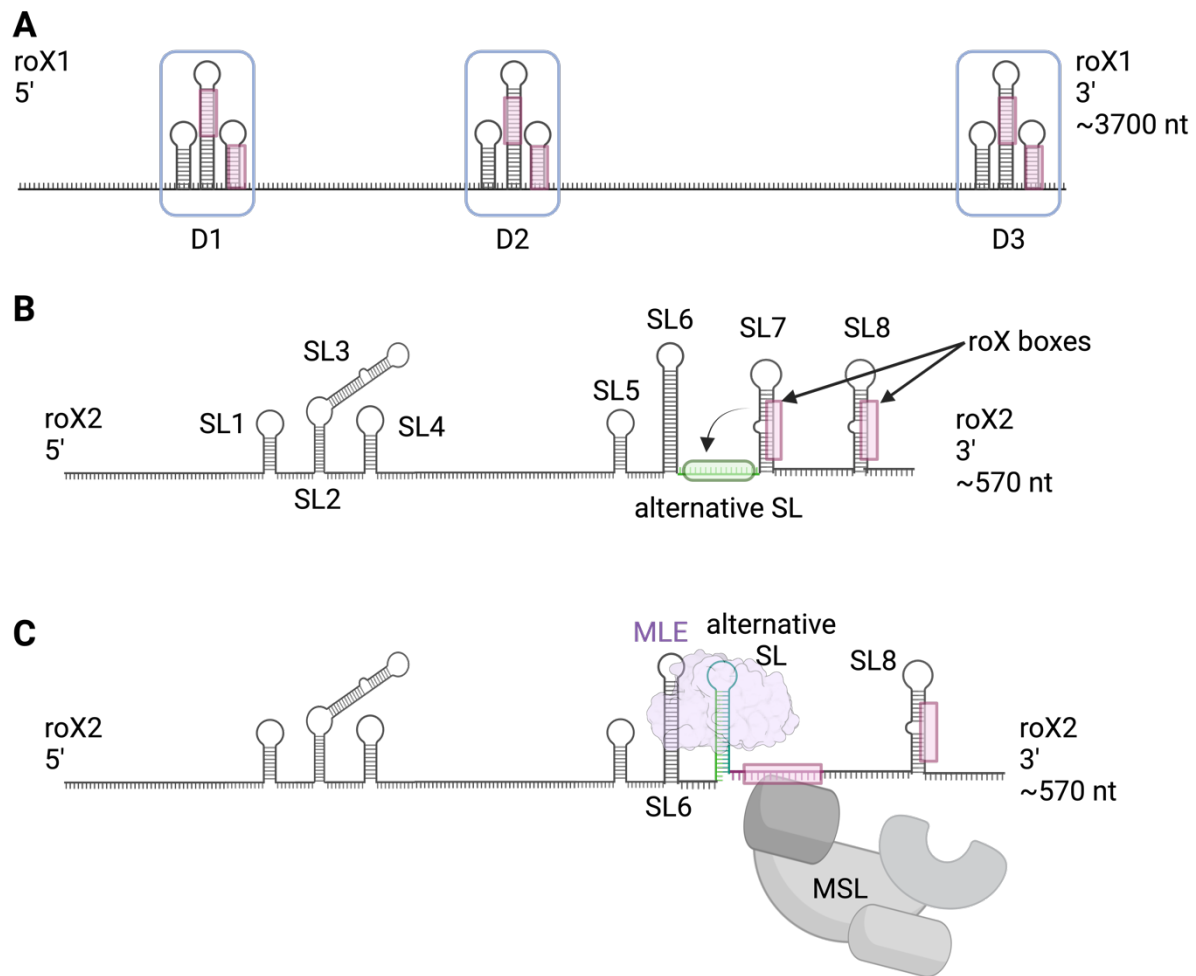
Figure 10: roX lncRNAs contain secondary structures. **A** roX1 lncRNA (3700 nt) contains three regions with stem loops and roX boxes (magenta), D1, D2 and D3, which are functionally important [155]. The stem loops contain 'roX boxes', which are sequences rich in U/A bases. **B** roX2 lncRNA (570 nt) contains eight stem loops with U-rich roX-boxes. Detailed view of the stem loops in roX2 [137, 144]. **C** roX2 can be remodeled by MLE and form an alternative stem loop (green) between SL6 and SL7. At the 3' end a roX box sequence becomes accessible and can be bound by the MSL complex [137, 156]. Figure created with Biorender and adapted after Maenner *et al.* with permission [137].

In roX1 there are three domains (D1, D2 and D3), which contain stem-loop structures with roX boxes and are important for the dosage compensation function of roX1 (Figure 10 A, [155]). Curiously, roX1 was not enriched with the MSL core complex *in vitro [143]*, for unknown reasons. It may suggest that roX1 has another role in dosage compensation, possibly during embryonic development, where it is detected earlier than roX2 [157-160].

The secondary structure of roX2 has been analyzed, revealing eight stem-loops along the 570 nt long sequence (Figure 10 B, [137, 144]). At the roX2 3' end, stem-loop SL7 and the sequence 5' of SL7 can be ATP-dependently remodeled by MLE into an alternative stem-loop (Figure 10 C, [156]). A 'minimal roX' of 240 nt length has been described, which contains a 5' SL and the 3' SLs with the possibility to convert into the alternative SL [156].

The MSL complex incorporates at least one of two long non-coding RNAs, roX1 or roX2 [29, 143]. RoX2 RNA has been found colocalized to the same genomic loci on the X chromosome as the MSL proteins [161, 162]. However a complementary Watson-Crick base pairing could be excluded [161], which indicates that the interaction with the genome is indirect. Due to the co-localization and co-immunoprecipitation with the MSL proteins, roX2 appears bound and integrated into the DCC complex *in vivo [162]*.

Despite many studies working on MSL-RNA interactions, MSL complex targeting and its evolution as well as its vital role for male flies remains elusive [70, 143, 163]. The molecular mechanism of how the lncRNA influences the MSL complex functionally is not understood. I aim to shed light on the structural and functional relationship between MSL complex and roX RNA *in vitro*.

3.8   Mass spectrometry

To address specific questions on molecular interactions (Figure 4) and potential structural changes upon roX RNA interaction (section 3.7.2) mass spectrometry was applied.

3.8.1   Crosslinking mass spectrometry as a tool in integrative structural biology

To shed light on the molecular mechanisms underlying dosage compensation, researchers sought to determine the structure of the MSL complex. Despite facing challenges due to the large size, numerous IDRs, and instability of MSL proteins when isolated, considerable progress has been made in the decade of 2010 [49, 50, 52, 66, 137]. As an alternative to X-ray crystallography, cryo-EM techniques are very powerful in structural analysis [118]. To support or validate data, so-called 'integrative structural biology' makes use of data from different techniques. One supportive technique is crosslinking mass spectrometry (XL-MS, Figure 11) [164-166].

Introduction
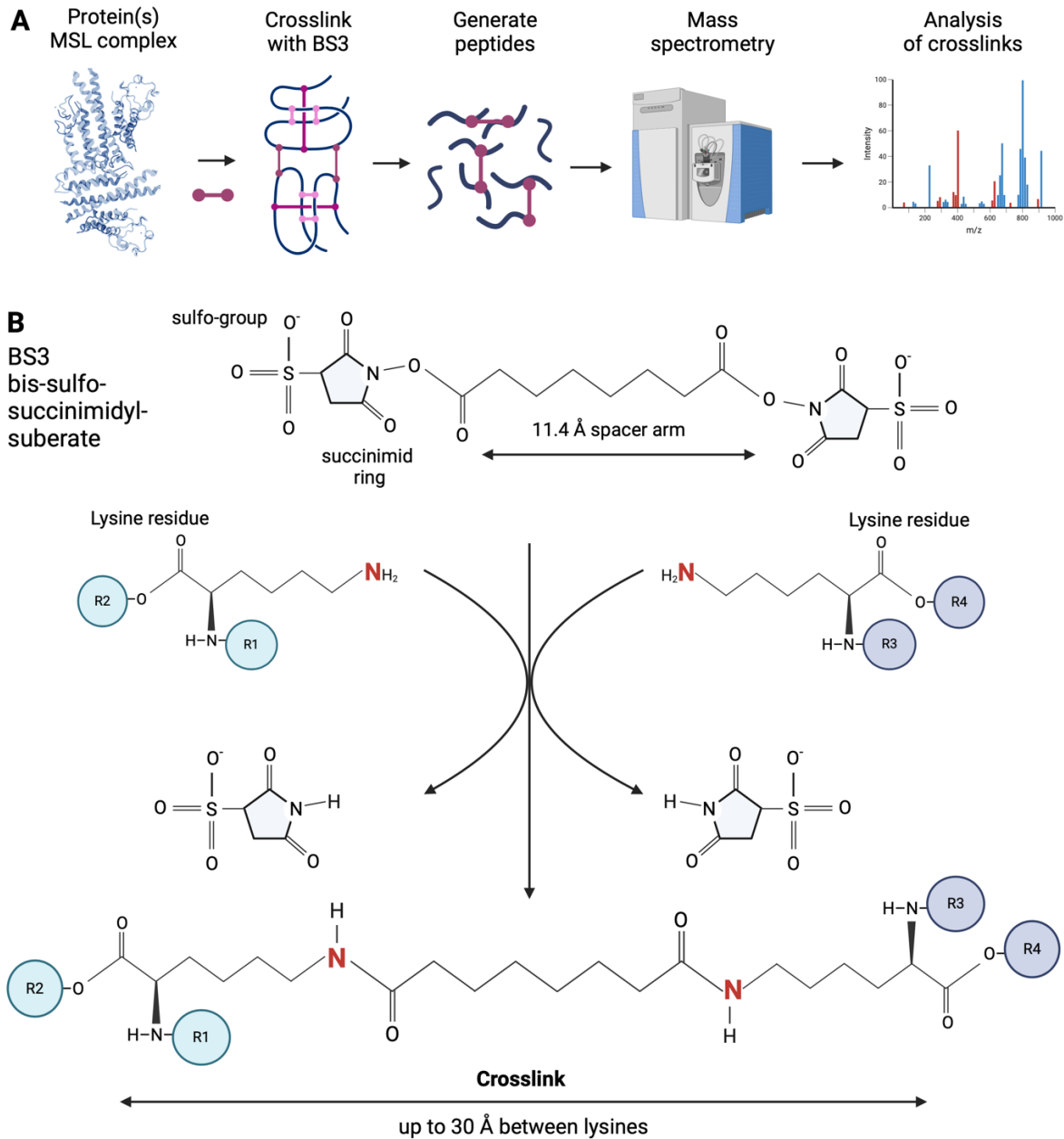
Figure 11: Crosslinking mass spectrometry can give valuable insights into interactions within proteins and protein complexes. **A** General workflow of XL-MS [166]. **B** Crosslinking reaction with BS3 at lysine residues of a protein of interest. Figure created with BioRender.

XL-MS employs crosslinking agents like BS3 bis(sulfosuccinimidyl)suberate to form covalent bonds between close-by residues, aiding in identifying interacting partners (Figure 11A, [166]). BS3, for instance, contains two reactive sulfo-succinimidyl groups linked by a 11.4 Å spacer arm, reacting readily with two lysine residues in spatial proximity (Figure 11B, [167, 168]). These crosslinked peptides are then analyzed by mass spectrometry, with the very defined mass measured for two crosslinked peptides allows tracing back the connected lysines, offering insights into intra- and inter-molecular interactions within multi-subunit protein complexes [169-171]. Additionally, validation or modeling approaches consider the spacer arm length and lysine side chain dimensions, accepting spatial distances of around 30 Å for BS3 (Figure 11B, [164, 168, 172]).

Even though it might not have the power to shine light on the entire structure of a protein complex, XL-MS is useful to explore interaction surfaces, which can then be further studied by orthogonal techniques.

3.9   Histone acetyltransferases

The process of protein acetylation involves the action of acetyltransferase enzymes, while its reversal is catalyzed by deacetylases (Figure 12, [173-175]). Specifically, when histones are the target proteins, the acetyltransferases are referred to as HATs, and the deacetylases as HDACs (Figure 12).

Figure 12: Protein acetylation is catalyzed by acetyltransferase enzymes, the reverse reaction is catalyzed by deacetylases, respectively. An acetyl-group is transferred from an acetyl-coenzyme A (Ac-CoA) to the amine group of a lysine residue of the target protein (e.g. histone). In the case of histones as target proteins, these acetyltransferases are called HATs and the deacetylases HDACs. Oftentimes, the HATs can acetylate other target proteins additionally to their histone substrates. Chemical formulas created with https://molview.org/ (27.03.2024).

Notably, HATs may also acetylate other proteins besides histones [176]. Other acylations, such as propionylation, butyrylation, 2-hydroxyisobutyrylation, succinylation, malonylation, glutarylation, crotonylation and beta-hydroxybutyrylation have been uncovered in the past decades [177-179]. Lysines are the residues primarily acylated, as the side chain features a primary amine [177, 180]. Histone acetylation and acylation offer interesting new avenues of research connecting the field of epigenetics with the field of metabolism.

Introduction

Several histone acetyltransferases are conserved throughout evolution, for example the human HAT KAT8 shares conserved domains with the *D. melanogaster* HATs Tip60 and MOF, as well as with the *S. cerevisiae* enzyme Esa1 (Figure 8). HATs can be classified into different enzyme families according to the composition of their HAT domain and conceivably the molecular reaction mechanism (Figure 8, [181, 182]). One evolutionary conserved HAT domain is the MYST domain, named after the founding members of the MYST family: MOZ, Ybf2/ Sas3, Sas2, and TIP60 [14]. MYST HATs exert their acetylation function on chromatin typically when incorporated into larger complexes [105, 183, 184]. The accessory subunits of these larger complexes can, for instance, recognize DNA damage sites [182, 185], interact with transcription factors [59, 182, 184, 186] and typically bind histone modifications on the nucleosomes nearby [40, 73, 74, 182, 187].

For some MYST HATs the molecular structure has been studied and determined, for example Esa1 [188, 189]). Esa1 of yeast is the MYST HAT within the large NuA4 complex [188-190]). The smallest functional subcomplex consists of three subunits Esa1, Yng2 and Epl1, which is known as the Piccolo NuA4 complex [191]. Additionally, a fourth subunit Eaf6 (Esa1 associated factor 6) is required for full functionality [189, 192]. In early studies, the NuA4 complex was already likened to the MSL complex as both are HAT complexes of the MYST family [104]. However, Esa1 does not appear to have strict substrate selectivity [190, 191, 193]. Later, it was established that the NuA4 complex is conserved in *D. melanogaster* as the DOMINO-A complex, which integrates a homologous HAT, Tip60 (Tat-Interacting Protein 60 kDa) [194]. The DOM-A complex has a likely submodule, the subcomplex, named dTIP60[piccolo] complex, which consists of the homologous proteins in *D. melanogaster* the HAT (Tip60) (Figure 8), Enhancer of Polycomb (EPc) and Inhibitor of growth family, member 3 (ING3) [104, 194-197].

Esa1 recognizes its substrates by a 'shallow' binding pocket, such that general lysine residues can be bound irrespective of the surrounding amino acids [188, 189, 198]. Prerequisite for full activity is an autoacetylation of Esa1 [181, 188]. In fact, Esa1 is not only known to acetylate the H4 tail, but also the H2A variant H2A.Z, which is involved in DNA damage repair [189, 194, 199]. This function is conserved and exerted by Tip60 in the *Drosophila* DOM-A complex as well (Figure 8, [194]).

Despite efforts to dissect the individual functions of the *Drosophila* MYST HATs, their regulation and specificity remain poorly understood. The complexity of histone modifications *in vivo*, coupled with indirect effects, hampers definitive conclusions [200]. However, studies in cells by RNAi against specific HATs of *D. melanogaster* aimed to clarify the specificity and crosstalk between different HATs [200, 201].

It can be difficult to determine the substrate specificity of a HAT solely from *in vivo* studies. Antibody specificity limited previous studies on specificity of HATs [202, 203]. Especially combinatorial patterns of acetylations on the same histone tail are often recognized by the antibody without discrimination of the different combinations. Moreover, the influence of one specific HAT can be difficult to clarify *in vivo* due to redundancies and secondary effects. One of the important influencers *in vivo*, which regulate

the acetylation state, are the histone deacetylases (HDACs, Figure 12, [192, 201, 204]). These enzymes can remove acetyl groups form modified proteins, in this case histones. They are typically ubiquitous in the cell nucleus and highly active without special specificity [201].

3.9.1 Mass spectrometry for histone acetylation

Do different histone acetyltransferases acetylate different histone tails and different lysines within the tails? To quantitatively study the HAT activity and to define the exact acetylation site, mass spectrometry has been employed [201, 205, 206].

The method was established earlier in our laboratory for the analysis of histone modifications in *Drosophila* tissue culture cells, in response to 'knock down' of histone modifiers of interest [201]. It follows a simple principle: Histone acetyltransferases transfer an acetyl group to their substrate proteins, which adds a mass of 42 Da to the target protein (Figure 13A).
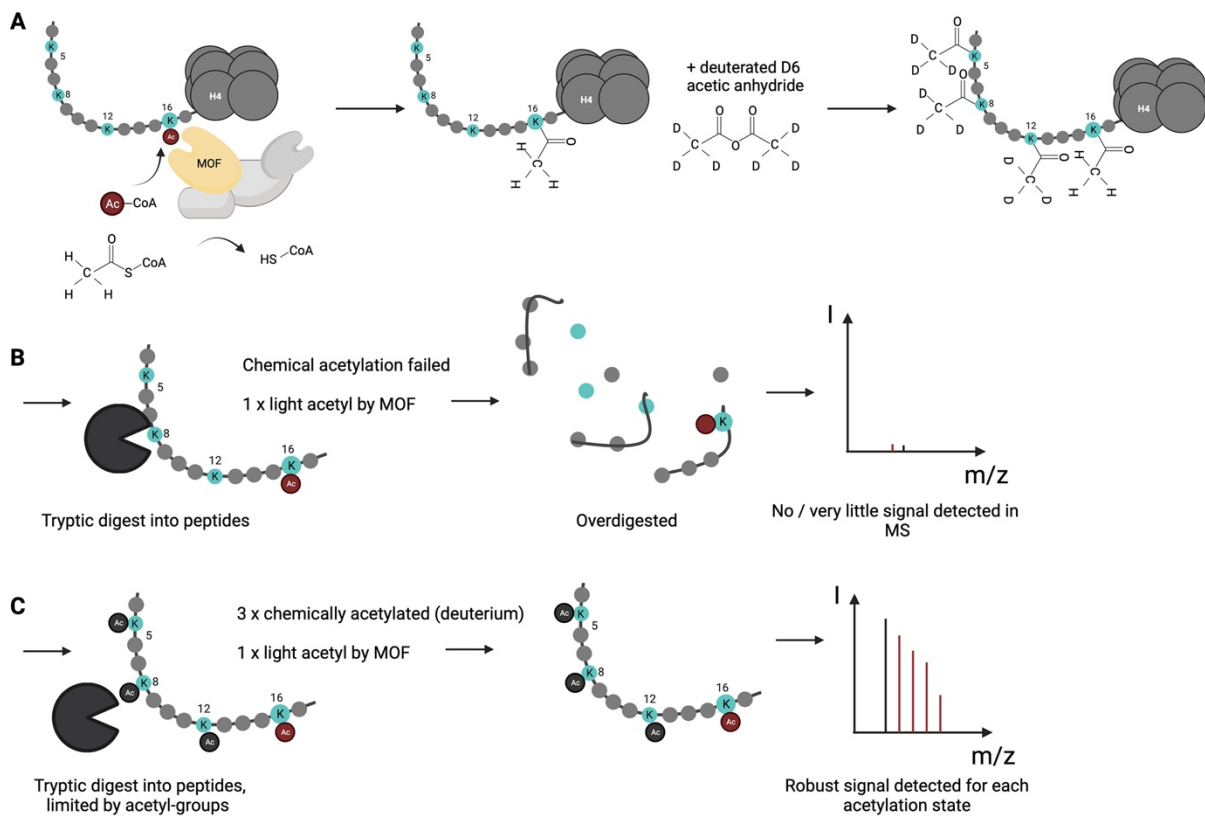
Figure 13: Schematic illustration of the mass spectrometric approach to detect and quantify the acetylated H4 tail peptide. **A** 4-MSL acetylates its nucleosome substrate with light acetyl-CoA. After the incubation time the remaining unmodified lysine residues are acetylated by deuterated acetic anhydride. The difference of 3 Daltons in molecular weight can be determined by mass spectrometry. **B** In case the chemical acetylation is poor, the lysines which are left unmodified by 4-MSL are a target for trypsin, which digests the peptide to very small peptides, which are not readily detected and identified by MS. **C** The deuterated acetyl groups protect the peptide from excessive trypsin digestion and the distinct masses can be readily detected in MS. Figure created with BioRender.

In case of cells, histones can be isolated, in the case of an *in vitro* reaction, the isolation of histones can be omitted. The remaining non-acetylated available lysine residues are acetylated chemically by deuterated acetic anhydride. The chemical acetylation protects peptides which are not enzymatically acetylated from overdigestion by the peptidase (Figure 13B). Due to the use of deuterium instead of

hydrogen, the mass will be 3 Dalton (Da) heavier, which means 45 Da are added to the protein mass (Figure 13A). Next, the completely acetylated protein can be cleaved into peptides, which then can be analyzed by mass spectrometry (Figure 13C). While this small change in mass can be precisely detected by mass spectrometry. Because all peptides are fully acetylated, the chemical properties and thus the behavior during the sample preparation and mass spectrometry run, e.g., ionization, are nearly identical. However, problems might arise if different data sets of different MS measurements need to be compared, due to batch effects of the MS itself, but also from the chemical acetylation and sample preparation steps. If the chemical acetylation fails due to any wrong handling, the only peptides that survive the sample preparation process are the ones which are highly acetylated by the enzyme already (Figure 13B).

All in all, the MS approach offers site-specific analysis of histone modifications: oligo-modifications can be faithfully detected and differentiated from singly modified residues. Moreover, all modifications can be quantified.

3.10 Aims

The overarching aim of this dissertation research is to characterize the role of roX RNA within the MSL complex under structural and functional aspects. This is addressed using XL-MS and HAT-MS assays.

3.10.1 roX2 RNA causes structural changes within the MSL complex

Incorporation of roX RNA is essential for male fly viability. How can roX exert this critical role? The main hypothesis of this first part of the project assumes structural changes of the MSL complex architecture occur upon roX RNA incorporation. To elucidate conformational changes of the MSL complex, the complex has to be reconstituted *in vitro*. The reconstitutions can contain roX2 RNA or be RNA-free. In both cases the structure or at least some structural features, like the interaction interfaces will be analyzed. To achieve this goal, the XL-MS approach is employed to first characterize conformations and interaction interfaces between the MSL proteins. In a further step roX2 RNA and MLE are added to the DNA-binding module of MSL1-MSL2 and the results are compared to the RNA-free complex. Any changes are taken as hint that roX2 RNA impacts the conformation of the MSL complex. These can be further investigated by structural modeling, biochemical immunoprecipitation assays and orthogonal structural methods.

3.10.2 roX2 RNA incorporation impacts the histone acetylation activity of the MSL complex

Independent of conformational changes induced by roX2 incorporation into the MSL complex, its functionality may be affected. Structural changes can have an allosteric effect on the enzymatic activity. The MSL complex has several catalytic subunits, however, I focus on the acetylation activity of MOF, because H4K16ac is the hallmark effect of dosage compensation. The envisioned hypothesis is, that in presence of roX2 RNA, the MSL complex rearranges for better substrate binding and higher MOF activity. H4K16ac would be placed at a higher rate and with increased specificity. Thus, the roX2 RNA could function as an 'on' switch for the acetylation activity.

4 Material and Methods

4.1 Material

4.1.1 Chemicals

Material	Manufacturer (catalog no. if applicable)
Acetic acid, puriss. p.a.	Sigma (33209-2.5L-M)
Acetonitrile, Ultra LC-MS	Carl Roth (HN40.1)
Acetyl-CoA	Sigma (A2056)
Acrylamide (30% and 40%)	Carl Roth (3029.1 and 3030.1)
ATP	Sigma (10127523001)
AMP-PNP adenylyl-imidophosphate	Sigma (10102547001)
Agarose	Bio & Sell (BS20.46.500)
Aprotinin	Genaxxon (M6361.0025)
Ammonium bicarbonate (ambic)	Sigma (09830)
Ammonium persulfate (APS)	Sigma (A3678)
Ampicillin (Amp)	Carl Roth (HP62.2)
Bis(sulfosuccinimidyl) suberate (BS3)	Life technologies (A39266)
Bis(2-hydroxyethyl) amino-tris (hydroxymethyl) methane (Bis-Tris)	Sigma (1.03252250)
Boric acid	Diagonal (131015.1214)
Bovine serum albumin (BSA)	Sigma (A9418)
Bromphenol blue	Sigma
Chloramphenicol (Chl)	Carl Roth (3886.2)
cOmplete protease inhibitor (EDTA-free)	Sigma (5056489001)
Coomassie brilliant blue G250	Serva (17542.02)
deuterated (D6) acetic anhydride	Sigma (175641)
DTT	Häberle Labortechnik (APA29480025)
EDTA	Diagonal (A2937.1000)
EGTA	Carl Roth (3054.3)
Ethanol puriss. p.a.	Sigma (1009832500)
Ethidium bromide (1%)	Carl Roth (2218.2)
Fetal calf serum (FCS)	Sigma (F7524)
FLAG peptide	Medchemexpress (HY-P0223)
Glycerol puriss.	Sigma (15523-2.5L-M)

Material and Methods

Glycine	Sigma (G8790)
Glycogen	Carl Roth (HP51.1)
Guanidine hydrochloride	Sigma (G4505)
HEPES	Serva (25245.05)
IGEPAL CA 630 (NP-40)	Sigma (I3021)
IPTG	Diagonal (A1008.0025)
KCl p.a.	Diagonal (131494.1211)
Leupeptin	Genaxxon (M6100.0025)
2-Mercaptoethanol	Sigma (M6250)
Methanol puriss. p.a.	Sigma (32213-2.5L-M)
MES (2-(N-morpholino) ethanesulfonic acid)	Cymit Quimica (02-H56472)
MgCl$_2$ Magnesiumchloride	Neolab (LC-5041.2)
Midori green advance	Nippon Genetics Europe (MG04)
NaCl	Serva (30183.02)
Native PAGE running buffer	Thermo Fisher Scientific (BN2001)
Nuclease-free water	Invitrogen
Orange-G	Sigma
Penicillin-streptavidin	Sigma (P4333)
Pepstatin	Genaxxon (M6359.0025)
Poly-deoxy-adenylic acid (poly-dA)	Merck (10223581001)
phenylmethylsulfonylfluoride PMSF	Carl Roth (6367.2)
phenol-chloroform-isoamylalcohol	Carl Roth (A156.3)
2-propanol	Sigma (34863-2.5L)
RNase away	Carl Roth (A998.3)
RNasin	Promega (N2515)
Schneider's *Drosophila* medium	Thermo Fisher Scientific (21720001)
Sepharose protein A and G	Helmholtz Zentrum Munich
Sodium acetate	Sigma (567422)
Sodium azide	Sigma (S2002-25g)
Sodium dodecyl sulfate (SDS)	Serva (20765.03)
Sodium metabisulfite NMBS	VWR
TEMED	Carl Roth (2367.3)
TRIS ultrapure	Diagonal (A1086.1000)
Trifluoro acetic acid TFA	Sigma
Triton-X-100	Sigma (T8787)
Trypsin Sequencing grade	Promega (V5111)

Tween-20	Sigma (P7949)
tRNA	Thermo Fisher Scientific (AM7119)
Urea	Diagonal (A1049.1000)
Zinc chloride	Sigma (208086)
Zinc sulfate	Carl Roth

4.1.2 Consumables

Material	Manufacturer (catalog no. if applicable)
15-mL tubes (Sarstedt),	Greiner (188271-N)
Amicon Ultra-4 centrifugal filter	Merck Millipore (UFC803024)
Anti-FLAG affinity gel	Sigma (A2220)
Anti-DYKDDDDK G1 Affinity Resin	GenScript (L00432-25)
Cassettes for acrylamide gels	Thermo Fisher Scientific (NC2010)
Cell culture flasks	Greiner (658170)
Deep-well plates 96-wells	Carl Roth (EN07.1)
Dialysis cups Slide-A-Lyzer	Life technologies (69552)
Dialysis membranes (Spectra/Por),	Carl Roth (0978.1)
Eppendorf reaction tubes 1.5 mL	Eppendorf (0030108116 and 0030108051)
50-mL falcon tubes	Corning (352070)
0.45 μm filters	Carl Roth (CY06.1)
Glass pipettes 5 mL and 10 mL	Hirschmann
Glassware	Schott
Gradient 3-12% (V/V) acrylamide Bis-Tris gels	Thermo Fisher Scientific (BN1001BOX)
Laboratory film (Parafilm M)	Häberle Labortechnik (9.170 002)
Nitrocellulose membrane	Sigma (GE10600002)
Polyacrylamide gels (7.5% and 4-20%),	Biorad
Polyacrylamide gels (8% and 4-20%),	Serva
Q HiTrap HP 5 mL	VWR/ Cytiva (17-1154-01)
SDB-RPS material 3 M	Empore
SP HiTrap HP 5 mL	VWR/cytiva (17-1152-01)
STAGE tips Attract SPE Disks Tips C18	Affinisep (C18.T2.200.96)
StrepTrap Streptactin column	Sigma/ cytiva (GE29401317)
Superdex 200 HiLoad chromatography column	Cytiva (289889335)
Syringes	Carl Roth / Braun
Whatman paper	Merck (588-3148)

Material and Methods

4.1.3 Instruments

Instrument	Manufacturer (Catalog No. if applicable)
Äkta pure	Cytiva (29018224)
Äkta purifier	Cytiva
Äkta Superloop 50 mL	Cytiva/VWR (18-1113-82)
Avanti JXN-26 centrifuge	Beckman Coulter
Branson Sonicator	Branson/Thermo Scientific
CASY cell counter	OMNI life science
Gel documentation for acrylamide gels	BioRad
Gel documentration for agarose gels	Peqlab
Incubators	Infors
Licor Western blot scanner	LiCOR
Lyophilizer freeze dryer	Christ
Microfluidizer	Microfluidics corporation
Nanodrop	Thermo Fisher Scientific
Peristaltic pump Minipulse evolution	Gilson
PCR PROFLEX 3x32-well PCR System	Life technologies (A28986)
Pipetboy	Integra biosciences (155000)
Pipetman	Gilson/ Neolab (CF-0415)
Power supply	BioRad (1645050)
Qubit fluorometer	Thermo Fisher Scientific
QexactiveHF	Thermo Fisher Scientific
RSLCnano Ultimate 3000 system	Thermo
Rotation wheel	VWR
Speed vacuum and centrifuge system	Eppendorf
Table-top centrifuge 5424 R	Eppendorf (5424000410)
Thermomixer	Eppendorf (5360000038 and 5308000003)
Trans-Blot Turbo Blotting System (semi-dry)	BioRad (170-4155)
Typhoon imaging system	GE Healthcare
Ultrasound bath H40	Carl Roth (CPK7.1)

4.1.4 Kits and enzymes

Material	Manufacturer (catalog number if applies)
100 bp and 1 kB DNA marker	NEB
Benzonase	Merck Millipore (71205-3)

Material and Methods

MEGAscript T7 transcription kit	Thermo Fisher Scientific (AM1334)
MEGAshort script T7 transcription kit	Life technologies (AM1354)
Maxi-prep kit	Macherey-Nagel (740414.10)
Midi-prep kit	Macherey-Nagel (740410.50)
MNase (Sigma),	Sigma (N5386)
NucleoSpin Plasmid Mini Kit	Macherey-Nagel (740499.250)
NucleoSpin Gel and PCR clean up kit	Macherey-Nagel (740611.50)
Proteinase K	BioCat (BIO-37039-BL)
Restriction enzymes	NEB
RNase A	Sigma (R4875)
Qubit dsDNA HS assay kit	Thermo Fisher Scientific (Q32854)
QuikChange site directed mutagenesis kit	Agilent (200523)
Taq DNA Polymerase	Biolabs (M0267S)
Triple color protein standard III (Serva),	Serva (39258.01)
Trypsin/Lys-C Mix Mass Spec Grade	Promega (V5073)
Turbo DNase I	Thermo Fisher Scientific (AM2238)

4.1.5 Cell lines

Strain	Origin	Source
Sf21	*Spodoptera frugiperda* embryonic cells	Gibco

4.1.6 Bacterial strains

Strain	Genotype	Source
E. coli DH5a	*fhuA2 (argF-lacZ) U169 phoA gln V44 80 (lacZ) M15 gyrA96 recA1 relA1 thi-1 hsdR17*	NEB Cat.-No c2987
E. coli BL21-Star (DE3)	*fhuA2 [lon] ompT gal [dcm] deltahsdS*	Agilent, Cat No. 230132

4.1.7 Buffers and solutions

Name of buffer	Composition
Agarose gel	TBE buffer, 1% (w/V) agarose, ethidium bromide (1:25000) / alternatively Midori green (1:25000)

Material and Methods

Ampicillin stock solution	100 mg/mL ampicillin (1000 X)
BC-0 (buffer C, no salt)	20 mM HEPES pH 7.9, 50 µM $ZnCl_2$, 10% (V/V) glycerol, 0.2 mM PMSF, 1 mM DTT
BCM-100 (buffer C, 100 mM salt, Mg)	20 mM HEPES pH 7.9, 100 mM KCl, 50 µM $ZnCl_2$, 3 mM $MgCl_2$, 10% (V/V) glycerol, 0.2 mM PMSF, 1 mM DTT
BC-200 (buffer C, 200 mM salt)	20 mM HEPES pH 7.9, 200 mM KCl, 50 µM $ZnCl_2$, 10% (V/V) glycerol, 0.2 mM PMSF, 1 mM DTT
Biotin elution buffer	50 mM biotin pH 8.0, 25 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 3 mM $MgCl_2$, 1 mM EDTA
Bis-Tris buffer (5x)	2.5 M Bis-Tris, 1.5 M HCl, pH 6.5 at 4°C
Bis-Tris gel 15% (V/V) acrylamide	15% (V/V) acrylamide, 1x Bis-Tris buffer, 0.05% (V/V) APS, 0.05% (V/V) TEMED
Bis-Tris gel 4% (V/V) acrylamide	4% (V/V) acrylamide, 1x Bis-Tris buffer, 0.05% (V/V) APS, 0.05% (V/V) TEMED
Blocking solution	3% (w/V) BSA in TBS
Chloramphenicol stock solution	34 mg/mL chloramphenicol in ethanol
Coomassie staining solution in fixing	50% (V/V) ethanol, 10% (V/V) acetic acid, 0.0025% (w/V) Coomassie brilliant blue G250
EMSA buffer	25 mM HEPES at pH 7.9, 60 mM KCl, 7% (V/V) glycerol, 0.5 mM DTT
HAT assay buffer (HABM-50)	50 mM HEPES pH 7.9, 3 mM $MgCl_2$, 50 mM KCl
High salt buffer	10 mM Tris-HCl pH 7.6, 2 M NaCl, 1 mM EDTA pH 8.0, 0.05% (V/V) Igepal CA-630, 0.1% (V/V) beta-mercaptoethanol
Laemmli buffer 5 x	250 mM Tris-HCl pH 6.8, 50%(V/V) glycerol, 10% (w/V) SDS, 0.05% (w/V) bromophenol blue, 0.5 M DTT
Low salt buffer	10 mM Tric HCl pH 7.6, 50 mM NaCl, 1 mM EDTA, 0.05% (V/V) Igepal CA-630, 0.01% (V/V) beta-mercaptoethanol
Lysis buffer MLE	25 mM HEPES pH 7.5, 300 mM NaCl, 2 mM $MgCl_2$, 5% (V/V) glycerol, 0.1% (V/V) Igepal CA-630, 1 mM DTT

Material and Methods

Lysis buffer dTIP60$^{\text{piccolo}}$	25 mM HEPES pH 7.5, 300 mM KCl, 5% glycerol, 0.05% NP-40, 3 mM MgCl$_2$, cOmplete EDTA-free protease inhibitor, 200 µg RNase A, 4 µL benzonase
MLE high salt wash buffer	25 mM HEPES pH 7.5, 1000 mM NaCl, 2 mM MgCl$_2$, 5% (V/V) glycerol, 1 mM DTT
MLE normal salt buffer	25 mM HEPES pH 7.5, 200 mM NaCl, 2 mM MgCl$_2$, 0.5 mM DTT, 5% (V/V) glycerol
MNase buffer	20 mM HEPES pH 7.5, 50 mM NaCl, 3 mM MgCl$_2$, 2.5 mM DTT, 0.5 mM EGTA, 1.5 mM CaCl$_2$)
MS sample buffer	0.3% (V/V) TFA, 2% (V/V) ACN in MS grade H$_2$O
MXB-50 (MLE XL buffer 50 mM salt)	20 mM HEPES pH 7.5, 50 mM NaCl, 2 mM MgCl$_2$, 1 mM DTT
NE-A (nuclear extraction buffer A)	10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.5 mM PMSF, 1 mM DTT
NE-B (nuclear extraction buffer B)	20 mM HEPES pH 7.9, 400 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.5 mM PMSF, 1 mM DTT
Orange-G loading dye 6x	60% (V/V) glycerol, 40% (V/V) TE buffer, 2 mg/mL Orange-G
PBS(-T) buffer 10x	1.4 M NaCl, 27 mM KCl, 100 mM Na$_2$HPO$_4$, 18 mM KH$_2$PO$_4$, (0.1% (V/V) Tween-20)
PBS 1x	140 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 1.8 mM KH$_2$PO$_4$ pH 7.4
Refolding buffer	2 M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT
Running buffer MES (20x)	1 M MES, 1 M Tris-base, 20 mM EDTA pH 8.0, 2% (w/V) SDS
SAU-200 buffer (Sodium acetate urea buffer with 200 mM salt)	40 mM NaOAc at pH 5.2, 1 mM EDTA at pH 8, 10 mM Lysine, 200 mM NaCl, 7.5 M urea, 5 mM DTT, 1 µg/mL of each Pepstatin, Leupeptin, and Aprotinin, 1 mM PMSF, 5 mM ß-mercapto-ethanol
SDS-PAGE running buffer	25 mM Tris, 192 mM glycine, 0.1% (w/V) SDS
Stop buffer (MNase digestion assay)	10 mM EDTA, 2% (w/V) SDS

StrepTrap wash buffer 200	25 mM HEPES pH 7.5, 5% glycerol, 3 mM MgCl$_2$, 200 mM KCl
Transfer buffer	25 mM Tris, 192 mM glycine, 20% (V/V) methanol
TBE buffer	100 mM Tris base, 100 mM boric acid, 2 mM EDTA
TBS(-T) 10x	250 mM Tris-HCl pH 8.0 30 mM KCl, 1.4 M NaCl, (1%(V/V) Tween-20)
TE buffer	10 mM Tris pH 8.0, 1 mM EDTA
Trypsin-LysC Mix for MS	trypsin (10 ng/μL), Lys-C (10 ng/μL), 1 M urea, 50 mM ammonium bicarbonate pH 8.0 (NH$_4$HCO$_3$)
Unfolding buffer	7 M guanidine chloride, 20 mM Tris pH 7.5, 10 mM DTT
5% urea PAGE gel	5% (V/V) Bis-acrylamide, 8 M urea, 0.5% (V/V) APS, 0.05% (V/V) TEMED, TBE buffer
Urea sample buffer USB	9 M Urea, 25 mM Tris pH 6.8, 1% (w/V) SDS, 1 mM EDTA, 100 mM DTT
Wash buffer 200 dTIP60^{piccolo}	25 mM HEPES pH 7.5, 5% (V/V) glycerol, 3 mM MgCl$_2$, 200 mM KCl
XL-buffer	20 mM HEPES pH 8.2, 100 mM KCl, 50 μM ZnCl$_2$, 5% (V/V) glycerol

4.1.8 Antibodies

Name	Species	Type	Application	Source
Anti-MSL1	Rabbit	Polyclonal	Western blot 1:2000	[64, 207]
Anti-MSL2	Guinnea pig	Polyclonal	Western blot 1:2000	[67]
Anti-MLE	Rat	Monoclonal 6E11	WB 1:500	[138]
Anti-MSL3	Rat	Monoclonal 1C9	WB 1:50	[57, 208]
Anti-MOF	Rabbit	Polyclonal SA4897	WB 1:2000	[95, 209]
Anti-H4K16ac	Rabbit	Polyclonal	WB 1:1000	Merck Millipore Cat. No. 07-329
Anti-H4K12ac	Rabbit	Polyclonal	1:5000	Merck Millipore Cat. No. 07-595
Anti-H3	Rabbit	Polyclonal	1:10,000	Abcam ab1791

Material and Methods

Anti-H3	Mouse	Monoclonal	1:5000	Abcam ab10799
Anti-gp, mo, rb, rt	Donkey	polyclonal	1:10,000 secondary antibody IRDye 680RD	Li-Cor Biosciences
Anti-gp, mo, rb, rt	Donkey	polyclonal	1:10,000 secondary antibody IRDye 800CW	Li-Cor Biosciences

4.1.9   Primers and oligonucleotides

Primer	Sequence (5'-3')	Purpose
dH4_K16R fw	GGGTGGCGCCcgtCGTCATCGCA	Cloning of the H4K16R mutant
dH4_K16R rv	TTTCCCAAGCCTTTGCCTC	Cloning of the H4K16R mutant
pACEBac1_F	TCTAGAGCCTGCAGTCTCG	biGBac cloning of dPiccolo
pACEBac1_R	ATATTTATAGGTTTTTTTATTACAAAACTG	biGBac cloning of dPiccolo
TwinStrep-Tip60_F	taataaaaaaacctataaatatGAGCGCATGGAGTCATCC	biGBac cloning of dPiccolo
TwinStrep-Tip60_R	tcgagactgcaggctctagaTCATTTGGAGCGCTTGGAC	biGBac cloning of dPiccolo
Ing3_PH_f	tcccggtccgaagcgcgcggaattcATGCTTTACCTCGAAGAC	biGBac cloning of dPiccolo
Ing3_PH_r	tcctctagtacttctcgacaagcttTCAGTTCTTTCGGTTGCC	biGBac cloning of dPiccolo
E(Pc)_PH_f	tcccggtccgaagcgcgcggaattcATGTCCAAGCTGTCGTTC	biGBac cloning of dPiccolo
E(Pc)_PH_r	tcctctagtacttctcgacaagcttTCATCTGTTGATGGTTGAC	biGBac cloning of dPiccolo
0N80.fw	CCTGGAGAATCCCGGTGCCGAG	Mononucleosome 0N80 forward primer
0N80_6FAM.rv	GGTACCCGGGGATCCTCTAGA	Mononucleosome 0N80 reverse primer
GFP-fw76ntRNA	CACGCCGTAGGTCAGGGTGGTC	PCR primer GFP-76 template for transcription
GFP-fw182ntRNA	CGTCGTCCTTGAAGAAGATGGTGCGCTC	PCR primer GFP-182 template for transcription
GFP-fw283ntRNA	CAGGATGTTGCCGTCCTCCTTGAAGTCGA	PCR primer GFP-283 template for transcription
GFP-fw391ntRNA	GTCCTCGATGTTGTGGCGGATCTTGAAGTTCA	PCR primer GFP-391 template for transcription
GFP-rv470ntRNA	CCTGAAGTTCATCTGCACCA	PCR primer GFP-470 template for transcription together with the fw GFP-RNAi primer
GFP-fw550ntRNA	GAGCTGGACGGCGACGTAAAC	PCR primer GFP-550 template for transcription
GFP-fw627ntRNA	CCACCGGTCGCCACCATG	PCR primer GFP-627 template for transcription
rvGFP-RNAi-primer with T7 promotor	TTAATACGACTCACTATAGGGCCTGAAGTTCATCTGCACCA	Reverse primer for the PCR of the GFP transcription templates together with the GFP fw primers listed
fwGFP-RNAi	TTAATACGACTCACTATAGGGTGCTGAGGTAGTGGTTGTCG	T7 fwGFP-RNAi primer together with the 470nt rv primer

MLE ΔRB1-all rv	CATGGATCCGCGCCCGAT	Cloning of MLE ΔRB1
MLE ΔRB1-70 fw	GGCAAACTGAACACCAATGATGTTC	Cloning of MLE ΔRB1-70
MLE ΔRB1-84 fw	TCGGGGGGCGGACCTAGG	Cloning of MLE ΔRB1-84
MLE ΔRB1-86 fw	GGCGGACCTAGGACTGGTC	Cloning of MLE ΔRB1-86
MSL2 Δpre-CXC fw	TGCCGCTGCGGTATCTCC	Cloning MSL2 Δpre-CXC in pFastBac
MSL2 Δpre-CXC rv	CTCGTCACTGTCCTCCTTCTG	Cloning MSL2 Δpre-CXC, 54 aa, in pFastBac
MSL2 ΔCC-pre-CXC rv	ATGACCCAATTGTGTTTTTGTGGC	Cloning MSL2 ΔCC-pre-CXC, 100 aa, in pFastBac

4.1.10 Plasmids

All plasmids cloned in this study have been Sanger sequenced at Azenta Life Sciences (Leipzig, Germany) and controlled for sequence integrity manually by Mac vector sequence analysis tool (version 18.5.1).

Name	Source	Description	Resistance
pFastBac-MSL2-ΔpreCXC-FLAG	This study	For mutant MSL2 expression with a 54 aa deletion N-terminal of the CXC domain, aa 470-524	Amp
pFastBac-MSL2-ΔCCpreCXC-FLAG	This study	For mutant MSL2 expression with 100 aa deletion N-terminal of the CXC domain, aa 424-524	Amp
pFastBac-MLE-ΔRB1-70-FLAG	This study	For mutant MLE expression with a 69 aa deletion of RB1, aa 2-70	Amp
pFastBac-MLE-ΔRB1-84-FLAG	This study	For mutant MLE expression with 83 aa deletion of RB1, aa 2-84	Amp
pFastBac-MLE-ΔRB1-86-FLAG	This study	For mutant MLE expression with 85 aa deletion of RB1, aa 2-86	Amp
pET3c-H4K16R	This study	Codon optimized for *E. coli* expression of mutant *D. melanogaster* H4 protein (untagged)	Amp

4.1.11 RNA oligonucleotides

RNA Name	Sequence 5'-3'	Source
GFP-76	CCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCA CCCUGACCUACGGCGUG	*In vitro* transcription from published plasmid [52]
GFP-182	CCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCA CCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGAC UUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGA CGACG	*In vitro* transcription from published plasmid [52]

GFP-283	CCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCA CCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGAC UUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGA CGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACC GCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUG	*In vitro transcription from published plasmid* [52]
GFP-391	CCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCA CCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGAC UUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGA CGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACC GCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGGCACAAGCUG GAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAU CAAGGUGAACUUCAAGAUCCGCCACAACAUCGAGGAC	*In vitro transcription from published plasmid* [52]
GFP-470	CCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCA CCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGAC UUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGA CGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACC GCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGGCACAAGCUG GAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAU CAAGGUGAACUUCAAGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACC ACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACU	*In vitro transcription from published plasmid* [52]
GFP-550	GAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGA UGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGC CCUGGCCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCC GACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGA GCGCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCG AGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGC AACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGC CGACAAGCAGAAGAACGGCAUCAAGGUGAACUUCAAGAUCCGCCACAACAUCGAGGACG GCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUG CUGCUGCCCGACAACCACU	*In vitro transcription from published plasmid* [52]
GFP-627	CCACCGGUCGCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAU CCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCG AGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUG CCCGUGCCCUGGCCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCG CUACCCCGACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUACG UCCAGGAGCGCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUG AAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUCAAGGA GGACGGCAACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUA UCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUCAAGAUCCGCCACAACAUC GAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGG CCCCGUGCUGCUGCCCGACAACCACU	*In vitro transcription from published plasmid* [52]
roX2 RNA	[137]	[137]
roX2-123	[137]	[137]
roX2-SL678	[118]	[118]
U$_{10}$	[118]	[118]

4.2 Methods

4.2.1 Cloning and protein expression

Cloning of *Drosophila* sequences for expression of MSL proteins and the respective complexes (MSL1, MSL2, MOF, MLE, 2-MSL, 3-MSL and 4-MSL) was described earlier [63, 64, 137, 143].

The 3-subunit dTip60[piccolo] complex was cloned and expressed by Mrs Silke Krause using the biGBac technology [210]. The cDNAs of full-length *Drosophila melanogaster* E(Pc), Ing3 and dTip60 fused to an N-terminal TwinStrep tag were combined in one pBIG1 vector. Primers are given in Chapter 0. The

Material and Methods

bacmid was transfected into *Spodoptera frugiperda* (Sf21) cells to produce one dTip60^{piccolo} -expressing baculovirus.

Drosophila 4-MSL complex, MLE and the dTip60^{piccolo} complex were expressed in Sf21 cells infected with 1/1000 (V/V) baculovirus for 72 h at 26°C. The cells were harvested by centrifugation, rinsed once with PBS buffer, flash frozen in liquid nitrogen, and preserved at -70°C.

4.2.2 Purification of proteins and protein complexes

Individual MSL proteins (MSL1, MSL2, MOF) were purified as previously described [63, 64]. MLE was purified as published [52].

The MSL complexes (2-MSL, 3-MSl and 4-MSL) were purified from isolated nuclei of 5×10^8 baculovirus-infected Sf21 cells as described [143] with minor modifications (all steps at 4°C). The nuclei were dissolved in 5 mL of NE-B buffer and allowed to incubate for 15 minutes with head-over-end rotation. The lysate was diluted with 1 volume of PBS supplemented with 1 mM DTT, 50 μM ZnSO$_4$ and 0.2 mM PMSF, and centrifuged for 10 min at 17,000 g. The resulting supernatant was subjected to incubation with 1.5 mL of a 50% suspension of FLAG-affinity beads (Anti-DYKDDDDK G1 Affinity Resin) for 1 h with head-over-end rotation, in the presence of 400 μg of RNase A. Beads were washed three times (PBS with 1 mM DTT, 50 μM ZnSO$_4$); PBS with additional 200 mM KCl; PBS only). The protein complex was eluted in three consecutive steps, each comprising 1 bead volume of elution buffer (consisting of PBS and 0.5 mg/mL FLAG peptide), with each step for 30 minutes. The combined elution fractions were supplemented with 10% (V/V) glycerol, flash frozen in liquid nitrogen, and preserved at -70°C. Protein concentration was assessed using SDS-PAGE and Coomassie staining, employing ImageLab software (version 6.0, Bio-Rad), with a BSA standard kit as reference.

FLAG-tagged MLE was expressed in Sf21 insect cells and purified by FLAG-affinity chromatography as described [137].

The dTip60^{piccolo} complex was purified from 2.5×10^8 baculovirus-infected Sf21 cells (all steps at 4°C). The cell pellet was resuspended in 25 mL lysis buffer dTip60^{piccolo}. The suspension was allowed to incubate for 15 minutes on ice. After sonication (performed using a Branson sonifier at 20% amplitude, 60 s), cellular debris was eliminated by centrifugation at 30,000 g for 30 minutes. The resulting supernatant was then transferred to a new tube and filtered twice through a Millex HPF filter with a pore size of 0.45 μm. The filtered cell lysate was loaded onto a 2 mL StrepTrap Streptactin column, equilibrated with wash buffer 200. The column was washed with 10 CV wash buffer 300 (300 mM KCl). Proteins were eluted in 10 CV biotin elution buffer and analyzed by SDS-PAGE and spectrophotometry at 280 nm. dTIP60^{piccolo}-containing fractions were pooled, concentrated using a 50,000 Da cut-off concentrator, flash-frozen in liquid nitrogen and stored at -70°C. Protein concentration was determined by SDS-PAGE and Coomassie staining using ImageLab (Bio-Rad, version 6.0) and BSA standards as reference.

The histone H4K16R mutation was introduced into codon-optimized *Drosophila melanogaster* H4 expression plasmid in a pET3c vector [211] using the QuikChange site directed mutagenesis kit and

primers as in Chapter 4.1.9. *Drosophila melanogaster* histones (wild-type and H4K16R mutant) were expressed and purified as described [211].

4.2.3 Cell lines

For amplification of recombinant baculoviruses and expression of recombinant proteins, Sf21 cells were cultured at 26°C in Sf-900 II medium supplemented with 5% FCS and gentamycin (1:100). Amplification of baculoviruses of all expression constructs and expression of dTIP60^{piccolo} complexes was performed by Silke Krause.

4.2.4 Crosslinking of MLE and RNA (in solution)

Protein and RNA constructs of interest were incubated in MLE crosslinking buffer MXB-50 in presence of the ATP-analogon adenylyl-imidodiphosphate (AMP-PNP 1 mM) and RNase inhibitor RNAsin (0.5 U) for 25 min at 4°C with head over end rotation. In the presence of 0.7 μM MLE, 1.4 μM RNA was utilized, yielding a molar ratio of 2-fold RNA over protein. Bis(sulfosuccinimidyl)suberate BS3 was dissolved in MXB-50, yielding a stock of 50 mM concentration. Of the crosslinker stock 1 mM was added to the protein-RNA samples and incubated for 30 min at 30°C and 950 rpm. The crosslinking reaction was quenched by addition of 50 mM Tris-HCl pH 8.0 and incubation for 15 min at 30°C and 950 rpm. A sample for Western blot analysis was taken to visualize the crosslinking degree of proteins of interest. Before tryptic digestion, 1 M urea was added to the in-solution crosslinked samples to allow partial unfolding of the protein. Trypsin Lys-C Mix for MS was added in a 1:50 ratio to the protein mass. 1 mM DTT was added and the tryptic digestion was carried out overnight at 37°C with 500 rpm agitation.

4.2.5 Crosslinking of MSL complexes for MS (on beads)

MSL complex proteins were purified from Sf21 nuclei as described in Chapter 4.2. However, proteins were not eluted. Instead, the buffer of the last washing step was discarded after centrifugation at 500 g for 2 min and 4°C. The beads were split into the desired number of samples. Per sample 50 μL of bead material were used. 50 μL of BC-100 including the desired concentration of BS3 were added to the beads and incubated for 30 min at 25°C with 1100 rpm shaking. The crosslinking reaction was stopped by 200 μL of 50 mM Tris-HCl pH 8.0, which has primary amine groups that react with the reactive groups of BS3. Quenching was performed for 15 min at 25°C and 950 rpm agitation. The beads were settled by centrifugation of 500 g for 2 min and the supernatant was discarded. The beads were washed three times with 100 μL of 50 mM Tris-HCl pH 8.0. After the washing steps, 100 μL of Trypsin-LysC mix for MS was added to the beads and the proteins were digested for 30 min at 25°C and 1400 rpm. After tryptic digestion, the supernatant was transferred into a fresh rection tube. Afterwards the beads were washed twice with 40 μL of 50 mM ammonium bicarbonate solution and per sample all supernatants were pooled in one reaction tube. 1 mM DTT was added to each sample and the tryptic digestion was incubated overnight at 25°C and 500 rpm. After 16 h of incubation, iodoacetamide was

added to a final concentration of 4 mM and the samples were incubated for another 30 min in the dark at 25°C and 500 rpm. Then DTT was supplied to 20 mM final concentration. After 10 min incubation, 3% (V/V) trifluoroacetic acid (TFA) was added. The sample preparation was continued as described in Chapter 4.2.6.

4.2.6 Mass spectrometry sample preparation for XL-MS

In-solution and on-bead crosslinked mass spectrometry samples were treated the same way. 4 mM iodoacetamide were added and the incubation was continued for 30 min at 25°C, 500 rpm in the dark. Iodoacetamide was quenched by the addition of 20 mM DTT and incubation for 10 min at 25°C and 500 rpm. Samples were acidified by the addition of 0.05% trifluoracetic acid (TFA), the pH was adjusted to pH 1. SDB-RPS stage tip material was equilibrated by washes with 100% acetonitrile (ACN), activation buffer (30% methanol, 0.2% TFA) and equilibration buffer (0.2% TFA). Trypsinized samples were loaded to the equilibrated stage tips and centrifuged for 15 min at 500 g. After subsequent washes with wash buffer (1% TFA in 100% isopropanol) and equilibration buffer, peptides were eluted into low protein binding Eppendorf reaction tubes with freshly prepared elution buffer (80% acetonitrile, 1.25% ammonia). After vacuum drying of the samples at 45°C, peptides were resuspended in MS sample buffer (0.3% TFA, 2% ACN in MS grade H_2O).

4.2.7 Mass spectrometry set up and conditions

Mass spectrometry was performed by Dr. Ignasi Forné (ZfP at LMU, Munich) as previously described [170]. Briefly, for LC-MS/MS analysis samples were injected in an RSLCnano Ultimate 3000 system and either separated in a 15-cm analytical column (75 μm ID home-packed with ReproSil-Pur C18-AQ 2.4 μm from Dr. Maisch) with a 50-min gradient from 5 to 60% acetonitrile in 0.1% formic acid or in a 25-cm analytical column (75 μm ID, 1.6 μm C18, IonOpticks) with a 50-min gradient from 2 to 37% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into a QexactiveHF operated in data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 375–1600) were acquired with resolution R=60,000 at m/z 400 (AGC target of $3x10^6$). The ten most intense peptide ions with charge states between 3 and 5 were sequentially isolated to a target value of $1x10^5$, and fragmented at 27% normalized collision energy. Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C; ion selection threshold, 33.000 counts.

4.2.8 Crosslinking data analysis

The raw data files were first converted by the proteome discoverer 2.2 (Thermo scientific) xlinkx workflow for crosslink detection into the .mgf file format. Secondly, the .mgf files were analyzed by crossfinder [169-171] with the filtering parameters for identification of cross-linking candidates: False-discovery rate (FDR) ≤0.05, number of fragment ions per spectrum ≥4, number of fragment ions per peptide ≥2, fractional intensity of assigned MS2 peaks ≥0.05, relative filter score: 95. Thirdly, after

reformatting the .txt output file into .csv file crosslinks can be visualized using the xvis web browser for arch plots (script in Appendix, [212]) or in the xiNET viewer (https://crosslinkviewer.org/, 02.02.2024). The raw data and .mgf files for the MLE and RNA project [118] can be found in the PRIDE database under the identifier PXD045725 [118, 213].

4.2.9 Sample preparation for mass spectrometry for the analysis of histone acetylation patterns

Samples for mass spectrometry were prepared as described [201]. Briefly, heat-denatured HAT assay samples were chemically acetylated using 25% fresh (V/V) acetic anhydride-D6 for 1 min at RT. The mass difference of three Daltons allows distinguishing chemical and enzymatic acetylation. Light and heavy isotope-labeled acetyl groups have the same chemical properties allowing for reliable MS quantification [201]. The pH was adjusted stepwise to pH 7.0 by addition of 1 M ammonium bicarbonate. D6-acetylation continued for 45 min at 37°C with 500 rpm agitation. Proteins were digested with 1 µg of trypsin for 16 h at 37°C with 500 rpm agitation. C18 STAGE tips were prepared for desalting as described [214]. They were washed with 60 µL each of (1)100% acetonitrile, (2) 0.1% (V/V) trifluoroacetic acid, 80% (V/V) acetonitrile in MS-grade water and (3) 0.1% (V/V) trifluoroacetic acid in MS-grade water before samples were loaded to the C18 tips. Liquid was passed through by centrifugation at 300-400 g for 3 min at RT. Trypsin digested samples were loaded to the C18 STAGE tips. The flow through was loaded once again. Bound peptides were washed three times with 0.1% (V/V) trifluoroacetic acid in MS-grade water. Peptides were eluted in three steps with 0.25% (V/V) trifluoroacetic acid, 80% (V/V) acetonitrile in MS-grade water. The elutions were pooled corresponding to the sample and dried under vacuum for 1 h at 45°C. Samples were resuspended in 17 µL MS buffer (0.1% (V/V) trifluoroacetic acid in MS-grade water), sonicated for 5 min in an ultrasound bath and stored at -20°C.

4.2.10 Mass spectrometry analysis of histone modifications

Mass spectrometry was performed by Dr. Ignasi Forné (ZfP at LMU, Munich). Samples were separated in an Ultimate 3000 RSLCnano system using a 25-cm Aurora column (Ionopticks) with a 50-min gradient from 2 to 37% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into a Qexactive HF. The MS instrument was programmed to target several ions as described previously [201] except for the MS3 fragmentation. For the analysis of the H4K16R mutant, the m/z values of the corresponding peptide ions [M+H2]2+ were adjusted accordingly (H4G4R17 monoacetylated adjusted from 724.9428 to H4G4R16 637.8759, diacetylated from 723.4329 to 636.3665 and triacetylated from 721.9221 to 634.8571). Survey full scan MS spectra (from m/z 270-730) were acquired with resolution R=60,000 at m/z 400 (AGC target of $3x10^6$). Targeted ions were isolated with an isolation window of 0.7 m/z to a target value of $2x10^5$ and fragmented at 27%

normalized collision energy. Typical mass spectrometric conditions were: spray voltage 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C.

4.2.11 Data analysis of MS data post-translational modification of histones

Raw data from mass spectrometry was analyzed using Skyline by Anuroop V. Venkatasubramani [215], v21.1. Peak integration was performed for H4 peptides for each of its corresponding modifications. Relative levels of each PTM were calculated from the obtained intensities using R, based on the formula given in [201]. Formula to calculate the relative levels of K16R mutated H4 motifs are published [216]. MS1 spectra of acetylated wild type and H4K16R mutant peptides were generated using Xcalibur 4.4 using the dataset of reference number Ref. 8241, replicate 3 [216]. The raw and processed HAT assay MS data can be accessed in the PRIDE database under the identifier PXD046636 [213].

4.2.12 Mathematical modeling

Mathematical modeling was performed by Dr. Dilan Pathirana and Prof. Dr. Jan Hasenauer (University of Bonn) as published [216]. All scripts are provided in the Zenodo repository: Zenodo (DOI: 10.5281/zenodo.10221453).

4.2.13 Immunoprecipitation assay for validation of interaction domains

Protein A/G beads were mixed 1:1 and equilibrated with BCM-100 buffer. Anti-MSL1 antibody (rb) was coupled to the protein A/G beads at 4°C for 3 h with head-over-end rotation, including 2% (w/V) BSA and 0.1 mg/mL yeast tRNA to block unspecific binding. The beads were washed three times with BCM-100 after the specific antibody was bound. MSL1, MSL2 (and mutants), MLE (and mutants) and roX2 RNA were pre-incubated in presence of 1 mM ATP at 4°C with head-over-end rotation for 25 min. 20% of this reaction were saved as 'input'. Subsequently the reaction was incubated with the anti-MSL1-antibody coupled A/G beads at RT with head-over-end rotation for 45 min. The beads were washed three times with BCM-100 with six inversions and centrifugation at 500 g for 2 min at RT to settle the beads. The supernatant was discarded and the beads were split into two reaction tubes at the last washing step. Diluted 1x Laemmli loading buffer was added to one half of the beads and they were heated for 5 min at 95°C. These samples were further analyzed simultaneously with the 10% of the input samples by SDS-PAGE or Western blot. The other half of the IP-beads, as well as the remaining 10% of the input samples, were incubated with proteinase K (100 µg) and 0.5% (w/V) SDS at 55°C for 45 min. The supernatant was taken, phenol-chloroform-isoamyl extraction was performed and the nucleic acids were ethanol precipitated. Nucleic acid content was analyzed by denaturing urea PAGE and ethidium bromide staining (1:10,000).

4.2.14 Histone octamer assembly

Histone octamers (both wild-type and H4K16R variants) were reconstituted following the method outlined in Luger *et al.* 1997 [11]. Briefly, individual lyophilized histones were dissolved in unfolding

buffer, then combined at a molar ratio of 1.4:1.4:1:1 (H2A:H2B:H3:H4) and dialyzed for 16 hours against 2 L of refolding buffer. Subsequently, they underwent two additional dialysis steps for 1 hour each against 1 L of refolding buffer at 4°C. Following dialysis, the samples were centrifuged for 30 minutes at 30,000 g at 4°C. The resulting supernatant was applied to a Superdex 200 HiLoad chromatography column that had been equilibrated with refolding buffer, which served as well as the running buffer. Octamer-enriched fractions were identified through UV absorption at wavelengths of 280 nm and 214 nm, and further analyzed via SDS-PAGE. Concentration was determined by UV absorption at 280 nm using a Nanodrop spectrophotometer. Fractions were consolidated, concentrated to >1 µg/µL, then flash-frozen in liquid nitrogen and stored at -70°C.

4.2.15 Assembly of nucleosome arrays by salt gradient dialysis

Nucleosome arrays were assembled by salt gradient dialysis [217-219] on a pUC18 plasmid comprising 25 repeats of a 197 bp long Widom-601 nucleosome positioning sequence [220-222]. The plasmid DNA was mixed with octamer in a 1.1:1 mass ratio in high salt buffer. The salt was gradually reduced by dialysis overnight at 30°C for 15-18 h to low salt buffer. The dialysis was continued on the next day for 1 h against fresh low salt buffer. Chromatin concentration was determined based on DNA absorption at 260 nm wavelength. Chromatin was stored at 4°C for up to 12 weeks.

Nucleosome array assembly quality was evaluated by micrococcal nuclease (MNase) digestion as described [218]. 500 ng of chromatin were mixed with 3 µL of 1.5 U/µL units MNase in MNase buffer and incubated for 30 s, 60 s or 5 min at 30°C. The reaction was stopped by addition of 10 mM EDTA and 2% (w/V) SDS. The sample was treated with 2.5 µL Proteinase K (10 mg/mL) for 30 min at 37°C. DNA was precipitated with 325 µL 100% ethanol and 65 mM NaCl for 20 min at -20°C, followed by centrifugation at 21,000 g at 4°C for 30 min and washed once with 400 µL of 70% ethanol. The pellet was air-dried and resuspended in 15 µL 10 mM Tris-HCl pH 8.0 and 3 µL of Orange G loading dye. The digestion degree was analyzed on a 1.5% (w/V) agarose gel in 1x TAE buffer, stained with Midori green.

4.2.16 RNA preparation

In vitro transcription of *roX2* and *GFP* RNA was performed using the MegaScript T7 RNA polymerase kit (Ambion) or T7 polymerase (NEB), following the manufacturer's instructions. The DNA templates of *roX2* constructs were prepared by restriction enzyme digestion with XbaI (NEB) as described [137]. GFP templates were amplified by PCR from an pHSP70-MLE-GFP plasmid [52]. After transcription (4 h, 37°C) and DNase I treatment (30 min, 37°C), RNA products were purified by denaturing PAGE (8 M urea, 5% acrylamide, 1x TBE), phenol:chloroform:isoamylalcohol extraction and ethanol precipitation. RNA was dissolved in 50 µL of nuclease-free water and stored at -70°C.

4.2.17 Histone Acetyltransferase (HAT) assay

0.8 µg of 25-repeat Widom-601 nucleosome arrays and protein complexes were incubated in HAT buffer and 10 µM acetyl-CoA at 26°C. Protein concentration (12.5–175 nM) and incubation times (2–180 min) varied between the different sets of experiments and are provided in the respective figures. HAT reactions with the 4-MSL complex and RNA included 1 mM ATP, 1 U RNasin and MLE at a 1:1 molar ratio to the 4-MSL complex. RNAs were titrated (25 nM – 200 nM) or applied in 2-fold excess as mass ratios. HAT reactions with dTip60^{piccolo} were performed with 50 nM protein complex and incubation times from 5 to 60 min. When specified, 100 nM of *GFP* RNA and 0.2 µg of RNase A were introduced. The reaction was halted by incubating at 95°C for 5 minutes. Histone acetylation analysis was conducted either through Western Blotting or mass spectrometry (Chapter 4.2.10). The denatured HAT reaction samples were supplemented with SDS-PAGE loading buffer and heated before being loaded onto a 15% (V/V) acrylamide Bis-Tris-MES gel in 1x Running MES buffer. Proteins were separated at a constant voltage of 140 V and then transferred (1 hour at 400 mA, at 4°C) onto a nitrocellulose membrane in transfer buffer. Subsequently, the membranes were blocked in 3% (w/V) BSA in TBS for 1 hour at room temperature (RT) and incubated overnight at 4°C with antibodies specifically targeting histone H3 (anti-H3) and histone H4 acetylation (anti-H4K16ac, anti-H4K12ac), respectively. Following three washes with TBS-T, the membranes were probed with species-specific secondary antibodies and images were captured using the LiCOR imager.

4.2.18 RNase A test of protein preparations to evaluate RNA content

A sample of 15 µL of the purified protein elution was pipetted into a fresh reaction tube and subsequently 1 µL of bovine pancreas RNase A (stock concentration 10 mg/mL) of the desired dilution in HAT assay buffer was added. Alternatively, in case of the control reaction 1 µL of ddH$_2$O was added. The reaction was incubated for 10 min at RT. After the incubation, 1.25% (w/V) SDS were added and the reaction was supplemented with 2 µg of proteinase K. Proteins were digested for 60 min at 55°C and 500 rpm. To perform phenol-chloroform extraction of the remaining nucleic acids, ddH$_2$O to complete 400 µL was added. 400 µL of phenol-chloroform-isoamylalcohol for RNA was added and the samples were vortexed for 30 s each at maximal vortex setting. Aqueous and organic phase were separated by centrifugation at 17,000 g for 30 min at RT. The aqueous upper phase was transferred into a fresh reaction tube and precipitated with double-volume, -20°C cooled, 100% p.a. ethanol at -70°C for 1 h. The precipitated nucleic acid was pelleted by centrifugation at 21,130g at 4°C. The supernatant was discarded carefully and the pellet was washed once with 70% (V/V) ethanol. The pellet was let to dry for 10 min at 70°C and then resuspended in 15 µL of nuclease-free water. The sample was supplemented with blue RNase-free loading dye and analyzed on a 5% (V/V) acrylamide 8 M urea gel in TBE buffer. After 90 min at 100 V separation, the gel was stained by Midori green 1:10,000 for 15 min at RT and imaged in the UV-gel documentation chamber.

Part I – Form follows function, function follows form

The molecular mechanism of how the MSL complex operates still remains elusive to date. Due to the intricate and reciprocal relationship of form and function, structural information can elucidate the molecular mechanism of the MSL complex. One long-standing theory in the field of dosage compensation established roX RNA as a key factor for dosage compensation. The exact structure of the MSL complex is not known, neither without nor with the lncRNA integrated. However, the idea of conformational change upon the integration of roX RNA is very intriguing. A different form might allow for a different function, which in the other conformation might be impossible. Such allosteric regulatory mechanisms are common in biological systems and thus are an attractive hypothesis to explain the critical role of the lncRNA roX for male fly viability.

The first aim is to characterize the interplay of the MSL subcomplexes in absence of roX RNA. To pursue this aim, I characterized the interaction interfaces by crosslinking mass spectrometry (XL-MS). Secondly, these results were compared to computational models of the MSL complex architecture. Furthermore, the effect of roX RNA on the structure of the helicase MLE and on the complex of MSL1-MSL2 after incorporation by MLE was explored. Finally, the data were used and can be further used for integrative structural biology approaches supporting other structural techniques such as Cryo-EM to investigate the full MSL complex or its submodules.

5 Results

5.1 Crosslinking mass spectrometry offers structural information

The MSL complex is a large multi-subunits complex, which is challenging to study by crystallography. Despite partial structures of conserved mammalian homolog proteins, the overall architecture and quaternary structure remains unresolved. Therefore, XL-MS was applied in this study to address the question of the interacting domains within the complex and to contribute to the structural analysis. The general workflow of the XL-MS approach uses purified protein complexes, which can be crosslinked by a chemical substance called the crosslinker, in this case, bis(sulfosuccinimidyl) suberate, hereafter BS3 (Figure 11, [166]). The covalently connected proteins are digested into peptides by peptidases, such as trypsin for the analysis in MS. The mass of the peptides and the fragmentation of the peptides is determined by Orbitrap MS. Lastly, the masses and fragmentation patterns are compared to a computed database of all hypothetically possible peptides by the program 'Crossfinder' [169, 171]. The additional weight of the crosslinker and the connected peptide, which is simultaneously fragmented gives insight on which exact amino acid reacted to form the chemical bond.

In the case of BS3, which contains two reactive sulfosuccinimidyl groups at each of its extremities, the covalent bond is formed between BS3 and primary amines (Figure 11B). In proteins, primary amines are frequently found as part of the lysine residue side chain. Additionally, the very N-terminus of a

protein can potentially react with BS3. If each end of the crosslinker reacts with a different lysine, a 'crosslink' between two peptides is formed. If only one group reacted, so called 'mono-links' are produced. Mono-links can offer information about surface exposure of a protein complex, as the only reaction partners available are BS3 and the solvent. The crosslinker contains a flexible spacer arm with a defined length. The extended spacer arm and the length of the lysine side chains, defines the maximal spatial distance that can be connected by BS3. For BS3 this length is usually assumed to be maximally 30 Å (Figure 11B). This distance can be evaluated further in structural models obtained by crystallography or by computational modelling, whether in the current model under investigation the crosslink would be plausible ('satisfied') or non-plausible ('violated'). If the XL-MS contradicts the crystallographic data, this can imply that another conformation is obtained in solution, which was not captured by the crystal structure. If the XL-MS contradicts an *in-silico* model, the XL-MS data can be used to further refine the model by using the covalent connections as distance restraints.

5.2   *In vitro* reconstitution of the MSL complex

To study the structural relations of the MSL complex components, it is central to aim for the optimal reconstitution *in vitro*. The superior reconstitution of the MSL complex contains all the described components of the *in vivo* system in the appropriate stoichiometry. To achieve in parallel a high yield with excellent stoichiometry, satisfying purity and quality of each of the subunits is challenging. This challenge was overcome through a *Spodoptera frugiperda* cell expression system [143]. By advanced baculovirus expression vectors it was feasible to co-express and co-purify four of the MSL complex subunits together with superior stoichiometry (Figure 14).
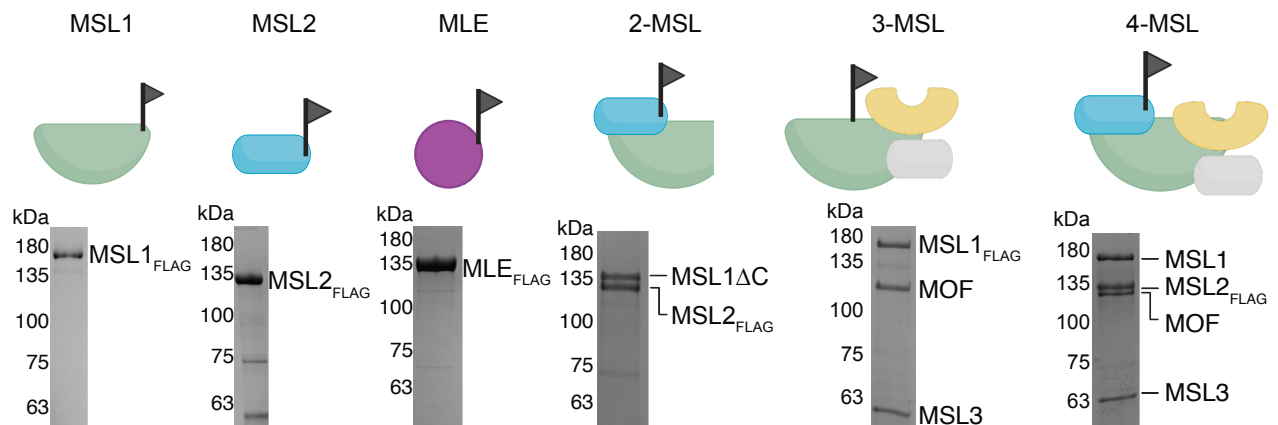
Figure 14: FLAG affinity chromatography-purified MSL proteins and MSL complexes of different subunit composition. Coomassie-stained SDS-PAGE gels demonstrate a satisfactory purity of the protein preparations and an equimolar stoichiometry for the 2-MSL, 3-MSL and 4-MSL complexes.

The assumed stoichiometry of 1:1 follows the principle of simplicity, not to overcomplicate a hypothesis without a given reason. Alternatively, a 2:2 stoichiometry is probable [49]. However, it would be impossible to differentiate between 1:1 and 2:2 on a simple Coomassie blue- stained SDS-PAGE image and more sophisticated methods such as small-angle X-ray scattering (SAXS), analytical ultracentrifugation or mass photometry would be necessary to investigate this question further.

Results

After expression in Sf21 cells, the MSL proteins or subcomplexes were co-purified using a FLAG affinity chromatography. The FLAG-tag is attached at the C-terminus of the desired MSL protein, to assure only full-length expressed proteins are purified. To prevent N-terminal degradation of the protein of interest after its expression, protease inhibitors are included in the lysis buffer. Larger MSL complexes are co-purified, which means that only one of them carries the FLAG-affinity tag. Usually, it is placed at the C-terminus of MSL2. Of course, in the subcomplexes lacking MSL2, such as the 3-MSL complex (consisting of MSL1, MSL3 and MOF), the FLAG tag is attached to another protein. In the case of the 3-MSL complex the tagged subunit is MSL1 (Figure 14).

In the two-subunit complex 2-MSL, consisting of MSL1 and MSL2, MSL1 was truncated at the C-terminus (MSL1ΔC, Figure 14). The MSL1 C-terminus was deleted after amino acid 865, because the resulting 2-MSL complex demonstrated a better stability and stoichiometry compared to the full-length MSL1. Moreover, the MSL1 C-terminus is known for its interactions with MSL3 and MOF, whereas the N-terminal coiled-coil region of MSL1 has been described to be critical for the interaction with MSL2 [49, 50]. Thus, when the interaction between MSL1 and MSL2 was analyzed, the C-terminus of MSL1 was not considered.

5.3   The 2-MSL subcomplex can be crosslinked by BS3 in solution and on beads

The 2-MSL subcomplex was chosen as a starting point to study MSL protein interactions, due to its small size and its previously described interaction at the N-terminus of MSL1 and the RING domain of MSL2 [49]. As a consequence of the known interaction, potential structural changes following the incorporation of roX-RNA can be studied.

To elucidate the interaction interfaces of the MSL protein subunits, XL-MS was applied. As the crosslinking agent bis(sulfosuccinimidyl) suberate was chosen, because it offers spatial flexibility due to its 11.4 Å long spacer arm. BS3 reacts with its two reactive groups on each end of the spacer arm with primary amines (Figure 11B). In the case of proteins, primary amines are present at lysine residues and at the very N-terminus of a protein. Moreover, previous experience at the laboratory with this crosslinker provided the basis for the mass spectrometric setup and data analysis [169-171, 223]. Initial experiments were undertaken in an in-solution crosslinking approach, using a range of 50-500 µM BS3 (Figure 15A). The mass spectrometry measurement of the in-solution samples did not report any crosslinked peptides. Thus, the approach was changed to an on-bead crosslinking protocol (Figure 15B). On beads, 50 µM – 4 mM BS3 were used, up to 8-fold higher than the published in-solution crosslinker amounts [165, 167, 170]. Presumably, a lot of the reaction is quenched by crosslinking merely the anti-FLAG antibodies, which are coupled to the agarose beads. A band of the antibody heavy chain protein and the crosslinking products can be observed on the Coomassie stained gel (Figure 15B). In the high molecular weight range as well as in the gel pockets the crosslinking products of the MSL proteins can

be observed as a smear. Usually, one would expect to obtain a clear band of the crosslinked product, however, the product could be very large and thus not enter the gel properly, hence the staining in the gel pockets. Alternatively, the products are very heterogeneous, thus leading to a smear instead of a clear band. Lastly, aggregation has to be kept in mind, which can be promoted by excessive crosslinking, so called 'over-crosslinking'.
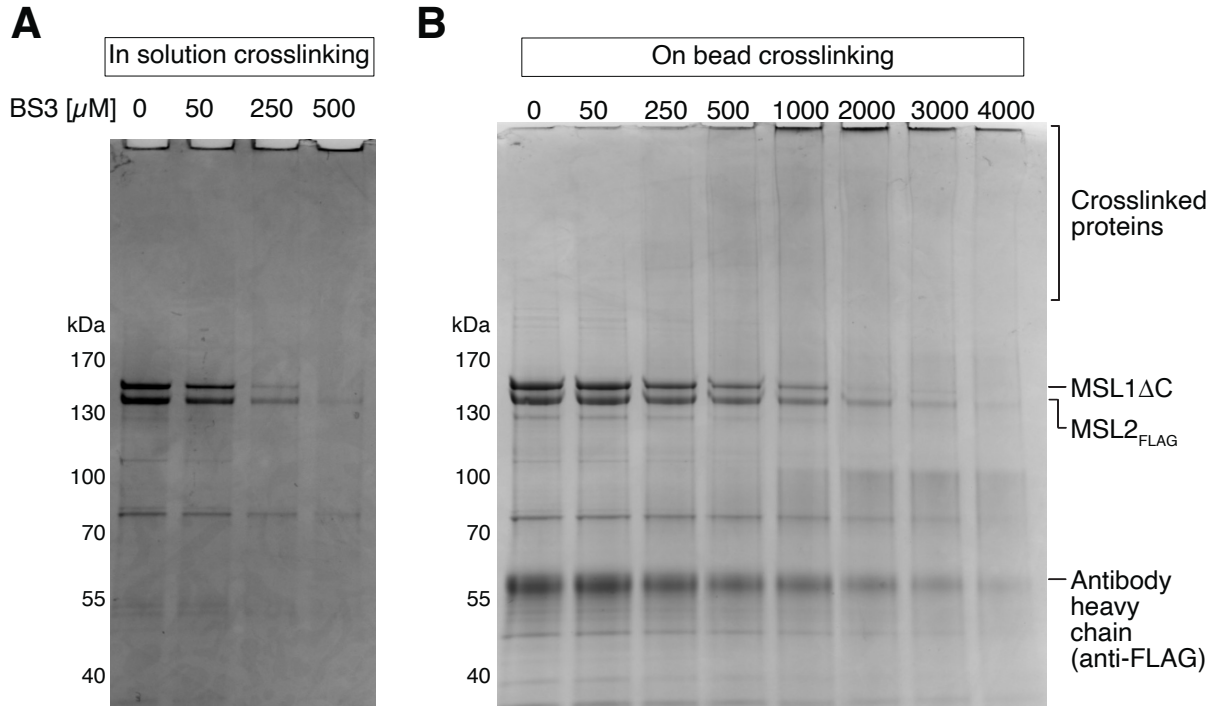
Figure 15: Crosslinker BS3 titration for the 2-MSL complex. **A** In solution 50 to 500 μM of BS3 were sufficient to crosslink the protein complex. **B** On beads the protein complex is crosslinked only at higher BS3 concentrations, above 500 μM. In the higher molecular weight range, as well as in the gel pockets, the crosslinking products can be seen as a smear (crosslinked proteins). Unexpectedly, there was no clear band visible for the crosslinked MSL1-MSL2 dimer.

After the initial titration, the BS3 concentration for on-bead XL-MS was settled at 0.75 – 1.5 mM, based on the shadow appearance in the high molecular weight area on the Coomassie stained PAGE gel, as well as the disappearance of the educt bands, MSL1ΔC and MSL2, on the gel image (Figure 15B). Preferably, a different and more sensitive staining method of the gel could be used, for example silver staining or Western blot staining with specific antibodies. That way, potential product bands of low quantity could be observed.

5.4 XL-MS identified novel interaction regions between MSL1 and MSL2

After applying BS3 in concentrations of 0.75, 1.00 and 1.50 mM to the 2-MSL complex coupled to FLAG-beads, the reaction was quenched by tris(hydroxymethyl)aminomethane (TRIS). TRIS is not only a commonly used buffer substance, but it also has a primary amine group, which reacts promptly with the BS3 reactive groups and thus stops the crosslinking reaction. The crosslinked proteins are then digested by trypsin proteinase overnight. The obtained peptides are then partitioned from the beads and

Results

enriched by styrene divinylbenzene-reversed phase sulfonate (SDB-RPS) stop-and-go-extraction tips (StageTips) for mass spectrometric analysis [214, 224]. The elution of these SDB-RPS StageTips is ready for injection into the LC-MS/MS workflow in an Orbitrap mass spectrometer [169, 170]. The raw data were processed by the ThermoFisher proteomics software "Proteome Discoverer" and then further analysed by an in-house, matlab-based software package called "Crossfinder" [171]. The principle of that tool is based on the calculation of a database, containing all possible peptides of a certain protein and all possible crosslinks with their respective masses. If there are many proteins in a protein complex, this calculation is computationally resource- and time-consuming, which is why an analysis on a computational cluster is highly advisable. In the following steps, the database can be searched for in the mass spectrometric data and peaks with the corresponding masses identified. Since there can be also false hits, a decoy-database with randomized amino acid sequences is searched. If a hit is scored in the decoy mass space, this contributes to a higher false-discovery rate (FDR). The FDR was generally limited to 5% in this project, however expert opinions diverged upon the inclusion of only peaks with lower FDRs (less than 1%) or rather including peaks, which have a trust-worthy fractionation pattern (manually evaluated), but could have FDRs of up to 25-30% [166-168, 225]. The obtained crosslink can be formed between two lysine residues within the same protein, referred to as an 'intramolecular crosslink', or between two different proteins in the same reaction, a so-called 'intermolecular crosslink'. These covalent bonds are spatial restraints for further modelling approaches and inform about protein interaction regions.

In the XL-MS dataset obtained of the 2-MSL complex 147 crosslinks of an FDR of < 5% were identified (Figure 16B, see Appendix 12.2.1). Of these crosslinks, 113 are intramolecular crosslinks within MSL1ΔC, most of them located in the intrinsically disordered, central domain of MSL1, approximately MSL1$_{358-799}$ (Figure 16A and B). 27 crosslinks are intramolecular crosslinks within MSL2, most of them located in the "pre-CXC" domain (MSL2$_{468-524}$). N-terminal of that region a coiled coil domain is predicted MSL2$_{424-468}$. Surprisingly, only 7 of all 147 crosslinks are intermolecular crosslinks, between MSL1 and MSL2. One of them confirmed the previously known interaction between the MSL1 coiled coil domain and the MSL2-RING domain (MSL1-K269:MSL2-K135). Notably, the crosslink was not found directly inside the respective domain, but next to it.

Moreover, two crosslinks were identified between the MSL1 coiled coil and a predicted coiled coil region in MSL2 (MSL1-K71:MSL2-K436 and MSL1-K102:MSL2-K436). At first glance, the simultaneous crosslink from MSL2-K436 to two different lysine residues in MSL1 seems to make no sense. However, due to the ensemble technique of XL-MS, it is possible that a population of proteins is crosslinked to one residue, whereas another population is crosslinked to another residue.

Furthermore, four crosslinks were identified in previously unknown interaction regions between MSL1 and MSL2, located in the MSL1 intrinsically disordered region (IDR, MSL1$_{358-799}$) and in the MSL2 pre-CXC domain (MSL2$_{468-524}$). In both domains the IDR score is very high, which indicates that

multiple, dynamic and flexible conformations exist as no rigid structure is known or predicted (Figure 16A).

To sum up, the XL- MS of the 2-MSL complex confirmed the known interaction of the N-termini and in addition revealed novel interaction domains in the MSL1-IDR and in the MSL2-pre-CXC domain.
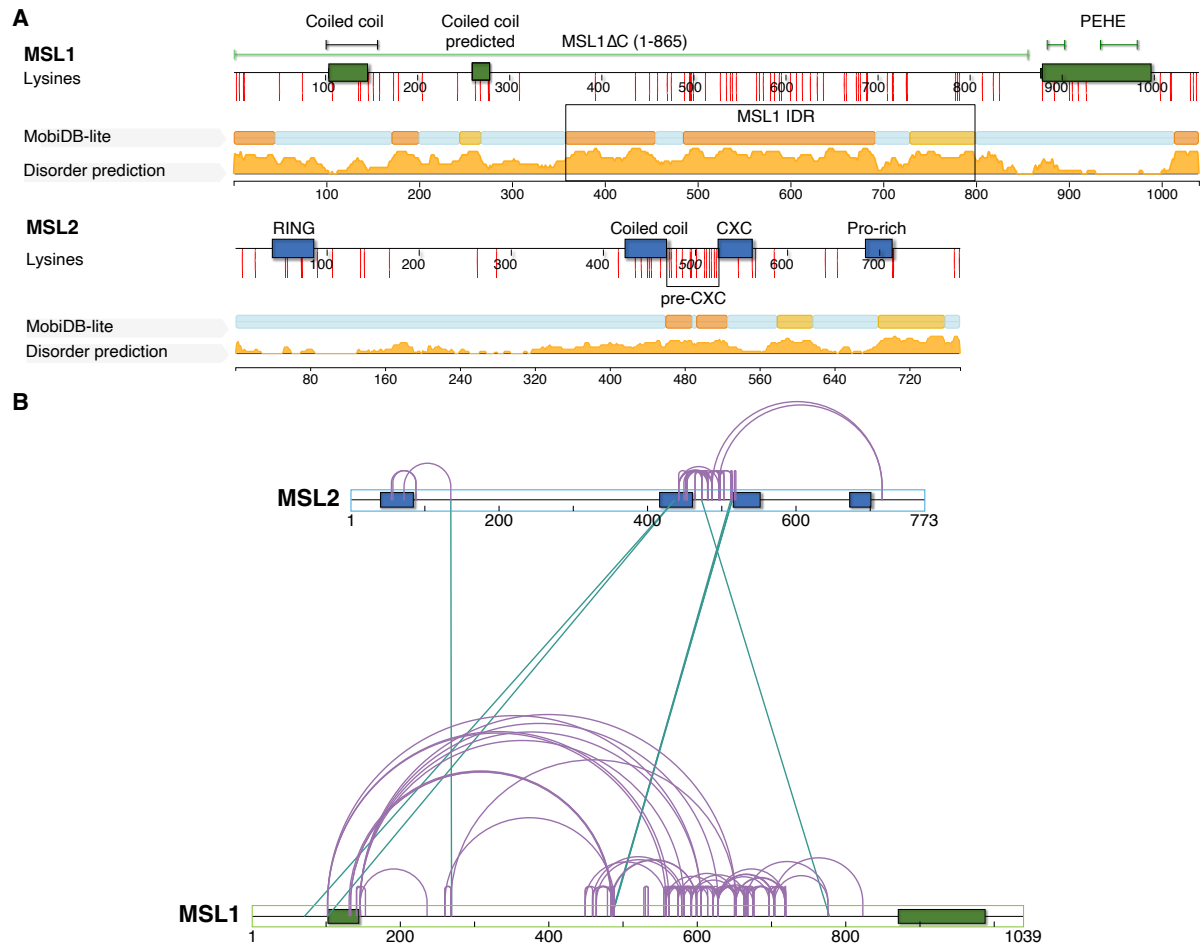
Figure 16: MSL1 and MSL2 contain intrinsically disordered regions which are engaged in intra- and intermolecular crosslinks. **A** Schematic representation of MSL1 and MSL2 domains with MoBiDB Lite IDR prediction [51]. Orange tracks represent the likelihood of IDR. Most of MSL1 is predicted to be disordered, particularly the central domain MSL1$_{358-799}$. Lysine residues are presented as red lines under the domain scheme. **B** Network representation of crosslinks within (purple arch) and between (blue line) MSL1ΔC and MSL2 by xiNet viewer (https://crosslinkviewer.org/, 16.10.2023, [226]). The known interaction between the MSL1 coiled coil and MSL2 RING domain was confirmed by one crosslink. Several new interactions were detected between IDRs and predicted coiled coil regions.

5.5   XL-MS calling, replicate samples and reproducibility

Usually, crosslinking mass spectrometry is not performed using replicates [168], nevertheless I aimed to reproduce each of my experiments at least once. For each of the data sets a fresh protein preparation from a different expression pellet was prepared, such that the protocol was repeated from the beginning accounting for variability during the expression and purification protocol. Likely, this introduces a lot of variability from the expression cells as well as from the purification procedure. Nevertheless, the overall crosslinking pattern was confirmed, however, some of the exact crosslinks were not always

detected. If a crosslink was found twice in these quite independent experiments, it would be assigned a 'Class 1', reflecting a good quality due to reproducibility. Within each 'data set' there are three 'technical' replicates, which are prepared from the same protein preparation, but using different amounts of crosslinker. If a crosslink was found more than once among the technical replicates, it would be assigned 'Class 2'. If a crosslink was found just once under a certain condition, it would be assigned 'Class 3' (not reproduced). For the calling of crosslinks, I used the standard filtering of the Crossfinder software (Materials and Methods [170, 171]). It was ensured that the false discovery rate (FDR) was below 5%. Certainly, the FDR can be lowered for stricter calling, if needed. However, in the field it is common to use manual evaluation of the crosslinks, regardless of the FDR [227]. Since the underlying principle of XL-MS relies on an in-solution dynamic system it may make sense that not each and every crosslink will be detected in each sample given that different conformations may coexist and be differently populated in the samples.

5.6 Identification of intramolecular crosslinks in individual MSL1 and MSL2 proteins

Due to the scarcity of intermolecular crosslinks, the next approach aimed to study, whether there were intramolecular changes in crosslinking pattern when MSL1 or MSL2 were crosslinked individually. Conceivably, the internal folding or the overall conformation of the proteins could change, when they are assembled into the 2-MSL complex.

5.6.1 XL-MS of MSL1 reveals dynamics and flexibility

First, the complete MSL1 protein ('full length') was crosslinked on beads with BS3 (concentrations 0.5; 0.75 and 1.0 mM). After MS analysis 596 unique crosslinks were detected in full-length MSL1 (Figure 17A, Appendix 12.2.2). Almost every lysine residue is crosslinked under these conditions. This suggests that MSL1 is a very flexible protein and can adopt many different conformations revealed by this ensemble technique. In comparison to the crosslinking of the 2-MSL complex, MSL1 showed many more crosslinks when alone (113 versus 596 intramolecular crosslinks). One explanation of the lack of crosslinks is the MSL1 truncation at the C-terminus in the 2-MSL subcomplex. However, association of MSL1 in a complex with MSL2 may limit the conformational space. Certain conformations are more stable while others are sterically not allowed. Moreover, MSL1 full-length on its own is a rather unstable protein and is prone to degradation. In complex with MSL2 its stability is much improved. Interestingly, MSL1 shows two crosslinks located at its C-terminus, which are crosslinked directly with the same position again, one in the PEHE domain, $MSL1_{879-996}$, and one in the very C-terminal region, (Figure 17A, indicated by red loops). This is of interest, because these links are usually an indication that the protein dimerizes in that region, which is so far not described for MSL1. However, the N-terminal region of MSL1 coiled coil domain is known to homodimerize in presence of MSL2 [49]. The expected "loop"

links at those domains were not found by crosslinking MS approach (Figure 16). It is not possible to conclude that the 2-MSL complex remains as a monomer under those conditions because the interactions might have actually formed but were not detected. Alternatively, the conformation of the proteins can be such that the lysines are oriented away from each other, such that the crosslink cannot be formed despite dimerization.

As a tentative conclusion, it appears that MSL1 in isolation could potentially be an unfolded protein, and it could require interaction partners to achieve a stable conformation when forming complexes. MSL1 certainly obtains numerous conformations in the ensemble method.
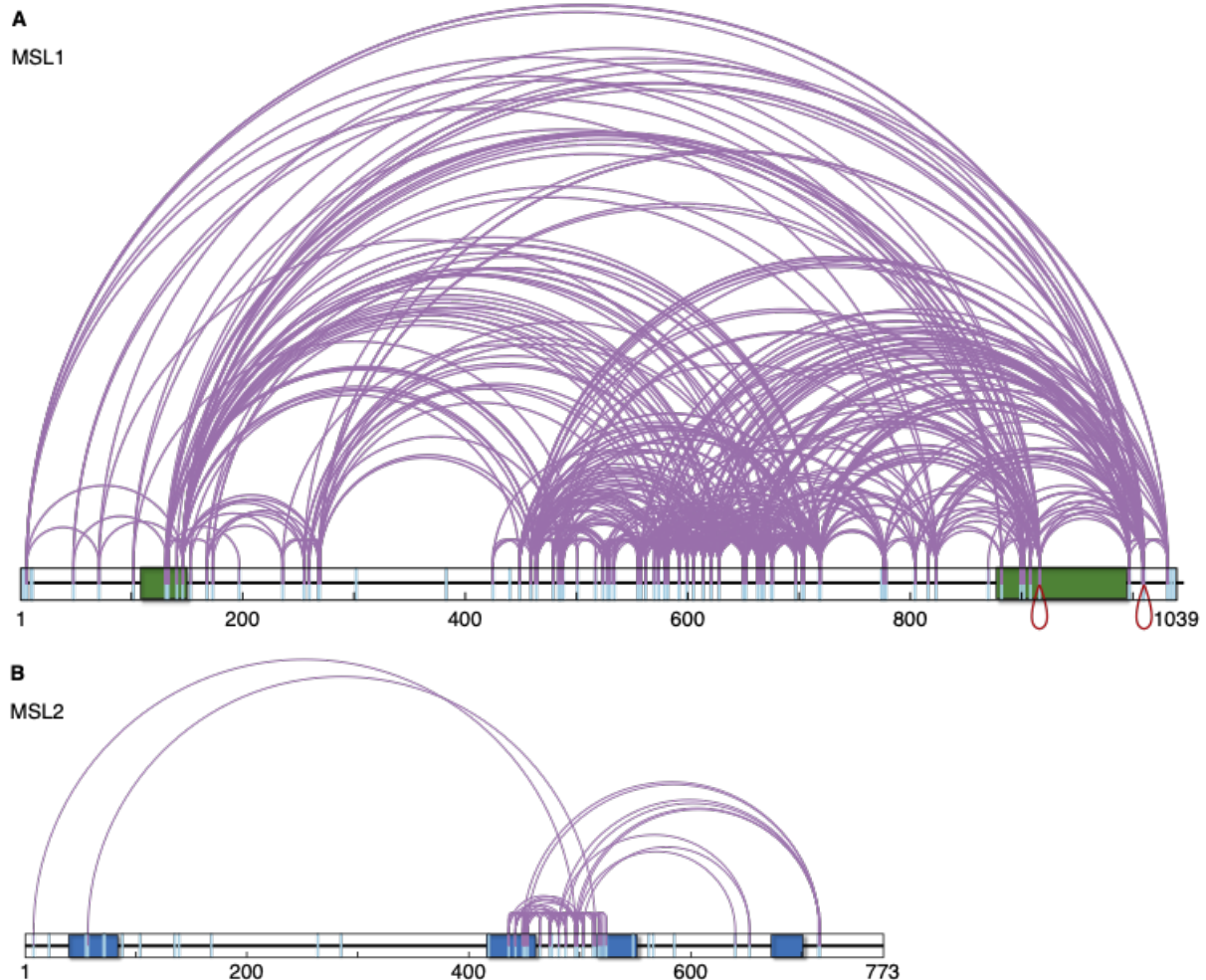
Figure 17: Intramolecular crosslinks of individual MSL proteins. **A** MSL1. **B** MSL2. Purple arc: intramolecular crosslink, red loop: crosslink with the same lysine residue again, light blue lines: lysine residues in the amino acid sequence of the proteins. Graph generated with xiNet viewer (https://crosslinkviewer.org, 19.10.2023, [226]).

5.6.2   XL-MS of MSL2 reveals internal interactions of domains

In order to validate the observed intramolecular crosslinks orthogonal to MSL1, the examination was extended to MSL2. Subsequently, MSL2 underwent crosslinking with BS3 on beads after FLAG affinity chromatography purification in the absence of MSL1 (Figure 17B, Appendix 12.2.3). Overall, there were 145 intramolecular crosslinks detected in MSL2, which is substantially more than the 27 intramolecular crosslinks detected in MSL2 within the 2-MSL complex (Figure 16). However, in MSL2, the crosslinking pattern was more structured than in MSL1, which gives a better insight into the fold of

the protein itself. Two lysine residues of the MSL2 N-terminus crosslink with the pre-CXC domain, both in the RING domain (MSL2-K8:MSL2-K496, MSL2-K57:MSL2-K514). This suggest that the pre-CXC domain at least transiently interacts with the N-terminus. The pre-CXC domain and the predicted coiled coil are once again the main regions where numerous crosslinks are formed. This crosslinking reactivity may be interpreted in two ways. Firstly, it may indicate high flexibility and the capacity to assume various conformations. On the other hand, it could result from solvent exposure of lysines due to the lack of secondary structures. However, this explanation falls short in clarifying why crosslinks form instead of just mono-links, where one of the reactive groups of BS3 does not interact with another lysine residue. Nevertheless, orthogonal methods are necessary to solve this puzzle. There are multiple contacts also reaching out to the proline-rich C-terminal domain of MSL2 ($MSL2_{685-713}$, Pro-rich). The proline-rich domain does not contain many lysine residues, yet the single pair of lysines (MSL2-K715, MSL2-K716) formed 11 crosslinks to the pre-CXC domain. Interestingly, the CXC domain itself did not form any crosslinks, despite three lysine residues are located in or adjacent to it. The absence of detected crosslinks can be interpreted in two ways. First, there are no crosslinks because the structure is rigid and far away from other lysine residues, that there are no possible ways of crosslink formation. Second, there are interactions with other domains, however, the crosslinks were not detected due to low signal intensity in MS.

To sum up, MSL2, similar to MSL1, showed more crosslinks, when analyzed individually compared to the 2-MSL complex. The more limited conformational space in the 2-MSL complex can be explained by occlusion of interaction surfaces by MSL1.

5.7 The interaction regions in the 3-MSL complex are confirmed and additional contact sites are identified

The next step towards understanding the interactions in the MSL complex was to increase the size of the complex. A stable, larger complex is the three-subunit reader-writer module called hereafter 3-MSL. This complex consists of three subunits, MSL1, MSL3 and MOF. It has acetyltransferase activity of MOF and can bind to nucleosomes. MSL1 is often thought to be merely a scaffolding protein; however, it has been shown that it is necessary for the efficient acetylation of nucleosome array substrates and for the efficient binding of the 3-MSL complex to mononucleosomes [184]. The MSL1-PEHE domain was earlier characterized to interact with MSL3's MRG domain in the conserved mammalian protein, so was the MSL1-PEHE and MOF-HAT interaction (Figure 4, [50]).

The FLAG-purified 3-MSL complex (Figure 14) was subjected to on-bead crosslinking with 1 mM BS3, and the crosslinked peptides were identified by MS.

Results

The crosslinking of the 3-MSL complex yielded 257 crosslinks in total, in other words the sum of intra- and intermolecular crosslinks in and between MSL1, MSL3 and MOF (Figure 18, Appendix 12.2.4).
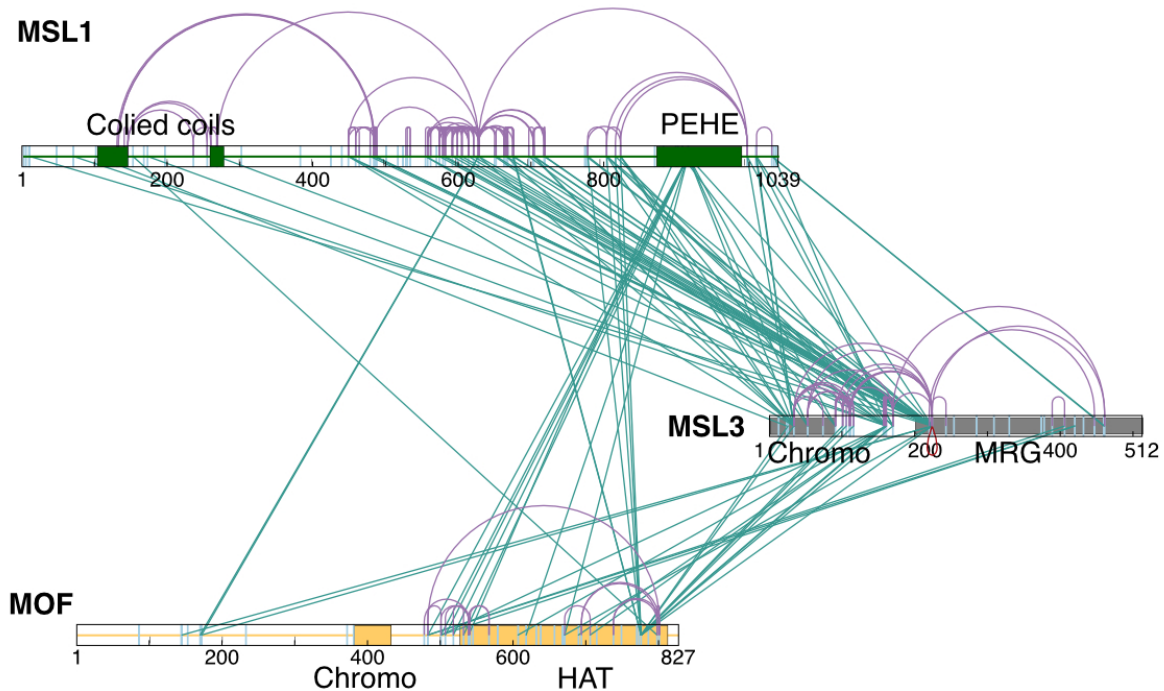
Figure 18: Known and novel interaction regions in the 3-MSL complex. 3-MSL crosslinking confirmed known interaction sites between MSL1-PEHE and MSL3-MRG as well as MSL1-PEHE and MOF-HAT domains. Many additional contact sites are detected. Purple arcs: intramolecular crosslinks, green lines: intermolecular crosslinks, light blue lines: lysine residues. Representation plotted with xiNet viewer (https://crosslinkviewer.org, 19.10.2023, [226]).

In MSL1 107 crosslinks were found, mainly in its IDR (MSL1$_{358-799}$, Figure 19A). Moreover, 33 intramolecular crosslinks were found in MSL3, predominantly in the more than 100 amino acid long linker region MSL3$_{91-195}$ (Figure 19B). Lastly, 15 intramolecular crosslinks were found in MOF, all of them C-terminal of the MOF-chromobarrel domain ('Chromo') (MOF$_{370-443}$, Figure 19C).

Results
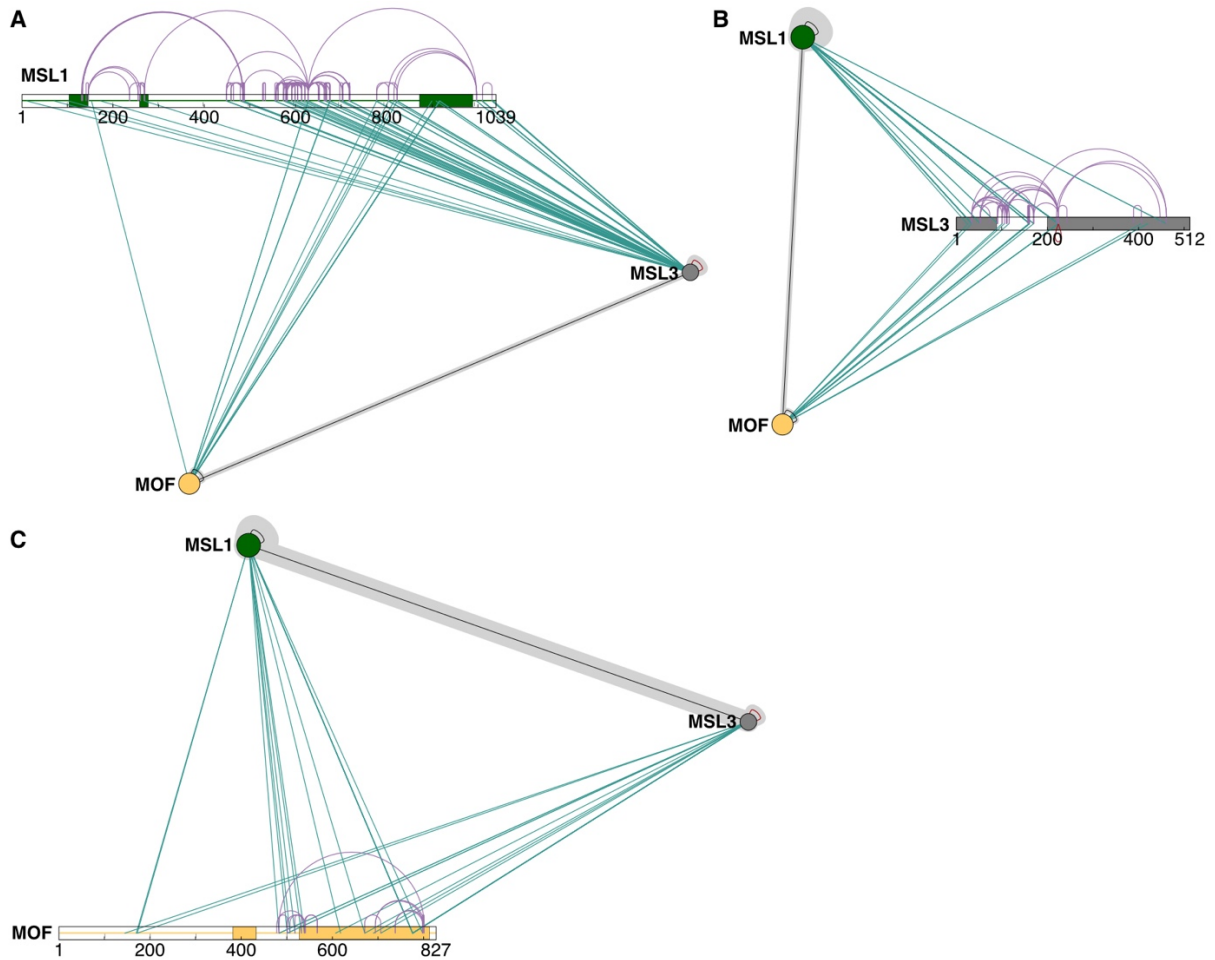
Figure 19: Crosslinks found in the 3-MSL complex. **A** MSL1 associated crosslinks intramolecular crosslinks as purple arcs, intermolecular crosslinks as green lines. MSL3 and MOF are collapsed for clarity. Grey line indicates that there are as well several crosslinks between MSL3 and MOF. **B** MSL3 related crosslinks, where MSL1 and MOF are collapsed for clarity. **C** MOF involved crosslinks, where MSL1 and MSL3 are collapsed for clarity. Representation generated with xiNet viewer (https://crosslinkviewer.org, 19.10.2023, [226]).

The expected interactions between the MSL1 PEHE and MOF-HAT domain and the MSL1-PEHE And the MSL3 MRG domain were identified, these are already known contacts published by the Akhtar laboratory on conserved mammalian domains of these proteins [50]. Nevertheless, the confirmation of the known interactions is a great reassurance that the reconstituted complexes and the crosslinking MS method yield consistent results.

Furthermore, 102 intermolecular crosslinks were detected (Figure 18, Appendix 12.2.4). First, the crosslinks found in known interaction regions are presented. These interactions have been described between the MSL1-PEHE domain ($MSL1_{879-996}$) and the MOF-HAT domain ($MOF_{538-813}$), as well as the MSL1-PEHE domain and the MSL3-MRG domain ($MSL3_{196-500}$) [50].

Three crosslinks were identified, which connect the MSL1-PEHE with the MOF-HAT directly (MSL1-K899:MOF-K539; MSL1-K916:MOF-K618; MSL1-K916:MOF-K671 (Figure 20)). Moreover, eight crosslinks connect the two domains adjacent to the domain limits, which is expected, considering that the BS3 crosslinker can span up to 30 Å between two lysine residues. Interestingly, four of these positions can be observed in the crystallographic structure model of the conserved mammalian protein

Results

[50]. When the distances between the conserved lysine residues were measured by pymol software tool (Schrödinger), the distance <30 Å was validated (Figure 20).
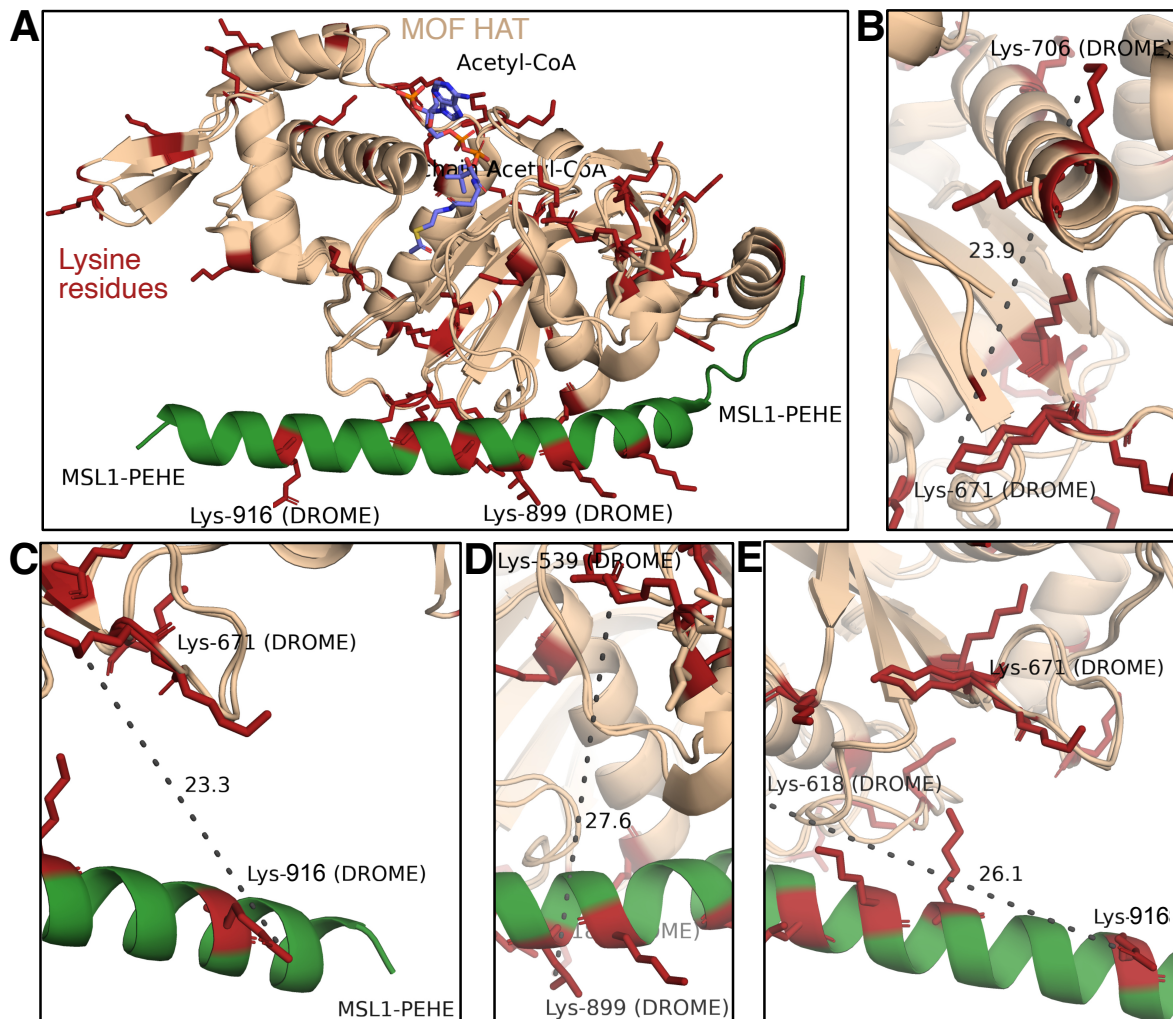
Figure 20: Structural information of the MSL1-MOF interaction was confirmed by crosslinking MS. **A** Human MOF HAT domain in complex with the mouse MSL1 PEHE domain overlayed with the MOF conserved HAT domain by pymol software, PDB: 2Y0M [50]. Murine MSL1-PEHE domain is shown in green, MOF in beige, lysine residues in the mammalian protein in red, acetyl-CoA in blue. B-E: Examples of crosslinks between the MSL1-PEHE and the MOF-HAT within the 30 Å distance restraint, residue numbers correspond to the *Drosophila melanogaster* positions. **B** Intramolecular crosslink in MOF MOF-K671:MOF-K706. **C** MSL1-K916:MOF-K671 **D** MSL1-K899:MOF-K539 **E** MSL1-K916:MOF-K618. Graphics with pymol (Schrödinger).

Another previously characterized interaction is the contact between the MSL1-PEHE domain and MSL-3-MRG domain [50]. One crosslink within the two respective domains was found, confirming the interaction (MSL1-K916:MSL3-K224). Likewise, there were 62 additional intermolecular crosslinks between MSL1 and MSL3, some of them in the neighboring regions to the MRG or PEHE domains. While MOF interacts with the PEHE domain N-terminally (MSL1$_{879-943}$), MSL3 binds to the C-terminal part of the PEHE domain (MSL1$_{944-996}$). The identified crosslinks could not be seen in the published crystal structure of the murine MSL1-PEHE with the human MSL3-MRG domain. The failure to detect these crosslinked may be explained in different ways. First, the corresponding peptides may be not detected in the MS. Second, the lysines may be buried and not accessible to the crosslinking agent BS3. Third, the interaction may be different in the crystal structure compared to in solution. Lastly, the

interaction in mammalian proteins may differ from the *Drosophila melanogaster* protein despite high evolutionary conservation of these domains.

Notably, most crosslinks identified by the crosslinking MS were novel interaction sites. Remarkably, two crosslinks connected the MOF N-terminus with the IDR region (MSL1-K628:MOF-K170; MSL1-K628:MOF-K172). Due to the IDR in MSL1 there might be a high flexibility and solvent exposure of that region, which makes multiple options for crosslinking possible. In this case, the two neighboring MOF residues are already in vicinity in the amino acid sequence, such that it makes sense that they might interact with the same region in MSL1. Further, crosslinks were detected between the MOF-HAT domain and the IDR of MSL1.

Remarkably, 17 novel crosslinks were found between MSL3 and MOF. This direct interaction has not been described before and might be caused by the close spatial distance of MSL3 and MOF both binding to the PEHE domain of MSL1 (Figure 18). Two of these 17 crosslinks were found between the MOF N-terminus and the MSL3-linker (MOF-K145:MSL3-K224 and MOF-K172:MSL3-K224). Three of them are located between the MSL3-MRG domain and a linker region of MOF between its chromobarrel domain and HAT domain (MOF-K501:MSL3-K406 and MOFK483-MSL3-K420). The remaining 12 crosslinks connected the N-terminal part of MSL3, its chromo domain and the linker region, to the MOF-HAT domain.

Curiously, in MSL3 often the same few lysine residues crosslink to various regions within MSL1-IDR or MOF-linker (MSL3-K34, MSL3-K161, MSL3-K170, MSL3-K224). Most likely these lysines are exposed in flexible regions and can form transient contacts with multiple other regions in the ensemble. Interestingly, there was one link that crosslinked MSL3-K224:MSL3-K224, which is compatible with a dimer of the MSL complexes (Figure 18, red loop below MSL3).

In conclusion, the XL-MS of the 3-MSL confirmed known interactions and highlighted potential new interaction regions. Especially flexible regions were crosslinked, caused by their ability to form dynamic interactions, which can differ between different molecules in the ensemble. It would be interesting to validate the interaction regions by orthogonal structural techniques such as cryo-EM or by classical mutagenesis. In classical mutagenesis the most straightforward way is deletion to analyze whether an interaction region is critical or not.

5.8 Modelling the 3-MSL complex structure by AlphaFold-Multimer showed many violated XL distance restraints

During the course of this PhD project the protein structure prediction tool AlphaFold emerged and developed into a powerful tool [228]. I evaluated the structural models of the 3-MSL complex provided by Dr. Sebastian Eustermann (EMBL, Heidelberg) for their compatibility with my crosslinking data. The six best scoring models are overlapping in the core regions, which can be predicted with high confidence (blue, Figure 21A). However, the loop regions and IDRs are predicted with low confidence

Results

(red) and this leads to a 'spaghetti'-like appearance of the models (Figure 21A). If only the highest-ranking model was taken into consideration, the picture became much more ordered (Figure 21A). Subsequently, the low-confidence loops were excluded and the proteins were color-coded (Figure 21B). The MSL3-chromo domain is displayed detached from the core of the complex, because the linker of MSL3 is predicted to be unstructured and thus invisible. Next, the crosslinking data were compared to this model by Xlink analyzer, a toolkit developed for such comparisons in the chimera software (Figure 21C) [229]. A subset of the crosslinks could be mapped to the model (73 crosslinks). The crosslinks which were linked from several loop-like regions to the core of the complex, were mostly > 30Å distant, so called 'violated' crosslinks (red). Forty crosslinks were violated (54.8%). Nevertheless, 33 crosslinks were satisfied (45.2%) the predictions, (blue), with < 30Å distance.
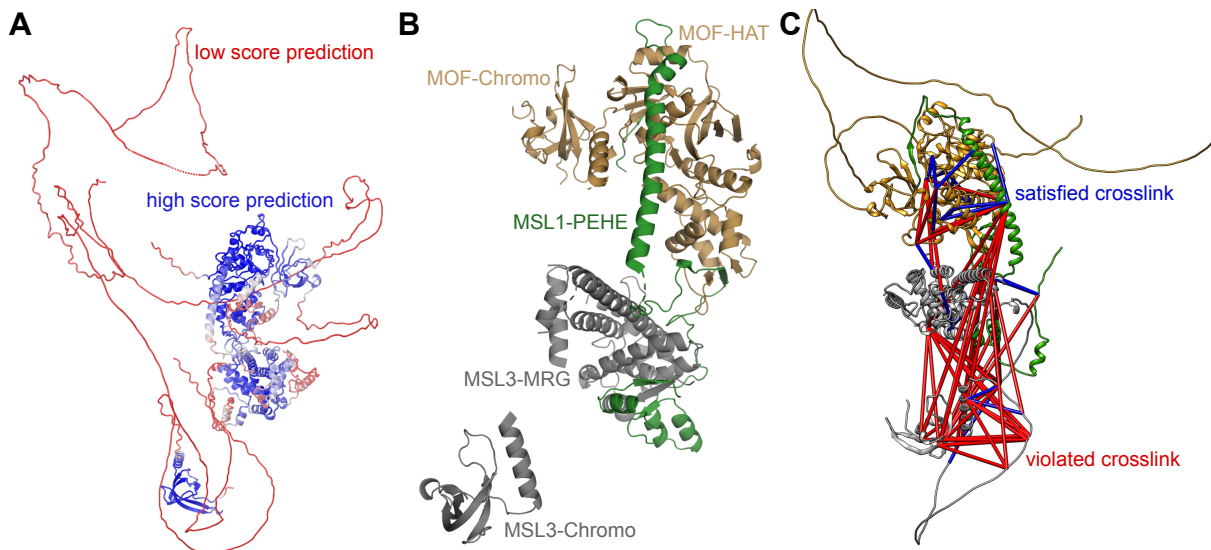
Figure 21: AlphaFold-multimer complex prediction for the 3-MSL complex performed by Dr. Sebastian Eustermann. **A** Best ranking model of the 3-MSL complex displayed with the b-factor coloring, blue: high confidence, red: low confidence. **B** Best ranking model displayed without any loops. Color coding of the amino acid chains MSL1 green, MSL3 grey and MOF sand. The MSL3-Chromo domain seems detached from the rest of the complex, because it is connected by a linker, which has a low prediction score and is hidden in this display. Visualization by pymol (Schrödinger). **C** Crosslinks of the 3-MSL analyzed by Xlink analyzer and ChimeraX. Red: violated crosslinks > 30Å distance, blue: satisfied crosslinks < 30Å distance.

Further, the crosslinks were categorized into three classes and compared to the model without the loops, because the loops led to many violated crosslinks (Figure 22). In the first panel all 31 crosslinks across two independent experiments were analyzed. Of these crosslinks, 17 were termed 'satisfied' (54.8%) and symbolized by a blue stick. Fourteen crosslinked were classified 'violated' (45.2%) and symbolized by a red stick. The middle panel shows the crosslinks that are found at least twice in the same experiment and can be considered technical replicates. The last panel shows only the reproduced crosslinks between two independent replicates, which are only four crosslinks in total. Three of them were satisfied and one was not (Figure 22). Especially, the crosslinks to the MSL3 chromo domain were violated, pointing out that this domain is potentially inadequately positioned.

In conclusion, many crosslinks were violated by the AlphaFold-Multimer structure predictions. The violation of many distance restraints indicates that the model does not correctly predict the conformation

in some regions. To improve the model the crosslinking data were included into the modelling process, which gave a different and potentially more realistic structural model. Regardless of all modelling strategies, the structure of the complex should still be investigated by structural biology techniques, such as crystallography or cryo-EM.
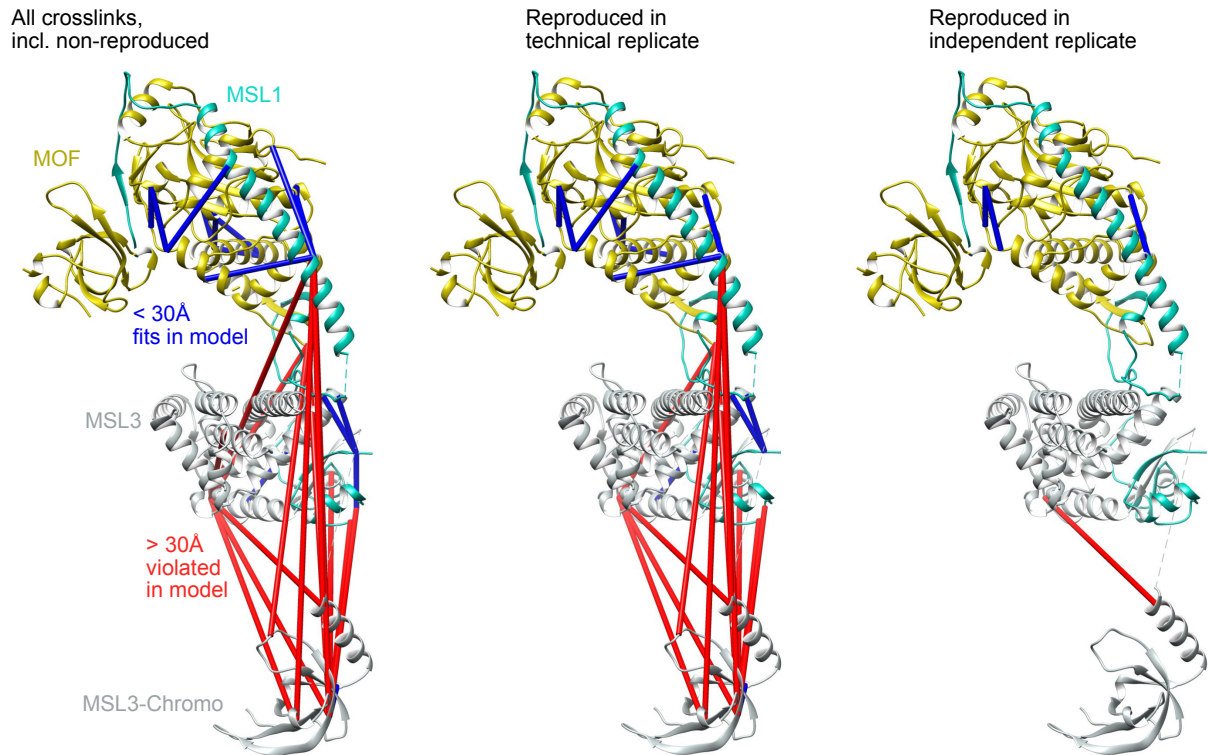
Figure 22: Comparison of crosslinking MS data with the AlphaFold-Multimer prediction model provided by Dr. Sebastian Eustermann (EMBL Heidelberg). Model displayed without loops for clarity. Fewer crosslinks are located in the regions without loops, yet still many of them remain violated. If only the crosslinks reproduced in an independent replicate are considered, only three satisfied intra-molecular crosslinks in the MOF HAT domain and one violated crosslink within MSL3 remain. Visualization by ChimeraX and Xlink Analyzer [228-231].

5.9   The 4-MSL core complex shows both expected and novel interaction regions in XL- MS

After analyzing the 3-MSL complex, I aimed to analyze the 4-MSL complex (MSL1, MSL2, MSL3 and MOF) by XL-MS. Within the 4-MSL complex not much is known about the potential interaction interfaces between MSL2, MSL3 and MOF. This subcomplex is another step towards the overarching goal of reconstituting the full complex, containing all MSL proteins and the roX RNA. It is also an interesting subcomplex to study, because it contains additionally the male-specific subunit MSL2, which can bind DNA, RNA and has enzymatic activity as a ubiquitin ligase.

Crosslinking was performed after FLAG affinity purification of the 4-MSL complex, bound to the FLAG-beads. One millimolar of BS3 was used, the peptides were obtained by tryptic digest and mass

spectrometric analysis was performed the same way as for the other MS experiments of the 2-MSL and 3-MSL complexes.
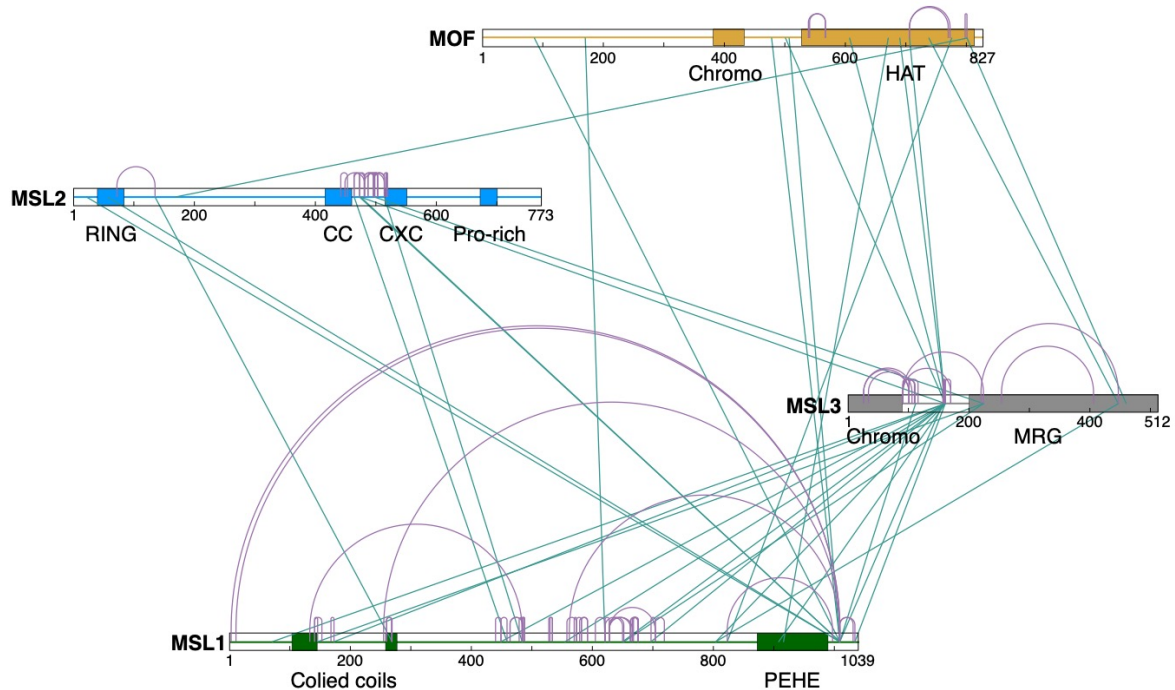
Figure 23: Crosslinking MS of the 4-MSL complex with the four core protein subunits MSL1, MSL2, MSL3 and MOF detected known and novel interaction sites by 194 crosslinks. Purple arc: intramolecular crosslink, green line: intermolecular crosslink. Protein domains are indicated by color-coded boxes. Visualization by xiNet (https://crosslinkviewer.org, 19.10.2023, [226]).

In the 4-MSL complex 194 crosslinks were detected (Figure 23, Appendix 12.2.5). There were several known interactions confirmed, these include the MSL1-PEHE domain interactions with MOF-HAT and MSL3-MRG and the MSL1-coiled coil interaction with the MSL2-RING domain. These already published interaction domains are also presented in the previous chapters. In the 4-MSL complex the number of crosslinks identified in those regions is lower, however this might be due to occlusion of lysines by other proteins. Alternatively, it might be due to the detection limit of the mass spectrometry. Despite the use of purified proteins, contaminants and un-crosslinked peptides could saturate the MS column and detector, albeit in the range of amount used (40-50 μg of protein) it is unlikely. Lastly, crosslinked peptides might be missed during MS detection due to stochastic variations in ionization of some peptides.

Nevertheless, interesting novel contact sites were found. In particular, previously unknown interactions between MSL2 and MOF or MSL2 and MSL3 were identified. One crosslink was found between MOF and MSL2 (MOF-K801:MSL2-K168). MOF-K801 is located in its HAT domain, whereas MSL2-K168 is located in a linker region between the N-terminal RING domain and the predicted coiled coil. Previously it was unknown, if MSL2 or MOF would directly interact and in which domain of the respective proteins. However, due to the long linker length of BS3 (30 Å), the crosslink could also indicate spatial proximity without direct interaction.

Results

The MOF-HAT showed eight crosslinks to MSL1, which link the MOF HAT domain with the N-terminal PEHE domain of MSL1, as observed for the 3-MSL complex. Two crosslinks N-terminal of the HAT domain link to the C-terminus of MSL1 (MOF-K478:MSL1-K1009 and MOF-K507:MSL1-K1009). One crosslink from the MOF N-terminus links to the IDR region of $MSL1_{K621}$ (MOF-K170:MSL1-K621). Another crosslink links the MOF N-terminus to the MSL1 C-terminus (MOF-K89:MSL1-K1009).

Nine crosslinks were detected between MSL3 and MOF. All of them were in the MOF-HAT domain or adjacent to the HAT domain. In MSL3 only three lysine residues were crosslinked to MOF (MSL3-K161, MSL3-K447 and MSL3-K460). The first one of them MSL3-K161, is located in the flexible linker region of MSL3 between the N-terminal chromo domain and the C-terminal MRG domain. The other two lysine residues are in the MRG domain of MSL3, which is known to interact as well at the MSL1 PEHE domain, just as MOF does. These MSL3-MOF crosslinks encompass both spatial proximity and direct interactions.

Moreover, the expected MSL3-MRG to MSL1-PEHE interaction was detected again by one crosslink (MSL3-K447:MSL1-K804). Several other crosslinks to the MSL1-IDR were found again. Curiously, these crosslinks originate from only four lysine residues of MSL3, again the highly reactive MSL3-K161, MSL3-K224, MSL3-K112 and MSL3-K157 (next to K161). The linker region of MSL3 seems to be highly dynamic and flexible, allowing for such a high number of different crosslinks. Both MSL4-K161 and MSL3-K224 seemed to be highly reactive ("hyperreactive"), which hints to a more transient and dynamic interaction than to a stable structural interface. Unexpectedly, two crosslinks were detected on these same lysine residues of MSL3 to MSL2, connecting to the similarly reactive "pre-CXC" domain of MSL2. If this interaction is direct or if the crosslink occurred due to spatial proximity, remains to be validated by other methods, e.g., mutagenesis and deletion co-IP assays.

Furthermore, MSL1 showed fewer intra-molecular crosslinks than in the smaller subcomplexes of the 2-MSL and 3-MSL complex, only 76, instead of 107 in the 3-MSL (Figure 17) or 113 in the 2-MSL (Figure 18).

One hypothesis drawn from the observation of reduced crosslinks in MSL1 assumes a more folded structure of MSL1. By integration of more subunits into the MSL complex, MSL1 might engage in more protein-protein interactions and change its extended conformation.

Lastly, MSL1 and MSL2 were found to have three crosslinks in the expected MSL2-RING to MSL1-coiled coil region (MSL2-K135:MSL1-K269). Interestingly, two crosslinks formed between MSL2-RING domain and the MSL1-C-terminus (MSL2-K22:MSL1-K1009 and MSL2-K57:MSL1-K1009). Four more crosslinks connected the MSL1 IDR to the MSL2 pre-CXC region.

In summary, XL-MS identified many previously unknown potential interaction surfaces within the MSL complexes that should be validated by applying other biochemical or structural biology techniques. Certain regions with many or novel crosslinks could be deleted or specific lysine residues could be mutated, followed by co-immunoprecipitation assays. Moreover, applying another crosslinking reagent

with a shorter or even no linker length could highlight the sites that are actually contact surfaces. On the other hand, it would be a great advancement in the field to solve the structure of the MSL subcomplexes or the full MSL complex by cryo-EM, which would be suitable for the size of the complexes. This way the interactions found by crosslinking MS could be validated, if they are direct interaction sites or just coincidental by spatial proximity.

5.10 Conformational changes within MLE upon RNA binding

The XL-MS approach was applied to study conformational changes within MLE on different RNA substrates (double-stranded dsRNA and single-stranded ssRNA) in a collaborative project with Dr. Pravin Ankush Jagtap and Prof. Dr. Janosch Hennig (EMBL Heidelberg/University Bayreuth). They provided purified full-length MLE and RNAs of interest for their structural studies. In parallel, I applied the XL-MS methodology to the MLE-RNA complexes. As MLE substrates the dsRNA a modified stem-loop 7 of roX2 (SL7) and as ssRNA a 10-mer repeat of uridine (U_{10}) were used. Crosslinking was performed in solution with 40 µg of MLE using 1 mM of BS3 crosslinker. Mass spectrometry analysis followed the established protocol similar to the MSL complexes. The data were published together with the cryo-EM and NMR structural data Jagtap *et al.*, 2023 [118] (Figure 24A, B). The experiment was replicated at least twice for MLE full-length without RNA, in the presence of a 10-mer poly-uridine RNA (U_{10}), stem-loop 7 double-stranded roX2 RNA with a modified uridine-rich single-stranded overhang (SL7-mod) (Figure 24C). To prevent ATP-hydrolysis and MLE dissociation after RNA remodeling, crosslinking reactions were conducted in the presence of 1 mM AMP-PNP, a non-hydrolysable ATP transition state analog.

In the NMR data interactions between the dsRNA binding domains dsRB1 and dsRB2 as well as the helicase core domain were observed, which could not be seen in the cryo-EM structure of the RNA-free MLE helicase. Moreover, the dsRB1 was not observed at all in the cryo-EM structure, which indicates that it might have dynamic and flexible conformations.

XL-MS was applied to underpin the findings on interaction between dsRB1, dsRB2 and the helicase (Figure 24C). In the RNA-free state, seven lysines from dsRB1 and a single lysine from the first linker region crosslink with eight lysines from the helicase module and the C-terminal G-rich region. However, dsRB2 lysine residues only crosslinked with three lysines located in HA2 (K936, K1020 and K1081), OB-like and linker 3 regions of the helicase module. Two of these lysines are within crosslinking distance of the previously known binding site of dsRB2 (K1020 and K1081), as demonstrated by the cryo-EM structure of MLEΔG (Figure 24D, E). In the presence of U_{10} ssRNA or SL7 dsRNA, the dsRBD1-linker formed fewer crosslinks, but with a similar distribution. The decrease in numbers of crosslinks could be due to a shielding or quenching effect if the crosslinker cannot reach the lysine residues occupied by RNA or when RNA itself can react with BS3. On the other hand, K256 (and neighboring K255, K254 and K253) from dsRB2 formed specific interdomain crosslinks with K1020 of the OB-like domain (Figure 24F). From the observed cryo-EM structures, these crosslinks are possible

if MLE binds to either ss or dsRNA. Therefore, these data are in agreement with the cryo-EM structures and suggest that dsRB1-linker interacts non-specifically with the MLE core independent of the presence of RNA and adenosine nucleotides, whereas dsRB2 has limited flexibility in the absence of RNA and is attached to the core of MLE. Upon RNA binding dsRB2 changes conformation and assumes a specific structure.

In contrast to previous findings that only dsRB2 interacted with the MLE helicase module, the NMR titration, XL-MS analysis, and cryo-EM structures revealed that dsRBD1 also interacts transiently and non-specifically with the helicase module. Although dsRB1 is essential for proper MLE localization with the MSL complex [135, 136, 138], the recent data suggest a potential role for dsRB1 in mediating protein-protein interactions during DCC assembly, possibly involving the MSL1/MSL2 module [143].
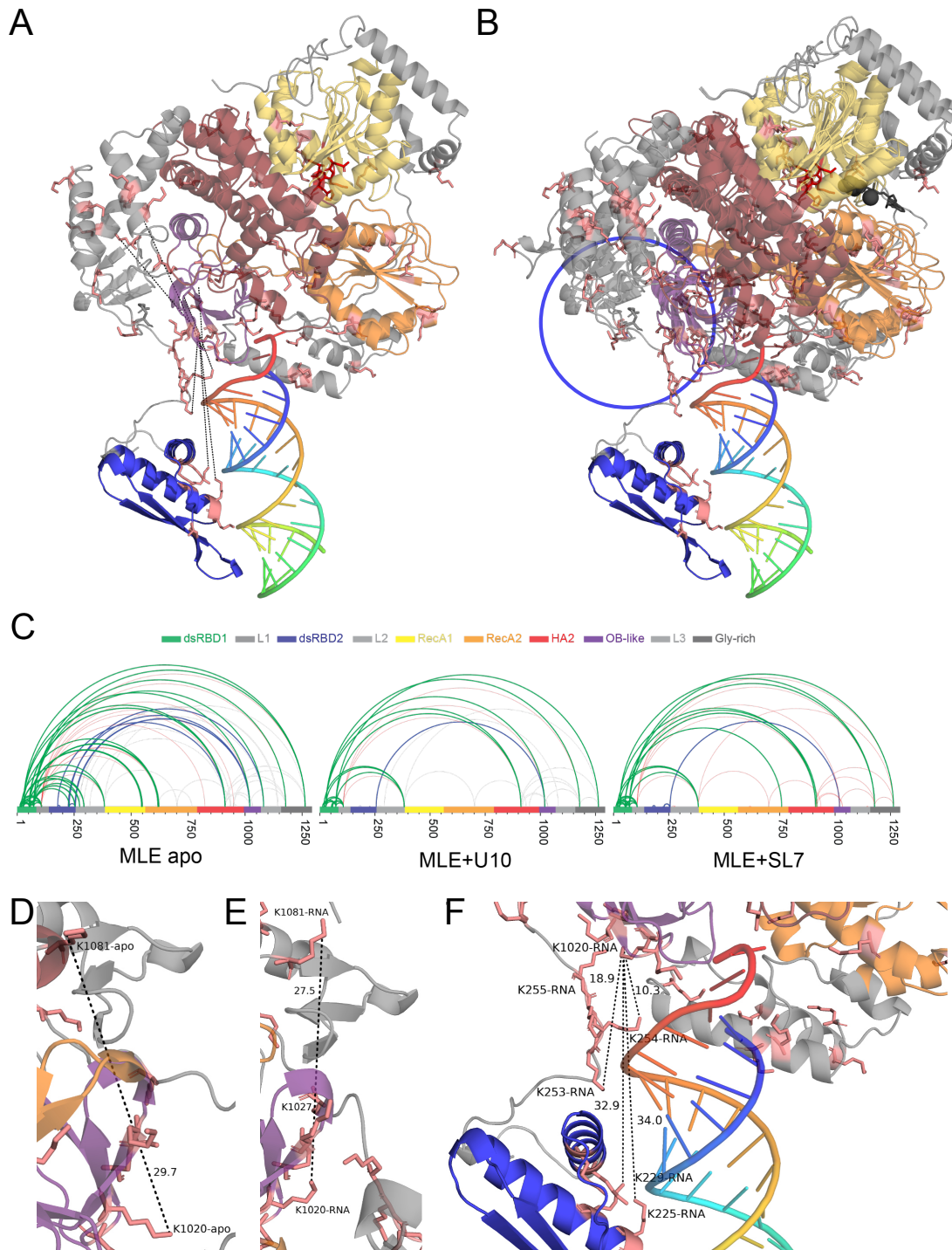
Figure 24: MLE undergoes conformational changes upon binding to dsRNA. **A** Cryo-EM model (PDB: 8B9K) depicts SL7 bound to MLEΔG, with domain coloring following specified code in (C). Lysine residues are highlighted in pink, and selected crosslinks are represented by dark dotted lines (detailed in D-F). **B** Cryo-EM structures comparing the dsRNA-bound conformation with the RNA-free conformation (PDB: 8B9J), revealing similar superposition but notably different positioning of dsRB2. **C** Interdomain crosslinks within MLE in different states (apo, U10 ssRNA-bound, and SL7 dsRNA-bound) are presented as arches. **D-F** Selected crosslinks in cryo-EM structures illustrate distances between selected lysines. **D** K1020 and K1081 apo-state and **E** dsRNA-bound state, **F** violations of the 30 Å distance restraint in the RNA-bound conformation. (Figure adapted with permission from Jagtap et al. [118]).

5.11 Conformational changes in MSL1-MSL2 with MLE and roX2 RNA

One overarching aim of this project was to identify and characterize the role of roX RNA within the MSL complex. One hypothesis in the field revolves around conformational and consequential functional changes upon roX integration into the MSL complex. I aimed to approach the question of conformational changes upon roX integration by XL-MS. To simplify the task, the 2-MSL complex was used, with fewer subunits the detection of crosslinks might be less technically and computationally challenging. Moreover, it was shown that the 2-MSL subcomplex is the required subunit of the MSL complex to incorporate roX RNA efficiently [143]. For the proper binding and integration, MLE and ATP are necessary [143]. I used roX2, the shorter one of the two roX RNAs, of a length of 552 nucleotides. Purified 2-MSL complex was incubated with MLE and ATP in absence or presence of equimolar amounts of roX2 RNA relative to the 2-MSL complex.

58 crosslinks with an FDR below 0.2 were identified (Figure 25A, Appendix 12.2.6). This FDR was high in comparison to other experiments; however, the crosslinks were inspected manually and the quality standards of the mass spectrometry fragmentation patterns were met. MLE showed only four intra-molecular crosslinks, two in the RNA binding domain 1 (RB1), and one in RNA binding domain 2 (RB2) and one in the linker C-terminal of RB2 (Figure 25A). Three crosslinks connected MLE to MSL2, all of them mapping to the pre-CXC domain (MSL2-K487). One crosslink was formed with the linker region of MLE between RB2 and the helicase domain and further two crosslinks were found in the C-terminal part of the helicase domain. There were six crosslinks found between MLE and MSL1. Five out of these six crosslinks were located in the flexible IDR of MSL1which crosslinked with different domains of MLE. RB2 of MLE crosslinked as well to the IDR of MSL1. In addition to that, MLE HA-domain and OB-like domains crosslinked to MSL1. Another crosslink was identified between MSL1-N-terminus and MLE-RB1 (MSL1-K48:MLE-K53).

Within MSL2 11 intra-molecular crosslinks were identified. Four crosslinks were detected between MSL2 and MSL1, in this experiment only the pre-CXC region of MSL2 crosslinked to the MSL1 IDR. The connection between the RING domain of MSL2 and the coiled coil of MSL1 could not be found under the experimental conditions. MSL1 showed 27 intramolecular crosslinks. One of the crosslinks crosslinked the same lysine residue with itself (MSL1-K704:MSL1-K704), supporting to the hypothesis that there might be a dimer of the 2-MSL complex.

Results

A No RNA added
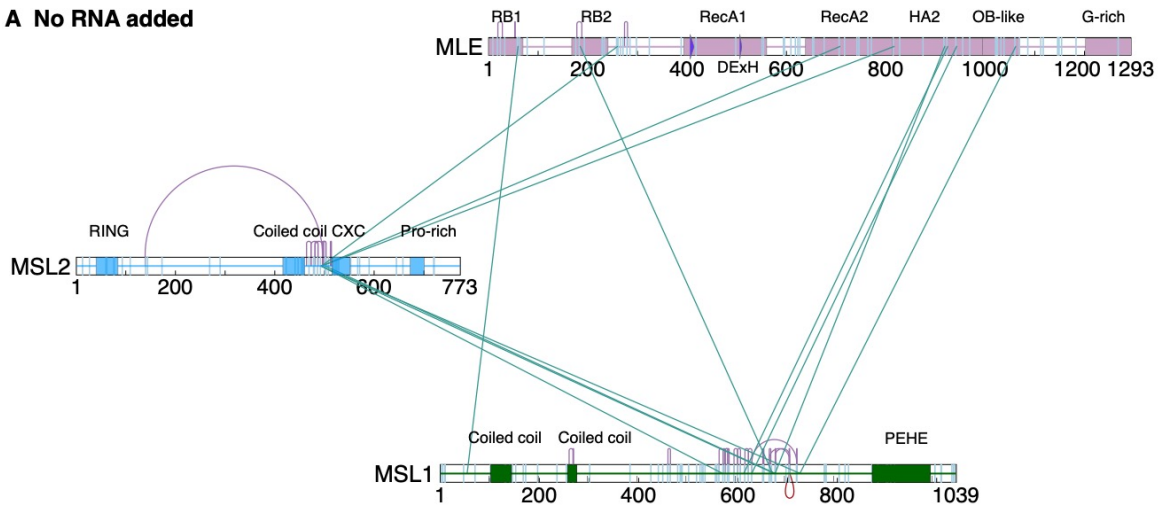
B 1:1 roX2 : 2-MSL
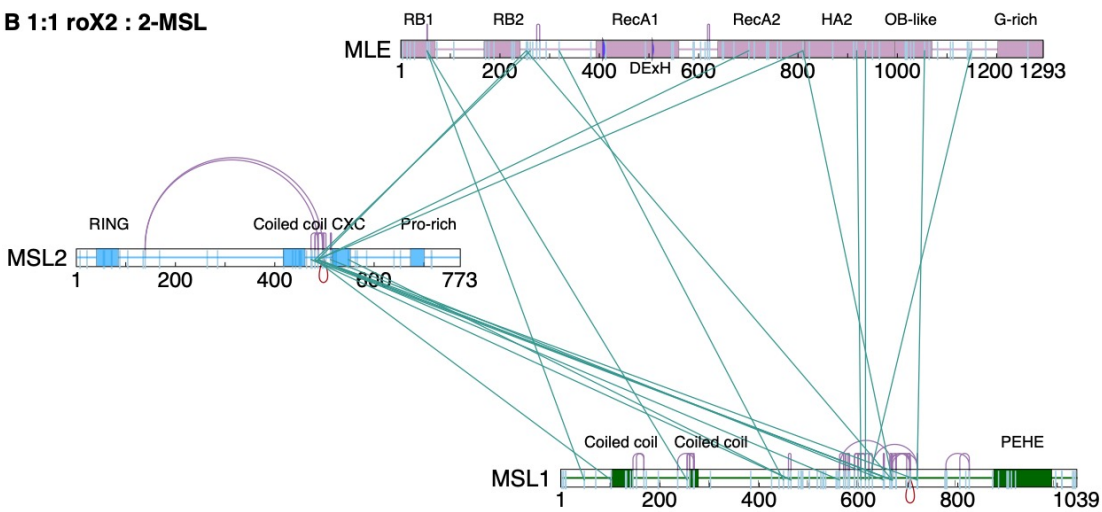
C 10:1 roX2 : 2-MSL
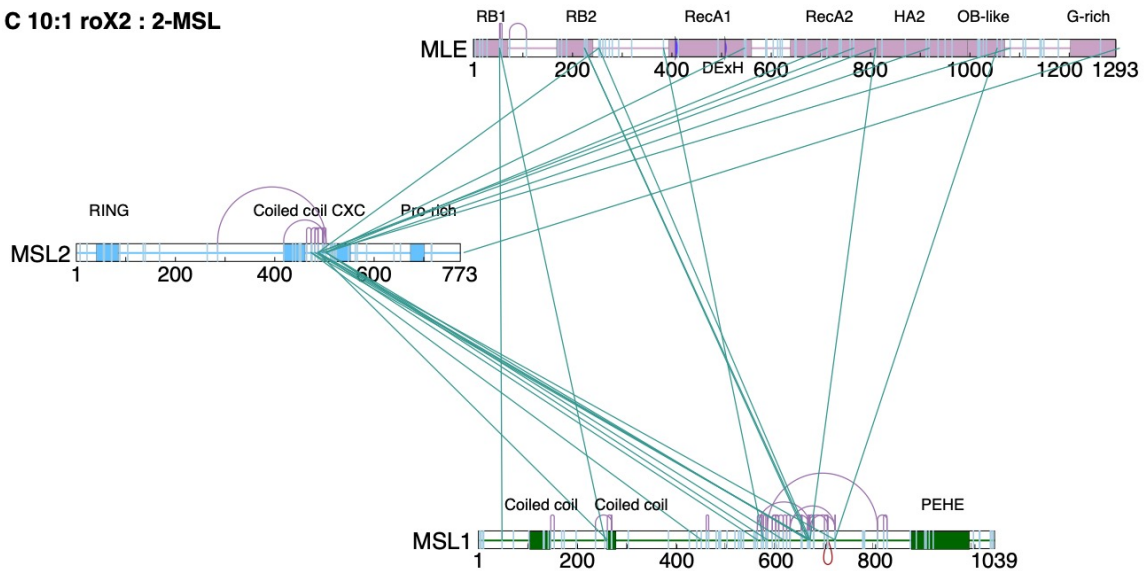
Figure 25: MSL1ΔC-MSL2 crosslinked with equimolar amounts of MLE in presence of ATP. Purple arches represent intramolecular crosslinks, green lines represent intermolecular crosslinks, light blue line represent lysine residues, red loop represents a crosslink at the same position, e.g., in case of dimers. **A** Without RNA, **B** equimolar roX2 RNA added, **C** 10-fold roX2 RNA added. Visualization by xiNet (https://crosslinkviewer.org, 24.10.2023, [226]).

Next, the 2-MSL and MLE reaction was incubated with two different amounts of *in vitro* transcribed roX2 RNA. In the first setup, the 2-MSL complex was incubated with a 1.3-fold excess of roX2 (for

Results

simplification called 1:1, Figure 25B). In parallel, the roX2 RNA was added in another reaction to 10-fold excess (Figure 25C).

In the equimolar stoichiometry setup (1:1), 94 total crosslinks were detected (Figure 25B, Appendix 12.2.6). In MLE, only three intramolecular crosslinks were found with the FDR of 0.2. One crosslink in RB1 and one in the linker region between RB2 and the helicase core were the same as in the absence of roX2. The third crosslink in the helicase domain in between the two helicase "lobes" was not detected in the previous sample, however it is only separated by few amino acids in the polypeptide chain such that spatial proximity is obvious and a crosslink can be expected (MLE-K618:MLE-K622). Four crosslinks were found between MLE and the MSL2 pre-CXC domain in presence of roX2. One additional crosslink was found between the MLE linker close to the RB2 and the MSL2 pre-CXC (MLE-K261:MSL2-K481). Between MSL1 and MLE 12 crosslinks were found, which is twice the number of crosslinks than were detected in the absence of RNA. Nevertheless, the overall pattern did not change much, seven crosslinks were found connecting the MSL1-IDR to MLE. Under 1:1 roX2 condition, two crosslinks between the MSL1-N-terminus and the MLE-RB1 were detected as well as before without RNA added. Moreover, the MSL1-coiled coil crosslinked to the MLE-RB1, which did not occur in absence roX2 RNA.

In MSL2 14 intramolecular crosslinks were detected, three more than in the absence of roX2. Eleven crosslinks were found between MSL1 and MSL2. The same crosslinks were identified between MSL1 and MSL2 earlier, in absence of roX2 RNA, connecting the MSL2-pre-CXC domain and the MSL1-IDR. One additional crosslink was found within the MSL2-CXC in the MSL1-IDR (MSL2-K548: MSL1-K464). Moreover, the MSL1-coiled coil region crosslinked with the MSL2 pre-CXC (MSL1-K102:MSL2-K481). Interestingly, in MSL2 a same residue loop link was identified at MSL2-K498 (red loop below the sequence, Figure 25B).

In MSL1 48 intramolecular crosslinks were identified, also twice as many as without RNA, which showed 27 crosslinks. Possibly, the 2-MSL complex can dimerize under these conditions, even though the interaction might be transient.

Stoichiometric addition of roX2 RNA did not majorly change the overall interactions. At least, considerable changes were not found in the crosslinking MS data. This led to the hypothesis, that the majority of MSL complexes did not bind or incorporate roX2 properly, because the reaction was not saturated. As a solution, the addition of roX2 in greater excess was applied. The changes in crosslinking patterns were hypothesized to be more pronounced if most MSL complexes would be saturated with roX2 RNA. Thus, roX2 RNA was added in 10-fold molar excess.

The 10-fold roX2 excess did not lead to the hypothesized drastic changes. Overall, 91 crosslinks were detected, almost the same number of crosslinks as in the 1x roX2 condition, 94 crosslinks (Figure 25B, C, Appendix 12.2.6). In MLE five crosslinks were identified, yet only three differed from the 1:1 ratio condition. These three crosslinks were found in the MLE-RB1 and the N-terminal MLE-linker (Figure 25C).

Results

More crosslinks were found between MSL2 and MLE, potentially indicating that more molecules interacted in presence of more RNA and thus the crosslinks were detected. The crosslinks were found in the same regions as in the 1:1 roX2:MSL ratio sample, namely, the MSL2-pre-CXC and the MLE-RB2. Moreover, crosslinks were found, which linked the MSL2 pre-CXC region connecting to the HA2 domain and to the OB-like domain of MLE. An additional crosslink was found between the MLE-RecA1 domain, which harbors the ATP binding site and the DExH site, to the MSL2 pre-CXC. The MSL2 lysine residues, which crosslinked to the various different domains of MLE, are always the same ones (MSL2-K487, MSL2-K496 and MSL2-K503). These sites are potentially highly reactive and might be dynamic and flexible.

Between MSL1 and MLE 11 crosslinks were detected, which is similar in the equimolar roX2 and 10-fold roX2 concentration scenario (Figure 25B, C). The crosslinks connected the MLE helicase domain RecA1 to the MSL1-IDR (MLE-K383:MSL1-K574).

Within MSL2 there were 13 intramolecular crosslinks formed, which differed from the crosslinking pattern under the equimolar roX2 or the no RNA conditions. The long-range interaction between the long linker region C-terminal of the RING domain with the pre-CXC region changed to a shorter distance crosslink (MSL2-K285:MSL2-K503). The intramolecular crosslink to the same lysine residue in MSL2 was not observed under the high roX RNA ratio.

In summary, the changes observed by the addition of roX2 to the complex were subtle. Thus, either the conformational change is subtle, or there might not be a conformational change upon roX2 RNA binding. However, the roX2 RNA *did* lead to a few changes in the patterns and an overall increase in crosslinks, which might hint at a stronger or closer interaction. The incorporation efficiency might be low despite roX2 being added as a 10-fold excess over the 2-MSL complex and MLE. Due to the technical limitations of roX2 RNA *in vitro* transcription, the scale-up was limited, such that no wide-range titrations or even higher ratios were feasible.

The identified crosslinks need to be validated by at least one orthogonal method. Structural techniques, such as cryo-EM or crystallography could be an option. Moreover, *in silico* predictions, such as AlphaFold-Multimer for complexes could serve as valuable sources of information. Alternatively, mutants can be created to assess functional changes in the MSL complexes and RNA incorporation by MLE. The mutants can be evaluated through classical immunoprecipitation assays.

5.12 Optimization of roX2 incorporation efficiency into the MSL complex

To further investigate structural and conformational changes of the 2-MSL complex or larger MSL (sub)complexes it is necessary to optimize the efficient integration of the roX-RNA into the complex. For easier *in vitro* handling, the experiments focused on roX2 for its shorter length. One straight-forward idea is to integrate the RNA already during the expression in the Sf21 cells. Another approach is to add the RNA during the purification process. The third option, which is the one already presented in the previous chapter, is to just add roX2 to the purified 2-MSL complex in the presence of the helicase MLE and ATP.

The prerequisite for co-expression is a baculovirus that expresses roX2 RNA. The baculovirus for co-expression of the 2-MSL complex (MSL1ΔC-MSL2-FLAG) with MLE and roX2 was were already prepared by Dr. Marisa Müller and Silke Krause. The expression was performed as described for the wild-type MSL proteins and the purification was performed from whole cell lysate and from nuclei, respectively. Furthermore, different buffers with varying salt and glycerol were assessed (Figure 26A, B).
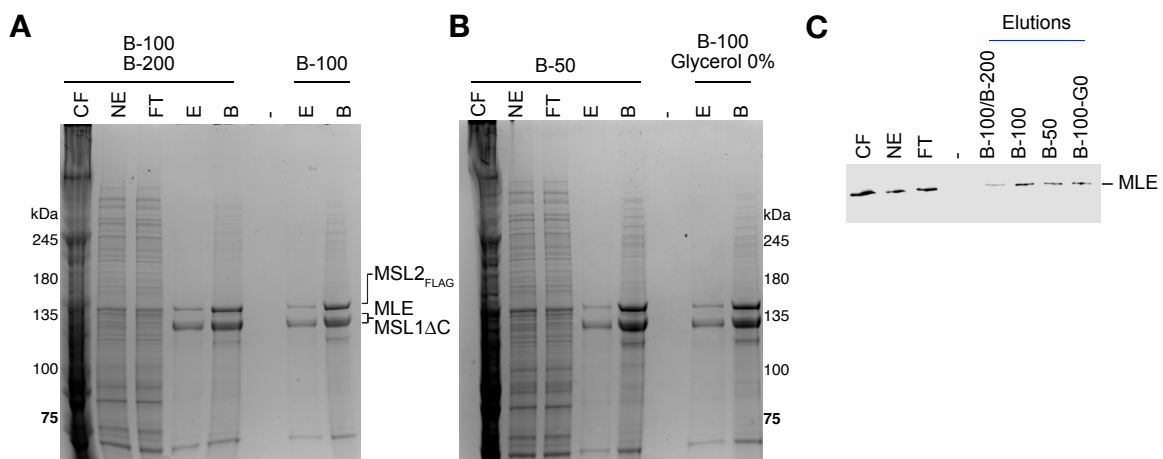
Figure 26: Test purification by nuclear extraction of the co-expression of MSL1ΔC-MSL2-FLAG with MLE and roX2 RNA. **A** Coomassie stained 8% acrylamide gel, purification with buffers containing 100 mM and 200 mM KCl (B-100 with 100 mM and B-200 200 mM KCl) or just one buffer. **B** Coomassie stained 8% acrylamide gel, purification with buffers with 50 mM KCl or 100 mM KCl and no glycerol. **C** Western blot probed with anti-MLE antibody (rat, 6E11) and secondary antibody anti-rat 800 nm (LiCOR). CF: cytoplasmatic fraction, NE: nuclear extract, FT: flow through of FLAG beads, E: elution, B: beads.

The purification tests after nuclear extraction demonstrate that MLE can be co-purified with the 2-MSL complex. However, in the Coomassie-stained SDS-PAGE gels a double band is visible for MSL1ΔC and MLE at the same molecular weight. To confirm that MLE is present in the co-purification, a Western blot against MLE of the elution fractions was performed (Figure 26 C). The double band is not well separated for a good quantification from the SDS-PAGE gel image, yet MLE seems to be very lowly abundant in the eluted fractions. Next, to improve MLE yield, I included *in vitro* transcribed roX2 RNA during the affinity purification.

Results
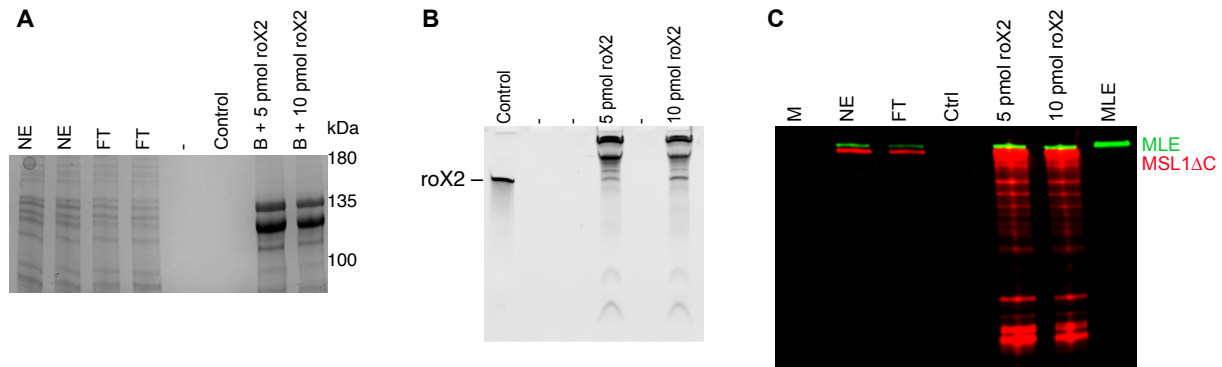
Figure 27: Purification test with roX2 RNA added to the MSL1ΔC-MSL2-FLAG MLE roX2 purification. **A** Coomassie stained 8% acrylamide SDS-PAGE gel of different purification steps. **B** 5% acrylamide and 6 M urea gel of the RNA co-purified after the spike-in of roX2 to the incubation with the FLAG agarose. Staining with ethidium bromide. **C** Western blot probed with an anti-MLE antibody (rat, 6E11) and an anti-MSL1 antibody (rabbit). Image taken by LiCOR. NE: nuclear extract, FT: flow through, B: beads, Ctrl: control roX2 RNA in buffer on FLAG beads.

In vitro-transcribed roX2 RNA was added to the nuclear extract of the 2-MSL/MLE co-expressing Sf21 cells. The complex was FLAG-purified and protein content was analyzed both by Coomassie-stained SDS-PAGE (Figure 27A) and Western blotting (Figure 27C), RNA content was analyzed after IP and proteinase K digest by denaturing PAGE and ethidium bromide staining (Figure 27B).

Coomassie staining of the SDS-PAGE gel showed wide band at 130 kDa, which is likely a not separated double band of MSL1ΔC and MLE-FLAG (Figure 27A). MSL2-FLAG runs at the apparent molecular weight of 135 kDa. roX2 RNA is co-purified efficiently with the bead fraction (Figure 27B). However, even after a 30 min proteinase K digest it shifts up on the gel, seemingly still bound by proteins. In the ethidium bromide-stained PAGE gel, there were two prominent bands visible running at a larger molecular size than the roX2 RNA. This hints to two major stable complexes that are not resolved well by the gel and they are not properly denatured, despite 6 M urea in the gel. One of them can be roX2 bound by MLE and the other one could be the roX2 RNA bound by the 2-MSL complex. Moreover, several degradation bands of the RNA are visible, indicating that despite careful handling RNases potentially contaminated the sample.

To distinguish MLE and MSL1ΔC a Western blot was performed of the purification steps (Figure 27 C). In the nuclear extract (NE) and flow through (FT) fractions both proteins are clearly detectable by their specific antibodies as a prominent band. However, in the samples of the spike-in roX2 bound to the FLAG beads MSL1ΔC shows a lot of bands at smaller molecular weight, indicating protein degradation despite including protease inhibitors during the protein purification and elution.

As a conclusion, 2-MSL, MLE and roX2 RNA can be co-purified. However, the stoichiometry is unclear and addition of excess roX2 was necessary to detect robust amounts of roX2. Potentially the Sf21 cells did not express roX2 to a detectable degree or it was not stable and degraded within the Sf21 cells. Since substantial amounts of properly reconstituted complex are required for the XL-MS, these yields were considered too low to further pursue them for the crosslinking sample preparation. Notably, in this approach ATP was not included, which might be necessary for a proper integration of the roX2 RNA into the 2-MSL complex.

Consequently, the larger complexes were not tested for XL-MS in presence or absence of roX2 RNA as proper integration into the complex was not quantitatively established.

5.13 Validation of interactions found by XL-MS applying deletion mutation and immunoprecipitation of MSL2 and MLE

Since MLE can be co-purified with MSL1 and MSL2 (Figure 26 and Figure 27) and many intermolecular crosslinks found in MLE were located at the RB1, I attempted to validate the crosslinking interaction sites in immunoprecipitation assays with mutant proteins. The putative interaction domains were deleted to create MLE-ΔRB1 and MSL2-Δpre-CXC (Figure 28A). Of MSL2 two mutants were created, the MSL2-ΔpreCXC domain (MSL2-Δ470-524), and the MSL2-ΔCC-preCXC domain (MSL2-Δ424-524). These mutants were expressed in baculovirus-infected Sf21 cells and purified by FLAG affinity chromatography (Figure 28B). Three deletions of the RB1 were created, the first one MLE-ΔRB1-70 deleted just the RB1 [52, 138], the second one MLE-ΔRB1-84 and the third one MLE-ΔRB1-86 [232]. All three mutants were cloned, expressed and purified (Figure 28A).

Next, it was analyzed, whether the mutations disturbed the interactions of MLE with the 2-MSL subcomplex. To study the interaction of MLE and the 2-MSL, the integration of roX2 RNA into the complex was scored after anti-MSL1 immunoprecipitation (IP). The 2-MSL complex, roX2, ATP and MLE sample were incubated, then the reaction was probed by anti-MSL1 antibody-coupled beads (Figure 28 B (input), C). The anti-MSL1 antibody-coupled beads were washed and the bound proteins were analyzed by Western blotting, as the proteins run at similar positions in the gel (Figure 28 C). The membrane was probed by the primary antibodies against MLE, FLAG (all proteins contain a FLAG-tag) and MSL2 (co-staining with MLE). MSL proteins and MLE were sticking to the beads in the absence of MSL1 (Figure 28 C, lane 2), giving a strong background signal. The MLE-ΔRB1-84 and MLE-ΔRB1-86 did not stain well with the anti-MLE antibody (Figure 28 C, lane 14-17). It seems as if there is less MLE co-immunoprecipitated. In the case of the MSL2-Δ407, which is the MSL2-N-terminus until aa 407, there is notably less MLE pulled down. MLE seems to interact with MSL2 more C-terminally and this interaction might be also required for the interaction with MSL1.

After the IP, the RNA was extracted after proteinase K digestion and phenol-chloroform extraction. The bound RNA and the input samples were visualized on a denaturing acrylamide urea gel (Figure 28 D). The negative control reaction, where MSL1 was omitted, showed roX2 clearly in the input sample (Figure 28 D, lane 1), However no roX2 in the IP sample (lane 2). The positive control sample, where all proteins were wild-type, showed a clear band for roX2 after the IP (lane 4). None of the mutants tested showed any roX2 band in the IP samples. This appears as if the RNA cannot be incorporated by the mutants properly. Whether this has to do with the interactions between the proteins or if the deletions disrupt the RNA binding capacity of the respective protein *per se*, remains to be clarified.

Results
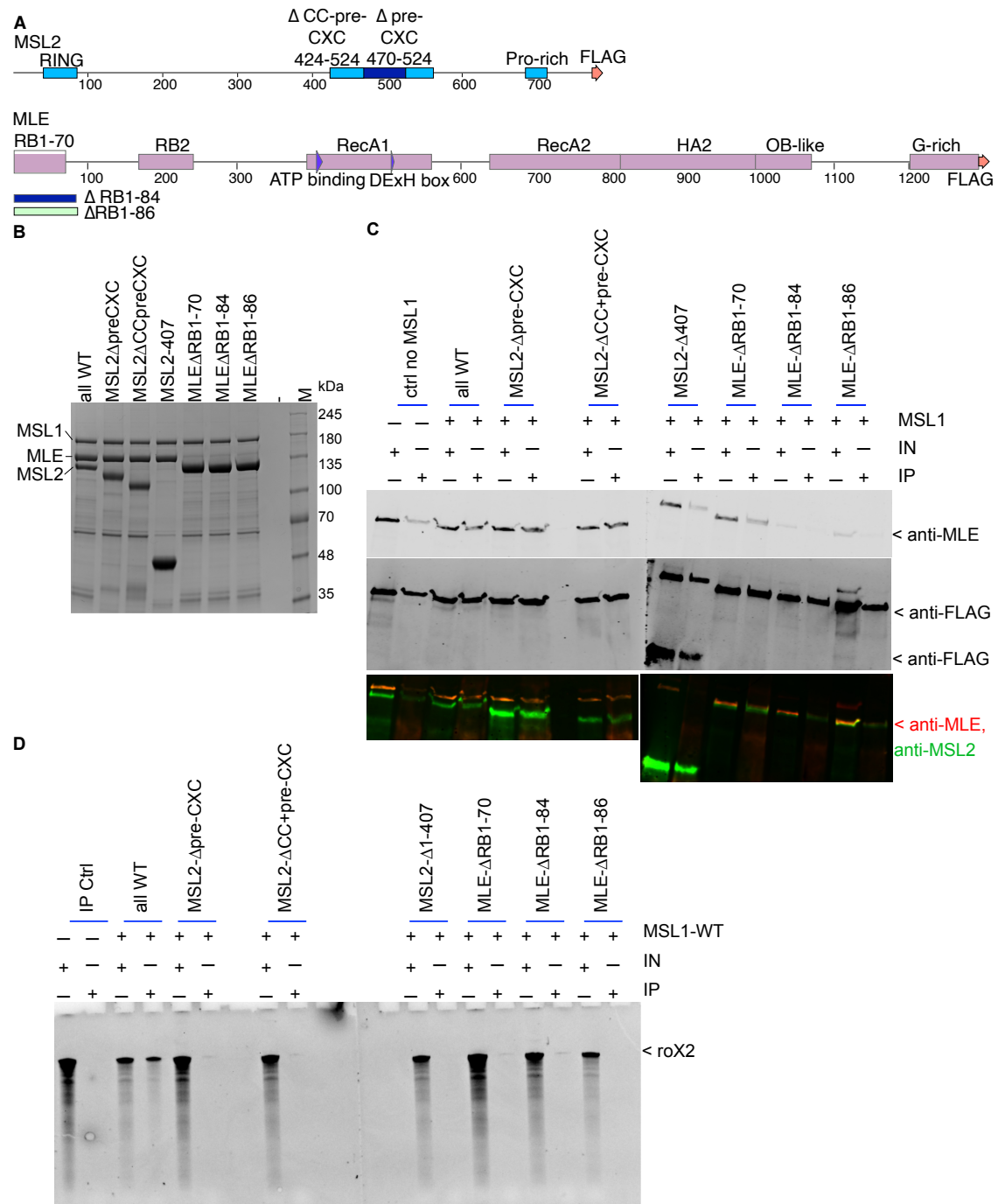
Figure 28: MSL2 and MLE deletion mutants based on the XL-MS data. **A** Constructs for MSL2 and MLE cloned and used for baculoviral expression. **B** Input samples before IP of the WT and mutant MSL proteins. Coomassie stained SDS-PAGE gels 8% acrylamide. **C** Western blots of input (IN) and anti-MSL1 immunoprecipitation (IP) reactions of the indicated WT and mutant proteins. If not indicated differently, the reaction conatin MLE-WT and MSL2-WT. MSL2Δ1-407 is the MSL2 C-terminal region only, provided by Silke Krause. In the control IP, MSL1 was omitted in the reaction. Three WB shown: anti-MLE probed, anti-FLAG probed and anti-MSL2 probed. **D** Ethidium bromide stained 5% acrylamide 8 M urea gel of input (IN) and anti-MSL1 immunoprecipitation (IP) reactions of the indicated WT and mutant proteins. If not indicated differently, the reaction conatin MLE-WT and MSL2-WT. MSL2Δ1-407 is the MSL2 C-terminal region only, which was kindly provided by Silke Krause. In the control IP, MSL1 was omitted in the reaction. Of all samples only the WT co-precipitated the roX2 RNA sufficiently.

To conclude, the IP assays of the deletion mutants hint towards a role of the deleted regions in protein-protein interaction as well as a potential role RNA binding. However, three technical challenges limit the conclusions drawn from these assays. First, the background of the IP is high, because MSL2 and MLE non-specifically stick to anti-MSL1 beads in absence of MSL1. Second, the antibody affinity decreases for the anti-MLE antibody in RB1 deletion mutants. Third, variable quality of the Western blots makes it difficult to prove the interaction based on this experiment only. Moreover, it remains unclear whether the loss in RNA binding is a direct effect of the deletion mutant or if it is inferred by the loss of protein-protein interactions and a decreased RNA 'incorporation'.

To further evaluate the interactions other experimental approaches should be tried. A repeated XL-MS experiment with the mutant proteins could be envisioned as well as cryo-EM. The RNA could be labelled by artificial nucleotides such as 4-thio-uridine (4-SU), a crosslinker for nucleic acid-protein interactions inducible by UV-light. The IP assay could be repeated with such a labeled roX2 RNA, UV-light crosslinked and the crosslinked protein domains could be analyzed by mass spectrometry [233].

6 Discussion

The structural composition of the *Drosophila melanogaster* MSL complex and its components is poorly understood due to challenges in reconstitution of the complex *in vitro*, a lack of suitable structural techniques, and the presence of intrinsically disordered regions in the MSL proteins. To address these issues, I utilized crosslinking/mass spectrometry (XL-MS) to gain insight into the complex's structure.

6.1 Advantages and limitations of XL-MS as a structural biology tool

Crosslinking involves introducing a covalent bond between two biomolecules or specific parts of the same molecule when they are in close proximity in the three-dimensional space (Figure 11). The crosslinker BS3 has reactive groups at each of its extremities connected by a flexible spacer, which allows to sample the space of up to 30 Å distance (Figure 11B). BS3 reacts with primary amines such as lysine residues in proteins, forming covalent bonds either within a single protein or between different proteins. These bonds are stable during trypsin digestion and can be analyzed by mass spectrometry.

XL-MS has three major advantages. First, the conformations are sampled in solution as opposed to crystallographic methods or on a cryo-EM grid. Therefore, the molecules can move freely and can obtain various as well short-lived, dynamic states, shedding light on the molecular dynamics and mechanisms of biochemical reactions. Second, the purity of protein preparations does not need to be extremely high, as contaminations can be addressed during mass spectrometric analysis (targeted detection of masses) and computational processing (database creation for the proteins of interest and crosslinks). This makes it possible to investigate crude extracts and cell lysates, potentially uncovering novel protein interactors. Third, the method is applicable to proteins that do not readily crystallize, which can be a challenge when studying proteins containing large intrinsically disordered domains or membrane domains, where crystallization and cryo-EM studies are often impractical.

Albeit giving valuable insights into the interactions within proteins or protein-protein interaction in protein complexes, XL-MS has several limitations as a structural biology technique.

The first limitation is a recognized problem in the field of XL-MS and revolves around the thresholding during the analysis of the crosslinks. To ensure the quality of identified crosslinks, stringent criteria and thresholds for their calling are applied and often manual evaluation of each individual crosslink is necessary. This manual evaluation is currently recommended as the "gold standard" by experts in the field to ensure the accuracy of crosslink data [164, 166, 168]. However, manual inspection can differ from person to person and might lead to non-reproducible results. Furthermore, additional potential sources of error in XL-MS emerge from false positives due to coincidental same mass, unprecise mass detection or even experimental variations during the MS run.

It is important to note, that XL-MS does not provide complete structural information on its own. Instead, it often complements other experimental techniques such as cryo-electron microscopy (cryo-EM), X-ray crystallography, or nuclear magnetic resonance (NMR) spectroscopy, which can offer more detailed

Discussion

structural insights. The integration of several or all structural techniques together with powerful artificial-intelligence (AI) driven predictions and models, are envisioned to revolutionize the field of structural biology [164].

Crosslinks inform about distance restraints for computational modelling, such as molecular dynamics (MD). In an MD approach the polypeptide chain of a protein of interest is placed in a virtual water or solvent box and using a force field let to relax into an energetic minimum *in silico*. The crosslinks limit the options of the overall folding in such a simulation. BS3 allows contacts with a maximum distance of 30 Å taking into account the length of the spacer and the length of the lysine residue side chains themselves (Figure 11B).

Moreover, crosslinking can be used for evaluation of structural models generated by AlphaFold (AF). Usually, modeling offers several models with sometimes similar energy minima. Crosslinking data can be used to check the different models, whether one might be superior that the others, e.g., by satisfying more detected crosslinks.

Similarly, experimental structures can be compared to the crosslinks. Since XL-MS is an ensemble technique and conducted in solution, it can access dynamic and transient states of the protein and protein complexes, which can be challenging to observe using static structural techniques, such as crystallography. If the distances do not match precisely, it does not necessarily follow that the crosslink is incorrect. An alternative explanation could be that during the crosslinking experiment, the protein(s) adopt transient conformations which might not be observable through static structural techniques like crystallography.

In conclusion, the results of XL-MS can be very insightful in integrative structural biology approaches.

6.2 XL-MS identified new interaction regions within the MSL complex

XL-MS was applied to study the interactions within the MSL complex and potential conformational changes when the complex composition was altered. To dissect the conformations obtained by the nucleoprotein complex, different subcomplexes, as well as individual proteins were analyzed, in absence and presence of roX2 RNA.

The majority of crosslinks were detected outside the classical, described interaction domains. Of the 2-MSL complex (MSL1ΔC-MSL2) the known interaction site is the MSL1 coiled-coil with the MSL2 RING domain. Many more crosslinks were found in the flexible regions of the MSL1-IDR and the MSL2 pre-CXC region. Both MSL1-IDR and MSL2 pre-CXC regions were extremely prone to form crosslinks in any constellation of subunits, even just MSL2 in absence of interaction partners formed crosslinks in that region (Figure 17, Figure 23 and Figure 25). Not much is known about the region N-terminal of the MSL2-CXC domain.

The MSL2-CXC domain is described to bind nucleic acids, especially the PionX sites of the X chromosome DNA [64, 65, 163]. Consequently, the MSL2-pre-CXC could be involved in that function, as well. More N-terminally there is a coiled-coil region predicted, which similarly showed an

Discussion

unexpectedly high number of crosslinks. This coiled coil region is both predicted by AlphaFold [228] and reported by the Uniprot database before AlphaFold was publicly available [234]. Both the N-terminal stretch and the C-terminal pre-CXC stretch of this coiled coil helix are predicted to have very long loops, hence it is very likely that the coiled coil and the pre-CXC domain can both make many dynamic contacts. These dynamic contacts may be transient, which would explain the variability in positions connected, i.e., the same lysine residue can form crosslinks with different other lysine residues in the ensemble. As there little is known about these regions of MSL2, it was of interest to follow these findings and to create deletion mutants (Figure 28). These mutant proteins, however, did not show any roX2 RNA binding. This could be due to two different reasons, on the one hand, the deletion abrogated the interaction between the proteins in a way that the "hand-over" and integration of roX2 by MLE to the MSL complex was not efficient. This hypothesis gains support from the fact that indeed the RB1 domain and the pre-CXC domain are necessary for the interaction of MLE and MSL2 and this interaction is essential for the roX2 integration. On the other hand, the folding of the proteins might be disturbed in ways that prevent general RNA binding. To exclude the second possibility further controls would be necessary to prove that the overall folding is intact and the general RNA binding ability is not perturbed. For instance, to assess the folding of the proteins, thermal unfolding by measurement of the melting curve could be performed. Additionally, structural methods such as cryo-EM, crystallography and XL-MS could be envisioned. The precise interaction sites could be narrowed down to specific peptides or even specific amino acids by gaining more structural information followed by mutational analysis.

The ensemble technique allows to sample different flexible conformations, which may be short-lived yet captured by crosslinking. Short-lived interaction states can also be investigated by biophysical methods, for example Förster resonance energy transfer (FRET). This fluorescence-based technique can elucidate the binding duration and dynamics.

Dynamic interactions are interesting to study as biological systems are dynamic and adaptable to environmental changes. Therefore, XL-MS is a valuable tool, because once the crosslink is formed it is covalent and non-reversible. Remarkably, the MSL1-IDR showed very many crosslinks. This may indicate a highly dynamic state, which may interfere with crystallization or cryo-EM analysis. Such IDR can be mutated, shortened or deleted. Based on the XL-MS data, a MSL1ΔIDR mutant (MSL1Δ358-799) was cloned, expressed and purified in the laboratory. The MSL1ΔIDR mutant could be co-purified as 3-MSL and 4-MSL complexes, proving that the MSL1-IDR is not essential for the protein-protein interactions. This finding raises the question whether the detected IDR crosslinks reflect functionally relevant interactions or merely non-specific encounters of the IDR due to structural flexibility. The MSL1ΔIDR mutant complexes were active in HAT assays and bound to mono-nucleosomes in EMSA. The MSL1ΔIDR complexes are invaluable for future structural studies, such as cryo-EM analysis, limiting the conformational space of the complexes by excluding this flexible region.

Discussion

Hence, the MSL1-IDR seems non-essential for protein-protein interactions, acetylation and nucleosome binding, it remains elusive, which function it may have in *D. melanogaster* as it is neither conserved in the mammalian homologue.

Newly suggested interactions, for example by the crosslinks between MSL2-MSL3, MSL2-MOF and the MSL3-MOF, need to be validated by structure-based mutagenesis, functional assays, and orthogonal interaction assays such as immunoprecipitation assays. I performed this analysis for the 2-MSL complex with MLE and roX2 by co-immunoprecipitation of the components when mutated in the respective domains of interest. Among the functional assays to be employed *in vitro* to test for the relevance of predicted interaction surfaces, roX2 RNA binding, histone acetylation (can be probed by HAT assays), co-immunoprecipitation of other MSL proteins, DNA binding and ubiquitylation come to mind. The *in vivo* functionality of DNA binding and recognition of the X chromosome as an essential function of the DCC for male viability and can be effectively studied in male *D. melanogaster* cells by immunofluorescence microscopy and chromatin immunoprecipitation (ChIP) and sequencing *[40, 68]*. Further, the integration of roX2 RNA could be analyzed by crosslinking the RNA with UV-light-activated crosslinker uridine, 4-thio-uridine (4sU). This artificial nucleotide can be incorporated into roX2 RNA both *in vivo* and *in vitro*, both incorporation in cells and *in vitro* transcription (ivt) were tested successfully. The crosslinked ribonucleoprotein complexes would be suitable for both the analysis of the RNA nucleotide contacts by sequencing, as well as the amino acids or peptides by MS in contact with the RNA [233].

6.3   Restraints for modelling

More than just the simple discovery of interaction sites, XL-MS data give great opportunities for molecular dynamic modelling and structure prediction. In that respect, especially the emergence of the AlphaFold prediction algorithm was a major milestone in the recent years [228].

There are two ways to use crosslinking data together with modelling. First, the crosslinks can be used as distance restraints: in a relaxation simulation the two connected lysine residues must have a distance of less than 30 Å in the case of BS3.

The AF-Multimer approach was applied to the 3-MSL complex (collaboration with Dr. Sebastian Eustermann, EMBL Heidelberg). A model of the 3-MSL complex was generated and the XL-MS data were compared to the model (Figure 21 and Figure 22). Mostly the IDR regions and loops of the proteins were poorly predicted, such that for further analysis of the crosslinking data they were excluded. Despite removing large parts of the proteins' sequences from the analysis and very strict selection thresholds for the crosslinks, there were some crosslinks satisfied. Notably, the most interesting model obtained so far, was a combination of the molecular dynamics modelling and the AF-Multimer prediction. Of the reproduced crosslinks, the best and strictest category of crosslinks, 70% were satisfied. In the lower categories the picture inverted, the non-reproduced crosslinks only 30% were satisfied. This gives an

insight that the truth actually might look quite differently. In solution, of course the proteins can obtain various conformations that are not accurately captured by one model structure alone.

Finally, both for the MD modeling approach, as well as for general validation of interaction sites, usage of different crosslinks with different linker length or even zero linker crosslinkers might be helpful to validate the contacts. Shortening the length would prevent to capture those crosslinks that are not direct contacts but just close in three-dimensional space. Crosslinks identified under different conditions, i.e., with different crosslinking agent are quite reproducible and reliable. The results from shorter length crosslinkers can be then integrated into the MD modelling approach, such that it limits the restraints. Certainly, the protocols for the crosslinking with another reagent need to be adapted, tested and the analysis parameters need to be changed.

6.4   Concluding remarks and outlook on XL-MS

XL-MS has revealed novel interaction surfaces among the MSL proteins and within individual protein domains. Incorporation of roX2 RNA could alter both the conformation of a single protein, MLE, and protein-protein interactions, in context of the 2-MSL, opening up intriguing avenues for research into the specific role of roX2 RNA. Notably, while the conformation of MLE undergoes significant changes upon roX2 RNA binding, the alterations in the 2-MSL complex are subtle and require further structural elucidation. Despite the structural changes were small, they can impact both the enzymatic and binding properties of the complex. Moreover, the upregulation of the transcription by the DCC is about twofold, which is a small change in biological terms. Oftentimes, change of transcriptional activity is defined by 10- or 100-fold changes. Given the current lack of knowledge about the complex's structure, it remains uncertain whether any potential structural change would be minor or major. Hence, even a slight alteration in the structure could be relevant. Computational modeling, including AF-Multimer, has been employed to track the crosslinks in the 3-MSL complex. Crosslink data have facilitated the design of MSL complexes, such as those with MSL1ΔIDR deletion, for cryo-electron microscopy studies. Moreover, exploring the *in vivo* effects of deleting specific interaction regions or analyzing enzymatic activity and binding affinity *in vitro* could provide valuable insights. Improved reconstitution of the complex with roX2 incorporation would enhance both structural studies and *in vitro* assays. Consequently, while roX2 may not dramatically alter the overall complex conformation, it could influence the enzymatic function of the MSL complex.

Part II – Functional roles of roX RNA within the DCC

The mechanistic role of roX RNA within the MSL complex still remains unresolved to date. The fact that roX RNA is crucial for male fly viability leads to the hypothesis that roX lncRNA can have a stimulating effect on the enzymatic activity of the MSL complex. MOF, the HAT within the MSL complex, could have an elevated activity for higher H4K16ac in presence of roX2.

To explore this hypothesis, I applied and adapted *in vitro* histone acetyltransferase assays to purified MSL complexes. Given that, *in vivo* MOF in the MSL complex marks H4K16ac on the X chromosome in male *D. melanogaster* flies, this assay offers a captivating approach to studying MSL complex activity. Prior research has shown that MOF alone cannot act on nucleosomal substrates; it requires MSL1 and MSL3 for full activity [184]. Additionally, roX RNA is essential for male fly viability, suggesting a functional role. I aimed to test, whether roX lncRNA has a stimulating effect on the enzymatic activity of MOF, leading to increased levels of H4K16ac.

Towards this end, I reconstituted both MSL complexes and nucleosome arrays *in vitro [11, 143, 211]*. The assay was adapted to allow testing and precise control of various conditions, including concentration, incubation time, complex composition, and RNA concentration. Instead of a non-specific readout using radiolabeled acetyl-CoA, acetylation was analyzed by WB and mass spectrometry [63, 184]. In this study, the method was refined to identify acetylation specifically using antibodies for certain histone acetylations. However, antibody-based assays raise concerns about antibody specificity [202, 203]. To address this issue, the Becker laboratory developed an orthogonal method employing mass spectrometry for histone modifications, offering both specificity and quantifiability [201].

7 Results

7.1 Characterization of the acetylation reaction by MOF based on Western blots

7.1.1 Chromatin

To characterize the acetylation on a near-physiological substrate, a nucleosome array was used. For this, a plasmid containing 25 repeats of the Widom-601 positioning sequence was amplified and purified [222]. This plasmid contains a 50 bp linker between each of the 147 bp long nucleosome positioning sequences. Thus, the nucleosome array will be evenly spaced and not packed too tightly, which could occlude binding sites for the MSL complex on the nucleosomes. Nevertheless, on the backbone of the plasmid, the 'pUC18' backbone, nucleosomes can assemble as well. There, the nucleosomes can be oddly spaced and potentially packed very tightly.

In vitro nucleosome array assembly was performed using the salt gradient dialysis (SGD) method [218, 235]. The theoretically perfect ratio is a mass ratio of DNA to octamers of 1:1.1. This should be titrated for each preparation of reconstituted octamers and DNA, respectively (Figure 29).
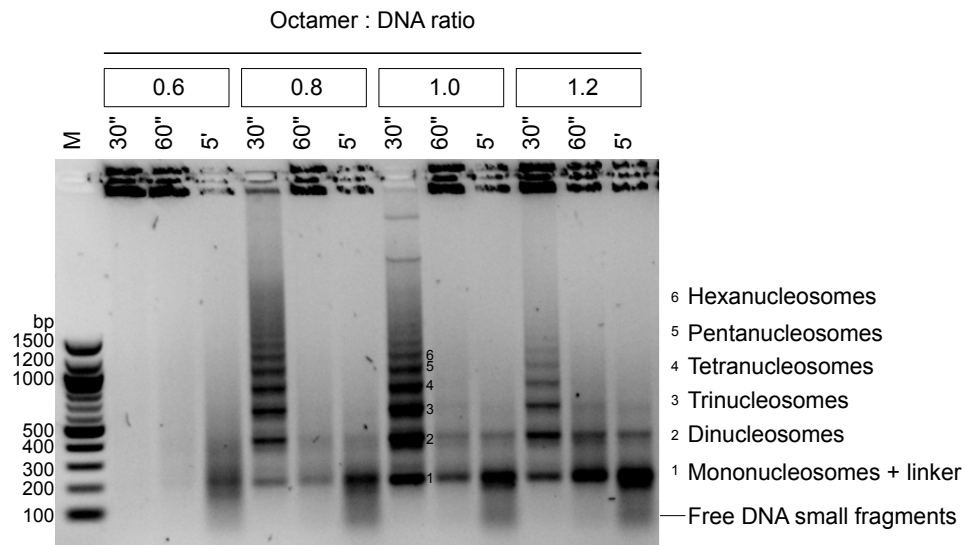
Figure 29: MNase digest of SGD nucleosome array assemblies. DNA was separated on a 1.5 % agarose gel stained with Midori green. The typical ladder-like appearance is an indication of appropriate nucleosome array assembly.

First, four different mass ratios of octamers and pUC18-25x601-plasmid DNA were tested (Figure 29). After the assembly by salt gradient dialysis, a digest by micrococcal nuclease (MNase) was performed for 30 s, 60 s and 5 min. The protein content of the samples was proteinase K-digested and the DNA was ethanol-precipitated. The MNase-digested DNA was loaded on an agarose gel, which was stained with Midori green to visualize the nucleic acids. A characteristic ladder pattern of the nucleosome array demonstrates a well-assembled array with nucleosomes on each Widom-601 positioning sequence. Both the DNA:octamer ratios of 0.8 and 1.0 ratios showed this ladder pattern (Figure 29). For practical reasons and to avoid free histone octamers in the chromatin assemblies a standard ratio of 1:0.9 was applied for all further experiments.

7.1.2   MOF acetylates chromatin in vitro in complex with MSL1 and MSL3

MOF was purified by FLAG affinity-chromatography as described previously for MSL2 [64], (Figure 30A). The 3-MSL and 4-MSL complexes were FLAG-affinity purified as described in part I of this dissertation for XL-MS experiments (Figure 14).

Results
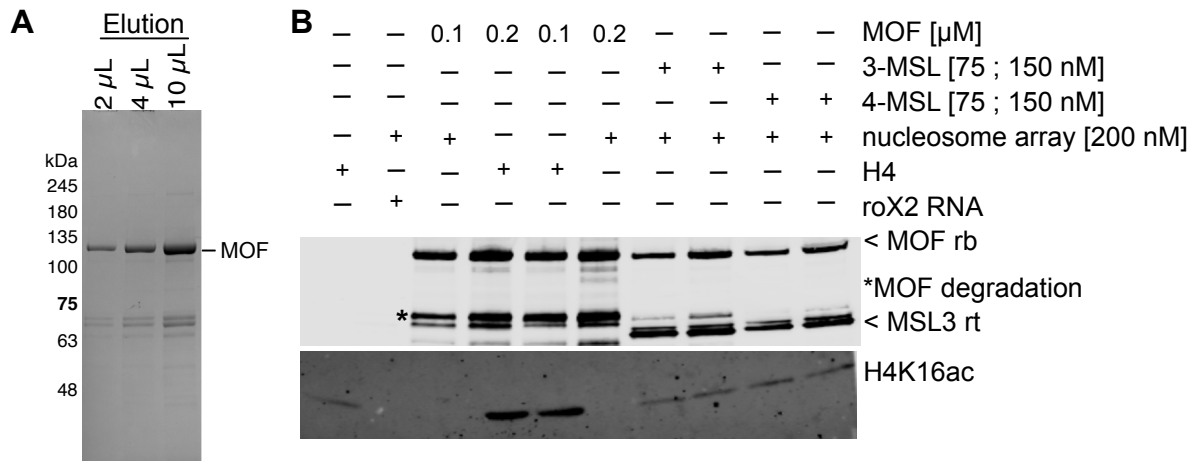
Figure 30: The 3-MSL and 4-MSL complexes acetylate H4K16, while isolated MOF does not. **A** Elution of the FLAG affinity material after MOF purification on a 7.5% acrylamide SDS-PAGE gel, stained with Coomassie blue G250. The expected MOF band is clearly visible between 100 and 135 kDa, degradation bands are visible <75 kDa. **B** Western blot of MOF in the HAT assay on H4 isolated histone and on nucleosome scanned with LiCOR instrument. Upper membrane shows MOF (anti-MOF, rb) and MSL3 bands (anti-MSL3, rt) for the MOF and MSL complex preparations. In the MSL complex MOF is more stable and does not show strong degradation bands (asterisk *). Lower membrane was probed with anti-H4K16ac antibody. MOF alone cannot acetylate H4K16ac. However, the isolated H4 histone is acetylated by MOF alone. In the 3-MSL and 4-MSL an acetylation band is detected as well for the nucleosome array substrate by the anti-H4K16ac antibody.

Histone acetylation assays were previously performed in the laboratory using radioactively labeled acetyl-CoA and classical detection on an X-ray film [63, 95, 184]. The assay was adapted to be executed without radioactivity but with a specific antibody against the H4K16ac mark and a fluorescently labelled secondary antibody that allowed detection by quantitative Western blotting (Figure 30B) [157].

First, acetylation by MOF alone was compared on isolated H4 versus on 25x601 nucleosome arrays. Two concentrations of MOF, 0.1 µM and 0.2 µM, were incubated with 10 µM acetyl-CoA in the HAT assay reaction buffer with 200 nM of the nucleosome array for 60 min. The reaction was stopped by heat denaturation of the enzyme. Subsequently, acetylation was analyzed by Western blot with anti-H4K16ac antibody. The upper part of the membrane was probed with anti-MOF antibody and anti-MSL3 antibody to control for loaded MSL amounts. MOF acetylated H4 (Figure 30B, lane 4 and 5), but failed to acetylate nucleosome arrays (Figure 30B, lane 3 and 6). H4 alone showed faint background staining by the anti-H4K16ac antibody, however the nucleosome arrays did not (Figure 30B, compare lane 1 and 2). When 3-MSL complex or the 4-MSL complex were used in the HAT assay, nucleosome arrays were acetylated, reproducing the previously known H4K16ac by MOF as part of the complexes [95, 184].

7.1.3 MOF within the 4-MSL complex acetylates both H4K16 and K12

To test whether MOF within the 4-MSL complex acetylates H4K16ac specifically, Western blot membranes were probed with an H4K12ac antibody. Further, to characterize the acetylation reaction, 4-MSL concentration was titrated at constant incubation time of 60 min (Figure 31). The 4-MSL acetylated H4K16ac with a Western blot detectable signal starting from 12.5 nM, reaching saturation at 150 nM (Figure 31A). The H4K12ac signal was weaker than the H4K16ac signal and started to be visible at the

lowest concentration of 25 nM of 4-MSL complex. The Western blot signal saturated at 100 nM 4-MSL concentration.
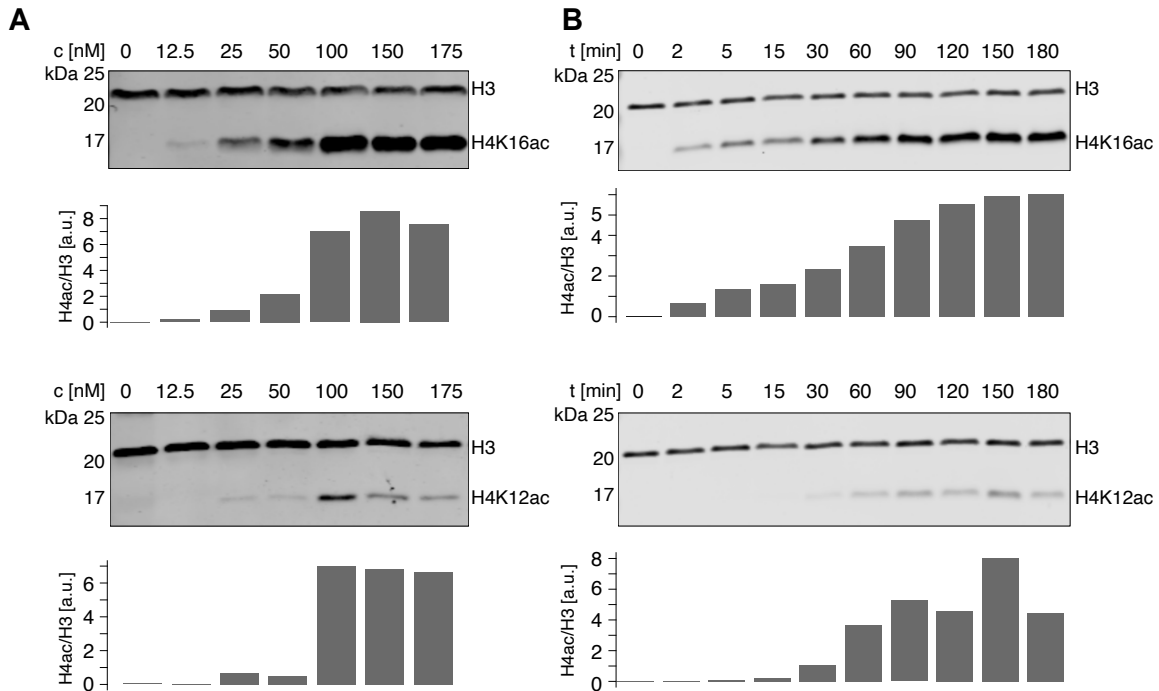
Figure 31: The 4-MSL complex acetylates both H4K16 and H4K12 on a nucleosome array substrate *in vitro*. **A** Representative HAT assay was conducted on a nucleosome array substrate using various concentrations (c) of purified 4-MSL complex for 60 minutes. Acetylation of histone H4 was assessed via Western blotting with antibodies specific to H4K16ac and H4K12ac. Histone H3 served as a loading control, and relative H4 acetylation was quantified and depicted in the bar plot as arbitrary units (a.u.). **B** A representative histone acetylation assay was performed using 50 nM 4-MSL complex over reaction times (t) ranging from 2 to 180 minutes. In the '0' lane, the 4-MSL complex was omitted. Detection and quantification of H4 acetylation were carried out as in (A). Figure reproduced as in [216].

Since these HAT assays were conducted at 60 min incubation time, a time course experiment was performed using 50 nM 4-MSL complex (Figure 31B). The H4K16ac signal is faintly detectable already after 2 min and saturates at 90 min. The H4K12ac signal is detectable after 30 min and saturates as well at 90 min. The signal of H4K12ac is around 10- to 20-fold less intensive than the H4K16ac signal if compared to the same H3-antibody staining. Of course, different antibodies cannot be compared in general. Albeit the 4-MSL complex acetylated H4K12, it might be to a lesser extent.

In conclusion, the 4-MSL complex acetylates both H4K16ac and H4K12ac based on Western blot analysis with the respective specific antibodies. The linear range was determined at 50 nM of 4-MSL and 5 to 60 min incubation time. The first hypothesis to explain this unusual finding, is to assume that the antibody of H4K12ac is not specific 'enough' to distinguish between H4K12ac and K16ac.

7.2 The impact of RNA on MOF within the MSL complex

Next, I wondered, whether roX RNA would enhance the acetylation. What led to this hypothesis? *In vivo* the MSL complex is present together with roX RNA on the active genes of the single male X

chromosome [161, 162]. These genes are marked extensively with H4K16ac by MOF [157]. Moreover, in part I of this dissertation, I showed that the addition of roX2 led to structural changes both within the helicase MLE as well as the two-subunit complex 2-MSL (MSL1ΔC-MSL2). These structural changes were subtle; however, they could directly or indirectly influence the enzymatic activity of MOF. To experimentally assess this hypothesis *in vitro*, roX2 RNA was introduced into the acetylation reaction, alongside MLE and ATP, and Western blot analysis was conducted to detect acetylation on H4K16 and H4K12.

7.2.1 RNA addition to the HAT assay reduced H4 acetylation

RNA introduction was tested in different setups, with different titrations and various controls. First and foremost, roX2 RNA was tested, as this is incorporated *in vitro* and *in vivo [47, 143]*. Apart from roX2 RNA, several shorter variants were tested, because they include critical stem loops for recognition by MLE and the 2-MSL complex. 'SL678' contains the stem-loops SL6, SL7 and SL8 with a total length of 236 bases, designed and provided by Dr. Pravin Jagtap [118]. In addition to that, roX2$_{123}$, a roX2 variant containing the stem-loop 7 (SL7) was tested [137]. The 4-MSL concentration was kept at 50 nM and the incubation time at 60 min as determined by pilot experiments (Figure 31). MLE was included at 50 nM concentration in a 1:1 ratio with the MSL complex and ATP was added at 1 mM concentration. RNA concentrations of roX2, SL678 and roX2$_{123}$ of 25, 50, 100 and 200 nM were tested, corresponding to the molar ratios of 1:2, 1:1, 2:1 and 4:1 compared to the 4-MSL complex. In neither of these conditions, the hypothesized boost of acetylation was observed (Figure 32A). On the contrary, the acetylation signal appeared diminished, especially H4K12ac was decreased compared to control reactions, where RNA was omitted (Figure 32B). The H4K16ac was also decreased, however, not to the same extent as the K12ac. This acetylation impairment was more prominently observed for longer RNAs, such as SL678 and roX2.

A

B
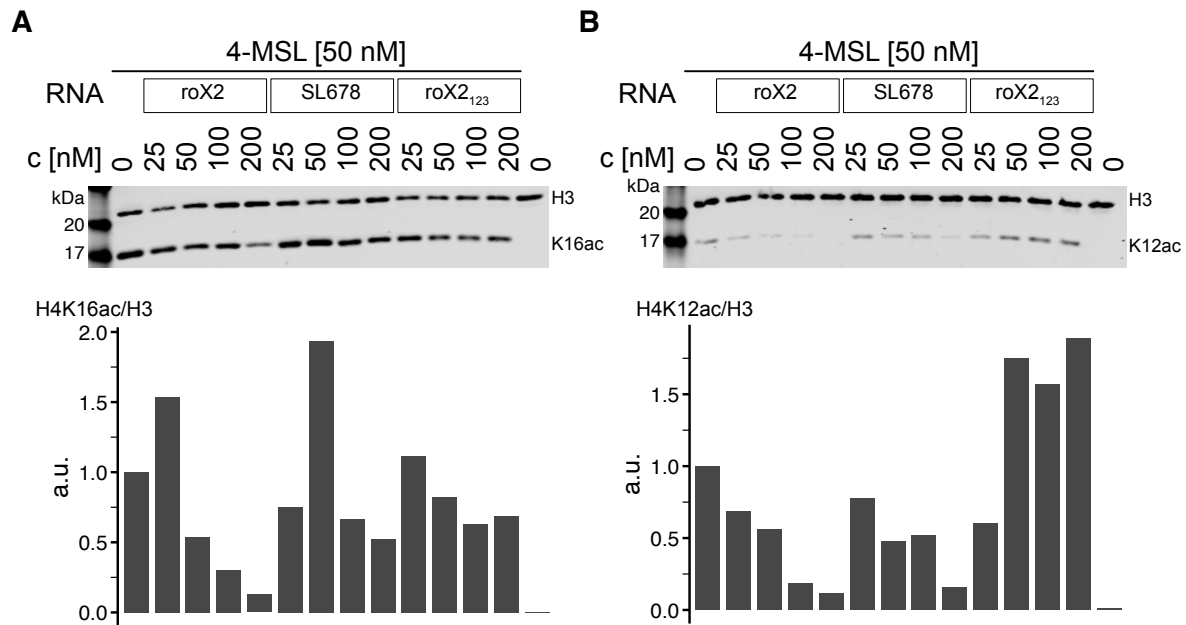
Figure 32: RNA does not increase H4K16ac, but rather impairs H4K12ac. **A** Titration with different roX2 constructs and different molar ratios (25; 50; 100; 200 nM) to the 4-MSL complex at 50 nM concentration at 60 min incubation time. Western blot of the HAT assay in absence or presence of RNA at indicated concentrations. H4K16ac was quantified in relation to the H3 signal. Higher excess of RNA lowered the H4K16ac signal. The H4K16ac signal was decreased upon roX2 RNA addition, but to a lesser extent as upon shorter RNA addition (SL678 or roX123). **B** Western blot of the same HAT assay samples probed in parallel on another membrane. The H4K12ac signal was strongly reduced by higher roX2 concentrations. Remarkably, at 200 nM H4K12ac was not detected anymore. roX2_{123} did not lead to suppression of H4K12ac. SL678 suppressed the acetylation but to a lesser extent than roX2 full-length (552 nt).

In summary, contrary to the initial hypothesis of an acetylation boost, inclusion of roX2 RNA led to a reduction of H4 acetylation of both H4K16ac and K12ac. H4K12ac was more prominently reduced than H4K16ac. Rox2 full-length reduced the acetylation signal more than the shorter variants, SL678 and roX2_{123}.

7.2.2   The suppression of H4K12ac is non-specific

The hypothesis that the effect on the reduction of the acetylation intensity is roX2-RNA-specific is very appealing, as the mechanistic role of roX RNA has been long sought after. In order to investigate this further, I constructed seven unrelated GFP RNA transcripts of various different length, in roughly 100 nucleotide steps. GFP has no sequence similarity with roX2 and is not assumed to fold into structurally related, uridine-rich stem-loops with roX boxes. The HAT assay was performed for 1 h on nucleosome arrays with 50 nM of 4-MSL, 50 nM of MLE and 1 mM of ATP. RNA was added in 2-fold molar excess to the protein complex.
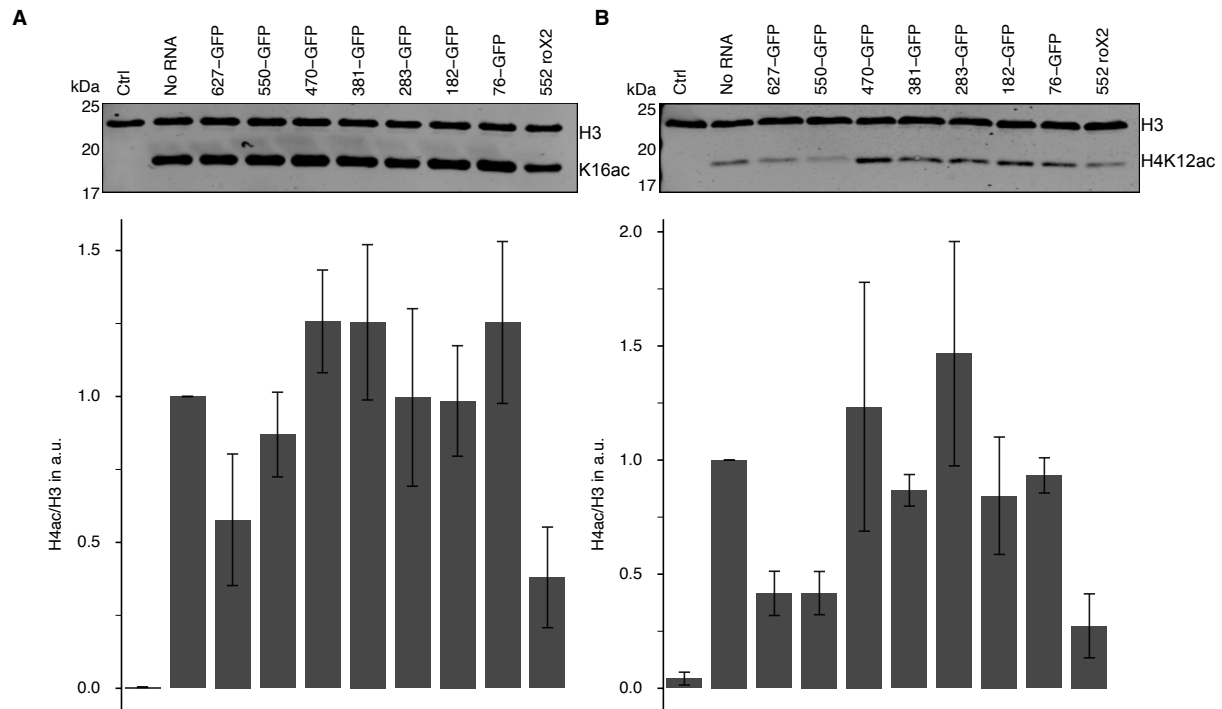
Figure 33: Long GFP transcripts suppress acetylation by the 4-MSL complex. **A** Representative Western blot with anti-H4K16ac staining and anti-H3 staining for loading control and quantification of three independent replicates below, error bar standard error of the mean. **B** Representative Western blot with anti-H4K12ac staining and anti-H3 staining for loading control and quantification of three independent replicates below, error bars represent the standard error of the mean (n=3, 4-MSL preparations).

Interestingly, only the longest GFP transcript (627 nt) reduced the H4K16ac signal to a similar extent as the roX2 full-length (here: 552 nt [137], Figure 33A). For the GFP RNA with matching length (550 nt) the H4K16ac signal was reduced in some replicates but variability was high and did not allow to claim a change in intensity (Figure 33A). Addition of GFP transcripts of intermediate or short length did not lead to a reduction of the H4K16ac signal (Figure 33A). Interestingly, the long GFP RNAs of 550 and 627 nt length revealed a reduction of H4K12ac levels similar to roX2, while transcripts shorter than 550 nt did not (Figure 33B).

In conclusion, the effect of roX2 RNA to reduce H4 acetylation on nucleosome arrays by the 4-MSL complex was not specific for roX2 and could be induced by unrelated RNA of similar length (550 nt) or longer.

7.2.3 The non-specific RNA effect depends on mass ratios of RNA

After the observation that short GFP RNAs did not reduce the acetylation signal as roX2 RNA or longer GFP RNAs do, I characterized this effect further by titrating different mass ratios of RNA. A molar ratio of 1.0 (4-MSL : roX2-full length) serves as a starting point, which corresponds to 280 ng of roX2 RNA. The corresponding masses of the GFP-470, roX2-123 and GFP-76 RNAs were included in the HAT assay with 50 nM 4-MSL, 50 nM MLE and 1 mM ATP. H4K16ac and H4K12ac was detected by Western blot and quantified relative to H3 (Figure 34).

Two independent experiments showed that the H4K16ac signal was mildly reduced by roX2 RNA and GFP-470 RNA in a concentration-dependent manner. The shorter RNAs did not show a clear reduction, except at the highest concentration of GFP-76 RNA (Figure 34A). Analogously the H4K12ac signal was detected and quantified. This revealed a good reduction for roX2, GFP-470 and at 1.12 µg of GFP-76 (Figure 34B). The Western blots suffer from high variability and a low number of replicates (n=2).

In summary, the reduction in acetylation effect was not linked specifically to roX2 RNA. Instead, RNA of a certain length and in high abundance yielded similar outcomes, suggesting a non-specific influence.
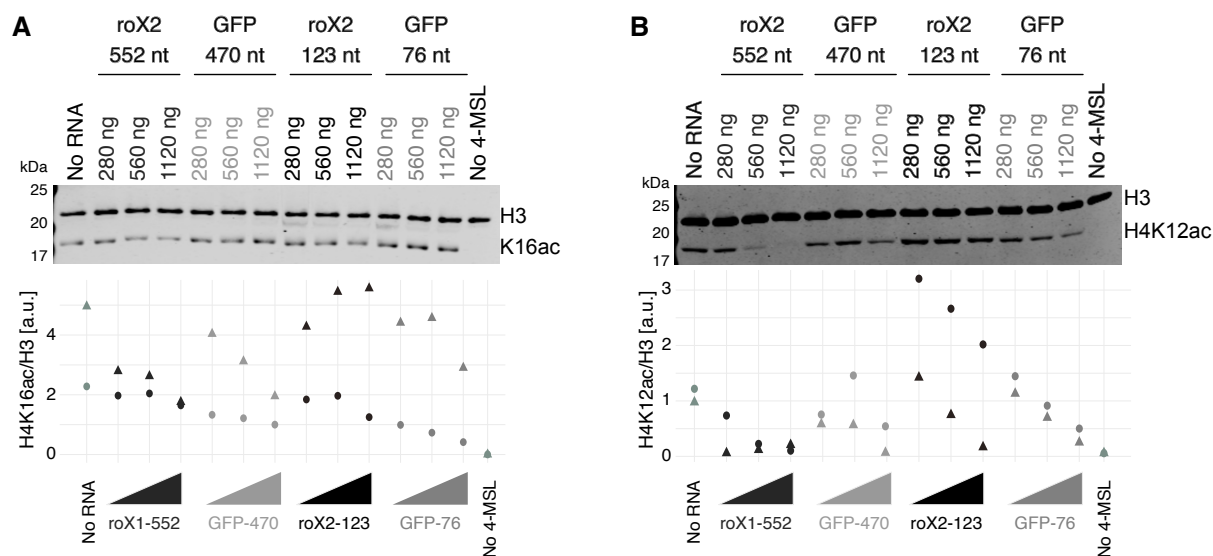
Figure 34: Long non-specific RNA reduces H4 acetylation. Mass ratio titrations of different *roX2* and *GFP RNAs* reveal that the acetylation reduction effect is not RNA specific. **A** Representative anti-H4K16ac Western blot with anti-H3 co-staining for loading control and quantification (n=2 independent 4-MSL preparations). 4-MSL concentration was 50 nM, incubation time was 60 min. Grey scales represent the type of RNA used. Oval and triangle represent the replicate. **B** Representative anti-H4K12ac Western blot with anti-H3 co-staining for loading control and quantification (n=2).

7.3   Mass spectrometry quantification of acetylation patterns

7.3.1   Motivation for mass spectrometric analysis of acetylation patterns

The Western blots offered an easy, fast and affordable method to analyze acetylation by the 4-MSL complex. Moreover, the specific antibodies allowed to identify the exact acetylation site. However, the Western blot readout had several limitations. First and foremost, the antibody specificity is questioned, especially upon observation of the H4K12ac staining, which is unknown for MOF [95, 184]. Moreover, the reproducibility of the Western blots was low, due to multiple technical reasons, which included gel casting, antibody stability, blotting variability and sample loading variability. This limited the reliability of the quantification, which depended on comparable intensity and staining quality between experiments.

The change in molecular weight by an acetyl group of 43 Da cannot be resolved by SDS-PAGE, even multiple acetylation events would not change the molecular weight enough to observe a distinct shift of a band on a gel. This means monoacetylation could not be differentiated from di-acetylation.

A technique, which can robustly distinguish even small changes in molecular weight, is mass spectrometry (Figure 13). Identification of post-translational modifications, such as acetylation, by mass spectrometry was previously applied in our laboratory [201]. The protocols were adapted to the *in vitro* approach of the HAT assays on nucleosome arrays. Due to the low complexity of the *in vitro* reaction, the time-consuming acidic extraction followed by gel purification of the histone proteins were omitted. Besides measuring all acetylation states and quantifying them relatively, MS is very sensitive and can detect small amounts [201]. Moreover, the MS approach allowed for parallel replication of the experiment with three independent protein preparations of the 4-MSL complex.

The analysis by mass spectrometry was conducted in collaboration with Anuroop Venkatasubramani and Dr. Ignasi Forné. I designed and conducted the HAT assays and prepared the peptide samples for MS analysis. Dr. Forné operated the mass spectrometer and Mr. Venkatasubramani performed the mass quantification for the peaks of the different acetylation patterns. These results are as well published in Kiss *et al.* [216].

7.4   Histone H4 tail acetylation patterns can be quantified by MS in a time-dependent manner

To determine whether the H4K12ac antibody reacted non-specifically with H4K16ac marks, MS was applied to unambiguously quantify these two modifications. Additionally, combinations of, for example, H4K12K16 di-acetylation can be detected and quantified. This approach yields a complete characterization of the acetylation reaction by the 4-MSL complex on nucleosome arrays *in vitro*.

The first step to characterize this reaction is to repeat the time course assays and quantify the different acetylation states at the time points in the linear range between 5 and 60 min. The HAT assay was performed with 50 nM of 4-MSL on 200 nM nucleosome array with 10 µM of acetyl-CoA, as established by Western blots for linear range reaction conditions (Figure 31). A reaction, in which the 4-MSL complex was omitted ('0 min') was included to control for background acetylation or chemical acetylation by acetyl-CoA in the reaction buffer. In the experiment three biological replicates, corresponding to independent 4-MSL preparations were processed in parallel. The mass spectrometric analysis revealed that the recombinant nucleosome array is devoid of any acetylation, as expected (Figure 35A, Appendix Figure 50A). After 5 min, the main acetylation is the H4K16 mono-acetylation at an average of 54%. In addition, 5% of H4K12K16 di-acetylated peptides were identified after 5 min. During extended incubation the H4K16ac levels did not increase further, rather the level remained constant and decreased at 60 min to 47%. Interestingly, the H4K12K16 di-acetylation increased over the time-course to 10% at 60 min. Additionally, the H4K8K12K16 and H4K5K12K16 tri-acetylations

started to accumulate as well as the tetra-acetylation, where all four lysine residues of the H4 tail peptide are acetylated. Curiously, the acetylation pattern of any combination that was lacking H4K16ac remained low (1% or <0.5%, noted as 0%), as well as mono-acetylation of H4K5 (<0.5%), K8 (1%) and K12 (1%), respectively.
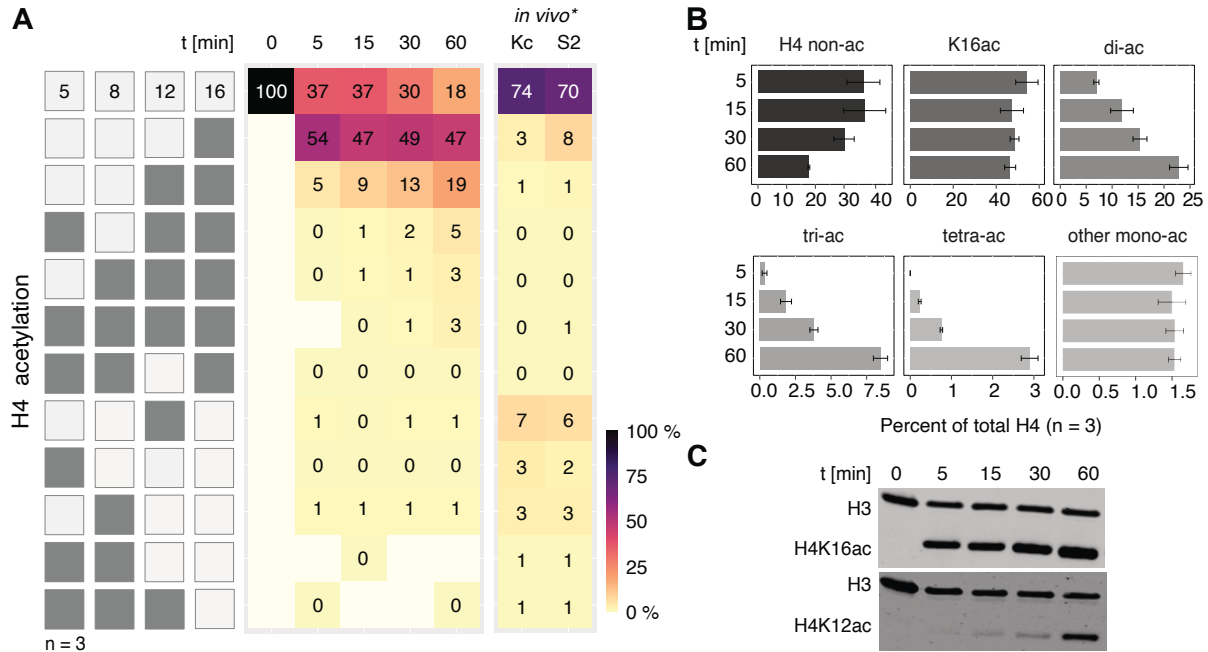
Figure 35: The 4-MSL complex sequentially acetylates H4 tail lysines *in vitro*. **A** A heatmap illustrates the acetylation patterns of H4 tail lysines on a nucleosome array substrate by 50 nM 4-MSL complex, quantified via mass spectrometry. The left panel depicts the various acetylation motifs on H4 lysines 5, 8, 12, and 16. Light shading indicates non-acetylated residues, while dark shading represents acetylation at the corresponding position. The heatmap displays all measurable combinations detected by MS2, albeit some diacetylated forms may not be included. The mean abundance of acetylated motifs at specified time points is presented. The top row displays the mean levels of non-acetylated H4 under different conditions. N=3 independent 4-MSL preparations. Motifs falling below the detection limit of 0.003% are indicated in white. '0%' refers to values < 0.5%, but above the detection limit. Additionally, the levels of H4 tail lysine acetylation in female Kc and male S2 cells ('*in vivo*') are provided for comparison [201]. **B** A bar plot summarizes the abundance of mono- and oligo-acetylated H4 tail motifs detected in (A). Grey shading indicates different scales. 'di-/tri-ac' represents the combined levels of all possible di- or tri-acetylated H4 tail motifs. 'other mono-ac' accumulates levels of K5ac, K8ac, and K12ac. The standard error of the mean of 3 independent 4-MSL preparations is provided. **C** Western blot analysis of one of the replicates with anti-H4K16ac antibody and H4K12ac antibody, respectively. Anti-H3 staining was performed as a loading control. Figure adapted from [216].

In comparison, in the Western blot, which was probed by anti-H4K16ac antibodies, there is a clear time-dependent increase of H4K16ac signal (Figure 35C). The increase is more pronounced, if the ratios between the H4K16ac and H3 signal are calculated (Figure 31A). The increase in H4K16ac signal on the Western blots over time can be explained by the accumulation of the 'oligo-acetylated' states, which carry the H4K16ac mark (Figure 35C). This means, when all patterns containing H4K16ac are summed up, irrespective of whether they are di-, tri or tetra-acetylated (oligo-acetylated), the signal increases. However, the antibody against H4K16ac cannot distinguish mono-acetylated H4K16 from the oligo-acetylated states. Along the same lines, the H4K12ac signal on the Western blots can be explained by the patterns containing H4K12ac, which especially start to occur at longer incubation times, whereas the H4K12 mono-acetylation remains low (Figure 35C).

These trends become more obvious if the sums of the acetylation patterns are taken into account (Figure 35B). The di-acetylation panel ('di-ac') sums up all possible di-acetylation patterns. In fact, H4K12K16 di-acetylation contributes most to the di-acetylation level. Similarly, the 'tri-ac' is the sum of all the possible tri-acetylation combinations. Over time a trend of linear increase for the combinatorial patterns can be observed, whereas the H4K16 mono-acetylation remains constant, as well as the 'other mono-ac' K5, K8 and K12, respectively.

There are stark differences to note when this *in vitro* analysis is compared to previous *in vivo* data (Figure 35A, [201]). First of all, the non-acetylated H4 tail is much more abundant both in female *Drosophila melanogaster* Kc cells (74%) and male S2 cells (70%), respectively. Secondly, the most abundant acetylation marks are mono-acetylation of H4K12 in female cells (7%) and mono-acetylation of H4K16 in male cells (8%), respectively. Finally, all di-, tri- and tetra-acetylation patterns are lowly abundant (around 1%, or lower). Of course, there are many differences between the *in vivo* and *in vitro* conditions. First, *in vivo* there are many HATs, not just MOF, such as Tip60 (H4K12ac and K5ac), chameau (H4K12ac), CBP (H4K8ac) and HAT1 (H4K5ac and K12ac) [14, 22, 201, 236, 237]. Second, *in vivo* HDACs constantly deacetylate and thus contribute to the non-acetylated H4 tail [16, 238-240]. Third, *in vivo* other modifications on the histone tail that counteract or prevent acetylation, for instance, histone methylation or acylation [16, 178, 193, 241]. And lastly, many other molecules exist in the cell nucleus, that are not reconstituted by this *in vitro* assay, which can regulate, occupy sites and limit the acetylation reaction [68, 242]. One molecule is of particular interest to this study, the lncRNA roX, as it is specific for male cells, which have a different acetylation pattern from female cells.

All in all, the analysis of the histone acetylation by mass spectrometry revealed that there is not only H4K16 mono-acetylation deposited on the nucleosome array by MOF in the 4-MSL complex, but also additional acetylation patterns occur over time. All of these patterns contain H4K16ac, followed by H4K12ac, hypothesizing that there could be a processive mechanism behind the acetylation by the 4-MSL complex.

7.5 RNA suppresses the 'oligo-acetylation' patterns of the H4 tail

The difference in the acetylation pattern between male and female cells could be mediated by the presence of roX RNA, which is a critical component of the dosage compensation complex (Figure 35A [201]).

In chapter 7.2.1 it was shown that, contrary to the initial hypothesis, RNA addition to the HAT assay reaction did not increase the acetylation by MOF within the 4-MSL complex. Rather, both the H4K16ac and the H4K12ac signal appeared to be decreased. To understand which acetylation patterns are impacted by RNA, I performed the mass spectrometric analysis of the patterns when RNA was included. Different RNAs were tested in different concentrations. First, roX2 RNA of 552 nt ('full length') was included in different molar ratios compared to the 4-MSL (25 nM, 50 nM and 100 nM) at 60 min incubation time (end point of linear range, near saturation). Additionally, MLE (50 nM) and ATP (1

mM) were included in the HAT assay reaction (50 nM of 4-MSL, 200 nM nucleosome substrate, 10 µM acetyl-CoA). In presence of 50 nM and 100 nM roX2 RNA the di-, tri- and tetra-acetylation were reduced compared to reactions, in which no RNA was added (Figure 36A). At the same time the H4K16 mono-acetylation is slightly increased (Figure 36A). Therefore, the roX2 RNA suppressed the oligo-acetylations in favor of the H4K16 mono-acetylation (Figure 36A, B).

The addition of MLE without RNA did not change the acetylation patterns compared to the 4-MSL complex without MLE (Figure 36A, B). Neither did the addition of other charged nucleic acids, such as polydeoxyadenylic acid (poly-dA, mixed length 250-500 bases) or yeast tRNA (70-90 nt), change the levels of oligo-acetylation compared to the RNA omitted control (Figure 36A, B).

Interestingly, when MLE was omitted and only 100 nM of roX2 RNA were included in the reaction with 50 nM of 4-MSL, the 'oligo-acetylation' was well suppressed (Figure 36A, B). This argues in favor of an unspecific RNA effect as roX2 RNA requires MLE for integration into the MSL complex [143].

Results

Figure 36: RNA inhibits the oligo-acetylation of H4 tails by the 4-MSL complex. **A** Heatmap illustrates the varying abundance of distinct H4 acetylation patterns in the presence or absence of nucleic acids, as assessed through mass spectrometry. Nucleosome arrays were subjected to a histone acetylation assay using 50 nM 4-MSL complex with or without 50 nM MLE, roX2 RNA, poly-deoxyadenylic acid (p-dA), and tRNA. roX2 RNA was added at 0.5-fold, 1-fold, or 2-fold molar ratios to the 4-MSL complex. p-dA and tRNA were included at a 2-fold mass ratio excess (280 ng). All reactions contained 1 mM ATP, and the reaction duration was 60 minutes. The left panel depicts the various combinatorial acetylation motifs on H4 lysines 5, 8, 12, and 16. Light shading represents non-acetylated residues, while dark shading indicates acetylation at the corresponding position. The map includes all detectable combinations measurable by MS2 but may lack some diacetylated forms. The first row illustrates the mean levels of non-acetylated H4 under different conditions. Data represent the mean of six replicates or, when denoted with an asterisk, of three replicates. **B** Bar plot summarizes the abundance of mono- and oligo-acetylated H4 tail motifs detected in the heatmap. 'di-/tri-ac' denotes the combined levels of all possible di- or tri-acetylated H4 tail motifs, while 'other mono-ac' aggregates levels of K5ac, K8ac, and K12ac. Standard error of the mean is provided. Figure from [216] with permission from NAR.

To investigate if the effect can be described more precisely during a time-dependent reaction, the HAT assay was repeated as a time series ranging from 5 to 60 min. roX2 RNA and GFP-RNA of 550 nt length

were included at a concentration of 100 nM. The GFP-550 RNA was chosen to avoid any length and molecular mass effects that could occur due to different length of roX2 and the control RNA. Compared to the control reactions lacking, the addition of roX2 RNA and GFP-550 RNA, respectively, demonstrated a suppression of oligo-acetylation states (Figure 37).
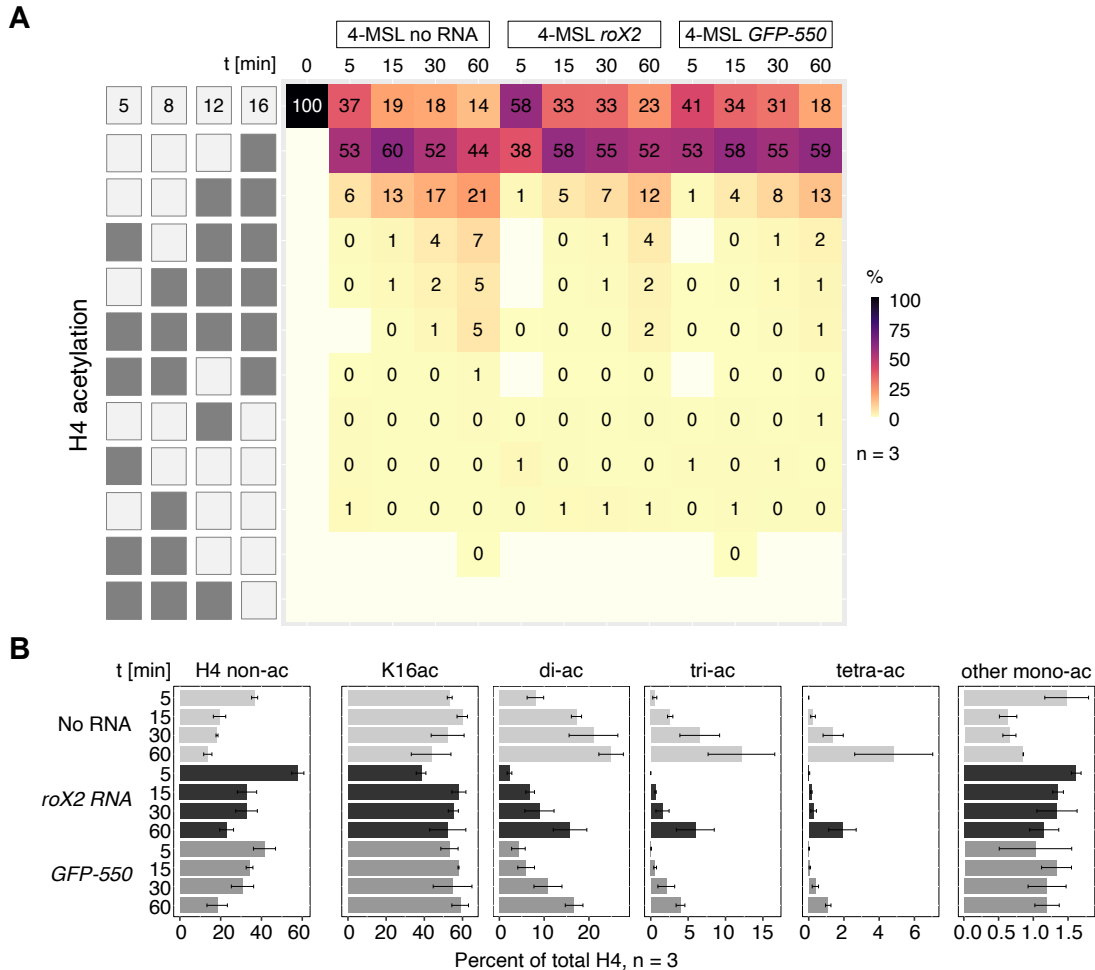
Figure 37: RNA impairs time-dependently oligo-acetylation of the H4 tail by the 4-MSL complex *in vitro*. **A** Heatmap shows H4 tail lysine acetylation by 50 nM 4-MSL complex with or without 50 nM MLE and a 2-fold molar excess of roX2 and GFP-550 RNA. All reactions contain 1 mM ATP. The left panel displays acetylation motifs on H4 lysines 5, 8, 12, and 16. The map includes detectable combinations measured by MS2. '0%' indicates values <0.5% but above the detection limit. **B** Bar plot summarizes mono- and oligo-acetylated H4 tail motifs. 'di-/tri-ac' represents combined di- or tri-acetylated motifs, and 'other mono-ac' denotes K5ac and K8ac. Standard error of the mean is shown for 3 independent 4-MSL preparations [216].

In conclusion, the mass spectrometry analysis clarified that the reduction of both H4K16ac and H4K12ac, which was observed in the Western blots, was caused by the reduction of oligo-acetylation patterns in presence of RNA. These patterns accumulate over time and can be effectively suppressed by long RNA. The effect was not roX2 RNA-specific, but could be provoked in absence of MLE and by unrelated RNA constructs (GFP-550).

7.6 The H4K16R mutation unveils MOF-mediated acetylation processes beyond K16ac

Is the H4K16ac a prerequisite for the acetylation of the other lysine residues in the H4 tail by the 4-MSL complex?

After the observation of the oligo-acetylation patterns, which always contained first H4K16ac and then H4K12ac, the hypothesis formed that H4K16ac is acetylated first and then the 4-MSL complex proceeds along the H4 tail outwards, like a zipper. To test whether the K16 is required for the oligo-acetylation reaction to take place, I created an H4K16R mutant, in which the arginine residue cannot be acetylated. Two extreme scenarios can be envisioned: 1) If the H4K16 binding and its acetylation was strictly required as a first step for the oligo-acetylation patterns to occur, one would not expect to find any acetylation of the mutant nucleosome arrays. 2) If MOF was not specific for H4K16ac at all, one would find just as much acetylation as of the wild-type nucleosome array, potentially starting from H4K12ac. Obviously, intermediate scenarios are also possible.

7.6.1 The H4K16R mutant histone assembles into nucleosome arrays

The H4K16R mutation was introduced into the H4 construct via Quik Change mutagenesis. The mutant histone was expressed in *E. coli* and purified like wild-type histones using a tag-less method involving anion/cation exchange columns. Octamers were formed and purified by size exclusion chromatography (Appendix Figure 49). The mutant octamers were assembled into nucleosome arrays on the 25x 601 plasmid (Figure 38).

Results
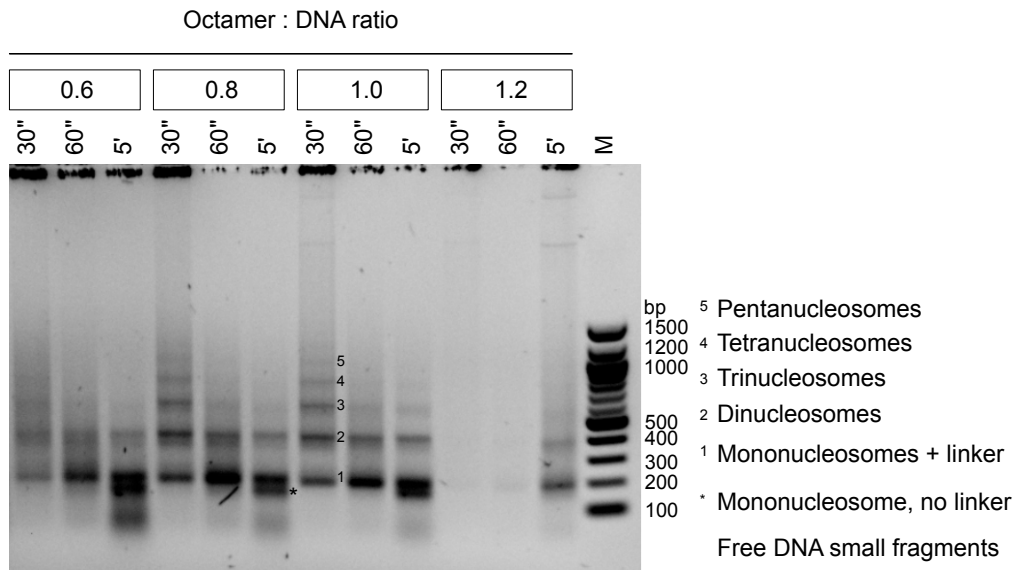
Figure 38: H4K16R octamers assemble into nucleosome arrays on the 25x 601 plasmid similar to wild-type octamers (Figure 29). The MNase partial digestion patterns demonstrate, that arrays containing several nucleosomes are formed. 1.5% agarose gel with Midori green staining.

7.6.2 H4K16R mutant nucleosome arrays can be acetylated by the 4-MSL complex at H4K12

The mass spectrometric analysis of 4-MSL on nucleosome arrays in the course of 5 to 60 min time points was repeated on the H4K16R mutant nucleosome arrays. The same conditions and concentrations were used as for the wild-type nucleosome arrays.

The H4K16R tail showed 4% H4K12ac after 5 min (Figure 39A, B, Appendix Figure 50B). H4K12 mono-acetylation reached up to 13% at 60 min. At longer incubation times low levels of the di-acetylations (H4K5K12ac and H4K8K12ac) occurred of up to 2% each and the tri-acetylation of H4K5K8K12ac reached around 1% on average at 60 min (Figure 39A). H4K12ac levels were low, yet they range in the same order of magnitude as the H4K12ac containing di-acetylations in the wild-type experiment (5% average at 5 min, 19% at 60 min (Figure 35)). As expected, the Western blot of the same samples showed H4K12ac, while the anti-H4K16ac antibody did not give any signal for the mutant nucleosome arrays (Figure 39C). Notably, this leads to the following two conclusions: 1) The H4K16ac antibody detects only H4K16ac, irrespective of mono-acetylation or oligo-acetylations. 2) The mutant chromatin with H4K16R does not contain traces or contaminants of the wild-type H4. The H4K12ac signal is comparable between H4K16R mutant and wild-type nucleosome arrays on Western blot level at the corresponding time-points (Figure 39C).

In conclusion, MOF in the 4-MSL complex is able to acetylate H4K12ac in the H4K16R mutant nucleosome arrays. However, this acetylation is diminished compared to the H4K16ac mono-acetylation in the wild-type nucleosome arrays. Di- and tri-acetylations arise over time and reach similar levels as the respective tri- and tetra-acetylation in the wild-type nucleosome array. Thus, H4K16R is a possible substrate for MOF and the mono-acetylation of H4K16ac is not strictly required for the proceeding along the H4 tail.

A

B

H4K16R nucleosome arrays

C
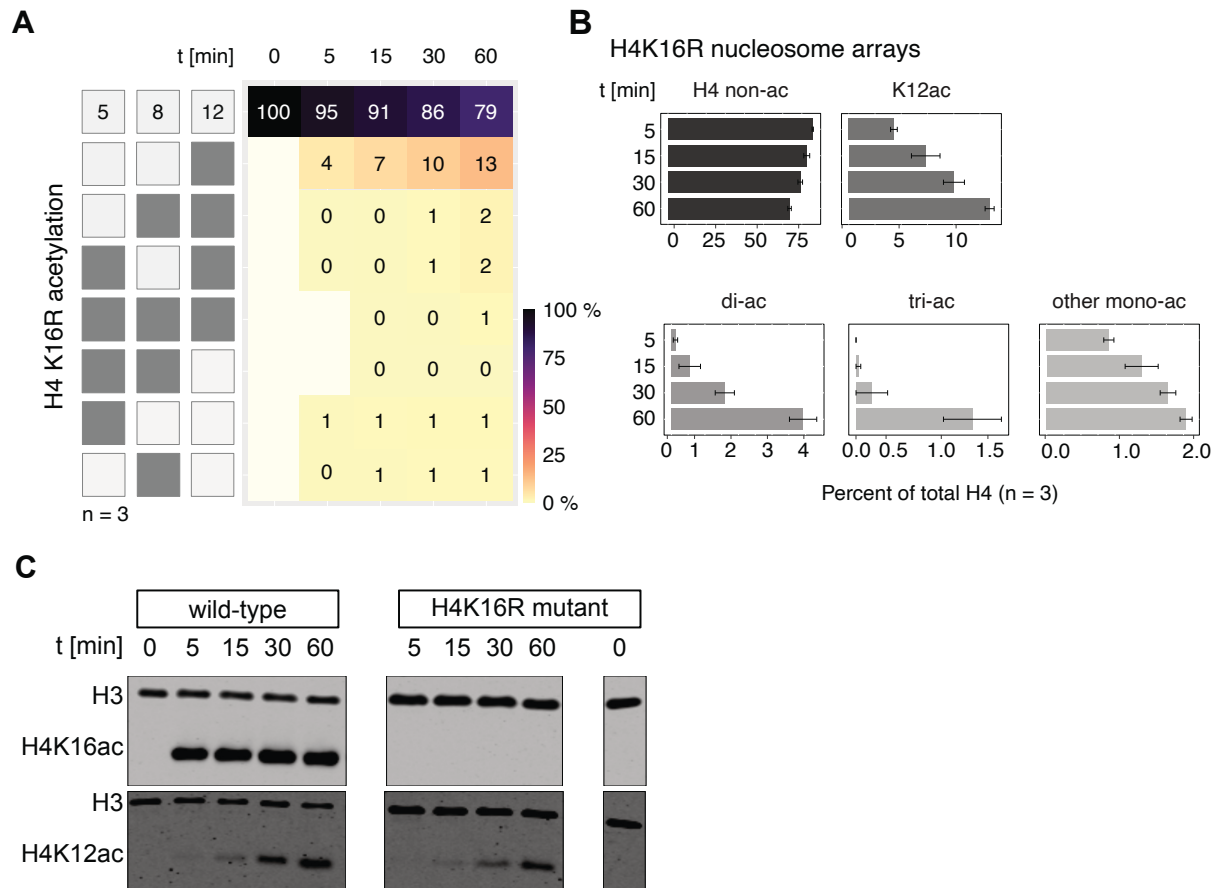
Figure 39: H4K16R mutant nucleosome arrays can be acetylated by the 4-MSL complex over time. **A** Heatmap of the MS quantification of the means of three independent 4-MSL preparations reveals that H4K12ac rises over time and few, but measurable amounts of H4K5K12 and H4K8K12 di-acetylations as well as H4K5K8K12 tri-acetylations appear at 60 min. **B** Bar graph representation of the data as in A with error bars standard error of the mean of three independent 4-MSL preparations. **C** Histone acetylation assay with 50 nM 4-MSL complex on wild-type or H4K16R mutant nucleosome arrays, respectively. H4 acetylation at indicated reaction times was analyzed by Western blot using antibodies specific for H4K16ac and H4K12ac. One representative Western blot is shown. Figure adapted from [216].

7.7 Mathematical modeling supports the hypothesis that MOF is a processive enzyme

The mass spectrometric data suggested that MOF first acetylates H4K16 and then starts to acetylate K12, K8 and K5 if time allows. To follow up on this hypothesis, I collaborated with the mathematicians Dr. Dilan Pathirana and Prof. Dr. Jan Hasenauer (University of Bonn). To delve into the kinetics of MOF activity, they employed mathematical modeling, testing three hypotheses: (1) Non-processive H4 acetylation with enzyme in excess ('mass action', Figure 40A) [243]. (2) Non-processive H4 acetylation with limiting enzyme ('Michaelis-Menten'). (3) Processive H4 acetylation, with the enzyme persistently bound to the substrate to acetylate additional lysine residues ('processive'). These models were calibrated against time-dependent H4 tail acetylation data and compared using the Akaike Information Criterion (ΔAIC). The processive model was strongly supported by the data, with the model suggesting a slower transition of the enzyme between nucleosome arrays being the most favored, indicating a spatial effect

in enzyme behavior. This spatial effect could be explained by a prolonged residence time of MOF on the nucleosome even after the acetylation of H4K16 is completed. The reaction mechanism would follow and inside-out 'zipper' model. The observation that all detected combinations of H4 tail acetylation feature K16ac indicates that the enzyme predominantly targets the K16 site.

To evaluate the enzyme's motif specificity, they conducted Bayesian uncertainty quantification [216]. Analysis of the relative acetylation rates for different motifs revealed that the acetylation sequence of H4 → H4K16 → H4K12K16 dominated much of the flux (Figure 40B). However, in rare cases where the first acetylated lysine is K5, K8, or K12, the second acetylated lysine is often not K16 (e.g., H4K12 → H4K5K12 and H4K12 → H4K8K12). Overall, the model supported the experimental data, indicating selective K16 acetylation by the 4-MSL complex over other acetylation sites. In the *in vitro* conditions of this project, conclusively demonstrated that this processivity is a characteristic of MOF and remains to be evaluated for other HAT enzymes.

The model was tested for the H4K16R data, where all reactions including K16 acetylations were excluded. The model agreed with the data, where then the first residue acetylated was K12 with an outwards progression [216].

In conclusion, the mathematical modeling supported the hypothesis that MOF is a processive enzyme, staying bound after acetylation of H4K16, with an outward progression reminiscent of a 'zipper' mechanism. It would be interesting to further apply the model to the data of the reactions including RNA, which was not performed due to time restraints.
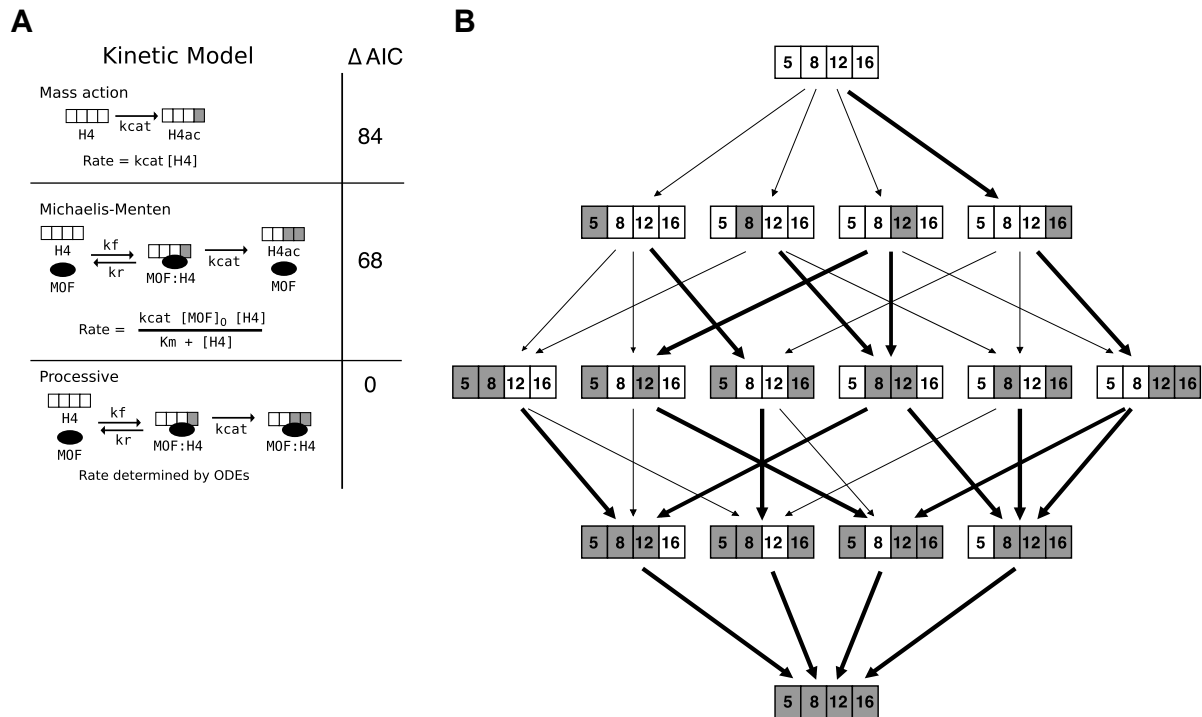
Figure 40: MOF is a processive enzyme. **A** The tested scenarios for MOF kinetics involved three models: the mass action model, the Michaelis-Menten model, and the processive model. The mass action model assumes a non-processive reaction where the rate of acetylation depends on the current concentration of the substrate, with the enzyme in excess. On the other hand, the Michaelis-Menten model also assumes a non-processive reaction but considers the enzyme-substrate complex concentration at equilibrium throughout the experiment. The processive model, however, explicitly models enzyme, substrate, and product interactions through a system of ordinary differential equations (ODEs), without assuming enzyme-substrate complex equilibrium. Additionally, this model does not presume enzyme excess. The rate in the processive model is implicitly determined by the ODEs system. These models were calibrated against time-dependent H4 tail acetylation data, with the difference in the Akaike information criterion (ΔAIC) relative to the AIC of the best model indicating model fit to the data, where lower AIC values suggest better support by the data. **B** Regarding site-specific acetylation of each motif on wild-type nucleosome arrays, relative probabilities were determined. The ensemble used contained samples of acetylation catalysis constants "kcat" (referenced from the processive model). In this representation, black arrows depict the relative probability for a particular motif to be acetylated at a specific site. The thickness of each black arrow represents the ensemble's median value for the corresponding reaction rate constant [216].

7.8 Long, non-specific RNA impairs nucleosome binding of the MSL complex

The acetylation data along with the modeling suggest that the 4-MSL complex stays bound for prolonged times on the nucleosome substrate and thus allows for additional acetylation along the H4 tail. The experiments including RNA in the HAT assay show a reduction of the acetylations apart from H4K16 mono-acetylation. The reduction of oligo-acetylation by RNA suggest that RNA might impair nucleosome binding by the MSL complex. To explore this possibility Dr. Marisa Müller performed electrophoretic mobility shift assays (EMSA) on mononucleosome substrates. These mononucleosomes were assembled on a DNA containing a 601-Widom positioning sequence with an 80 bp overhang (gift from Dr. Sebastian Eustermann, EMBL). The 4-MSL complex showed a clear band shift of the mononucleosome in absence of RNA (Figure 41).

Results

When roX2 RNA was added in stoichiometric amounts (50 nM), the band shift was abrogated, independent of the presence of MLE (Figure 41A). Upon addition of the control RNA GFP-550 the band shift was also no longer observed (Figure 41A). Including total RNA of Kc cells, which contain a mix of all RNAs extracted from the female cell line but no roX RNA, also suppressed the binding, demonstrating that the effect is not specific for roX (Figure 41A). However, yeast tRNA (70-90 nt) or single-stranded DNA (poly-dA, 250-500 bases) did not inhibit the binding to the nucleosome (Figure 41A).
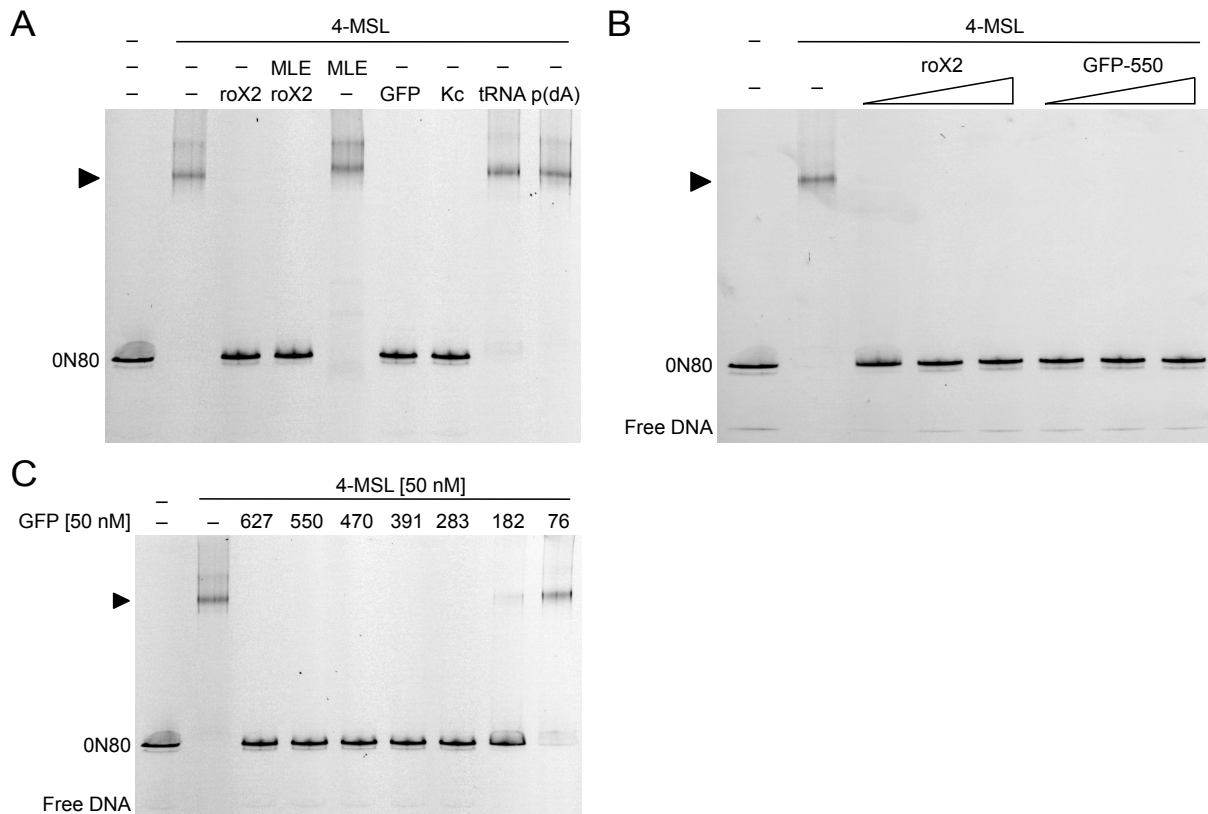
Figure 41: Interaction of the 4-MSL complex and mononucleosomes is destabilized by long RNA. **A** RNA disrupts the 4-MSL-mononucleosome interaction non-specifically, as shown by electrophoretic mobility shift assay (EMSA). EMSA was conducted by Dr. Marisa Müller using 50 nM 4-MSL and 10 nM 0N80 mononucleosomes in the absence or presence of MLE (50 nM), roX2 RNA (50 nM), GFP-550 RNA (50 nM), total RNA from *Drosophila* Kc cells (200 ng), yeast tRNA (200 ng), and poly(dA) polymer (200 ng), respectively. **B** RNA roX2 and GFP-550 concentration titration with 50 nM 4-MSL and 10 nM 0N80 mononucleosomes was assessed via electrophoretic mobility shift assay (EMSA). The presence of full-length roX2 (552 bases) and GFP (550 bases) RNA, each at concentrations of 25-50-100 nM, destabilizes these complexes. **C** Disruption of 4-MSL-mononucleosome complexes depends on RNA length. EMSA was carried out using 50 nM 4-MSL and 10 nM 0N80 mononucleosomes in the absence or presence of GFP RNA variants of specified lengths (in bases) at a concentration of 50 nM each. EMSAs performed by Dr. Marisa Müller. Figure adapted from [216].

When different concentrations of roX2 or of the GFP-RNA were tested, the 4-MSL showed no binding to the mononucleosomes even in sub-stoichiometric amounts of the RNA (Figure 41B). Subsequently, different lengths of the GFP-RNA were tested. The 4-MSL showed no binding to the mononucleosomes when the GFP transcript was longer than 283 nt (Figure 41C). For the 182 nt long transcript a faint band was visible for the bound nucleosome (Figure 41C). The shortest GFP transcript tested of 76 nt did not impair the band shift (Figure 41C). This agrees with the HAT assays: both long transcripts (GFP-550 or

roX2 RNA) were shown by mass spectrometry to inhibit the oligo-acetylation (Figure 37), while tRNA and poly-dA do not (Figure 36). The length titration of the GFP constructs showed that the acetylation, especially H4K12ac, which can be seen as a proxy for oligo-acetylation, was reduced by long transcripts, but not by the 76 nt GFP (Figure 33, Figure 34).

In conclusion, the EMSAs show that binding to nucleosomes by the 4-MSL is reduced in presence of long non-specific RNA. Therefore, if the complex cannot bind stably or for expanded time periods, the additional oligo-acetylation further than H4K16ac will be impaired.

7.9 The *Drosophila melanogaster* TIP60[piccolo] complex preferentially acetylates H4K12 and K5

Subsequently, I wanted to test whether the zip-like oligo-acetylation mechanism and the inhibition by long RNA is a common characteristic of MYST-type HATs. Thus, I focused on the Tip60 acetyltransferase, which is a member of the MYST enzyme family with a conserved HAT domain [14, 104, 197]. I aimed to compare the HAT activities of Tip60 and MOF in their respective complexes on nucleosome arrays. How does Tip60 acetylate the H4 tail as a function of time? Does it change, limit or boost its activity and specificity in presence or absence of non-specific RNA? Does it have a processive zipper-like mechanism of oligo-acetylation patterns as the MSL complex?

Tip60 also exists in a multi-subunit complex, in *Drosophila melanogaster* called the DOMINO complex [194]. As this complex is large and therefore difficult to handle *in vitro*, I focused on a smaller subcomplex, the trimeric dTIP60[piccolo] complex, which was described to have acetylation activity on nucleosomes [191, 196, 206]. Dr. Zivkos Apostolou and Mrs. Silke Krause cloned and expressed the trimeric dTIP60[piccolo] complex.

Drosophila Tip60 is an acetyltransferase specific for H4K5, H4K12 and the *Drosophila* H2A variant H2A.V [182, 194, 199]. The Tip60 homolog in *Saccharomyces cerevisiae*, Esa1, resides in a large multi-subunit complex, called the NuA4 complex, of which additional subunits are still discovered to date [197, 244, 245]. A smaller subcomplex, called piccolo NuA4 complex, consists of Tip60, Ing3 and the enhancer of polycomb (EPc) and is sufficient to acetylate nucleosome substrates *in vitro [197, 246]*. These proteins are conserved across species and are present in the *D. melanogaster* DOMINO complex, as well as in the yeast NuA4 complex or the mammalian ortholog [105, 194]. We named the trimeric complex of Tip60, Ing3 and EPc, dTIP60[piccolo].

7.9.1 The dTIP60[piccolo] complex co-purifies substantial amounts of RNA

The dTIP60[piccolo] complex was purified by affinity chromatography using the Strep tag on Tip60. Two prominent peaks were observed (Figure 42A). The first peak of the elution included a major contaminant of an acetyl-CoA carboxylase, as determined by MS (by Dr. Zivkos Apostolou) (Figure 42B). The

Results

second peak includes the three proteins of interest dTip60, EPc and ING3, confirmed by MS (Figure 42B, C). Additionally, a heat shock chaperone Hsp70 and actin 5c are co-eluted (Figure 42C).
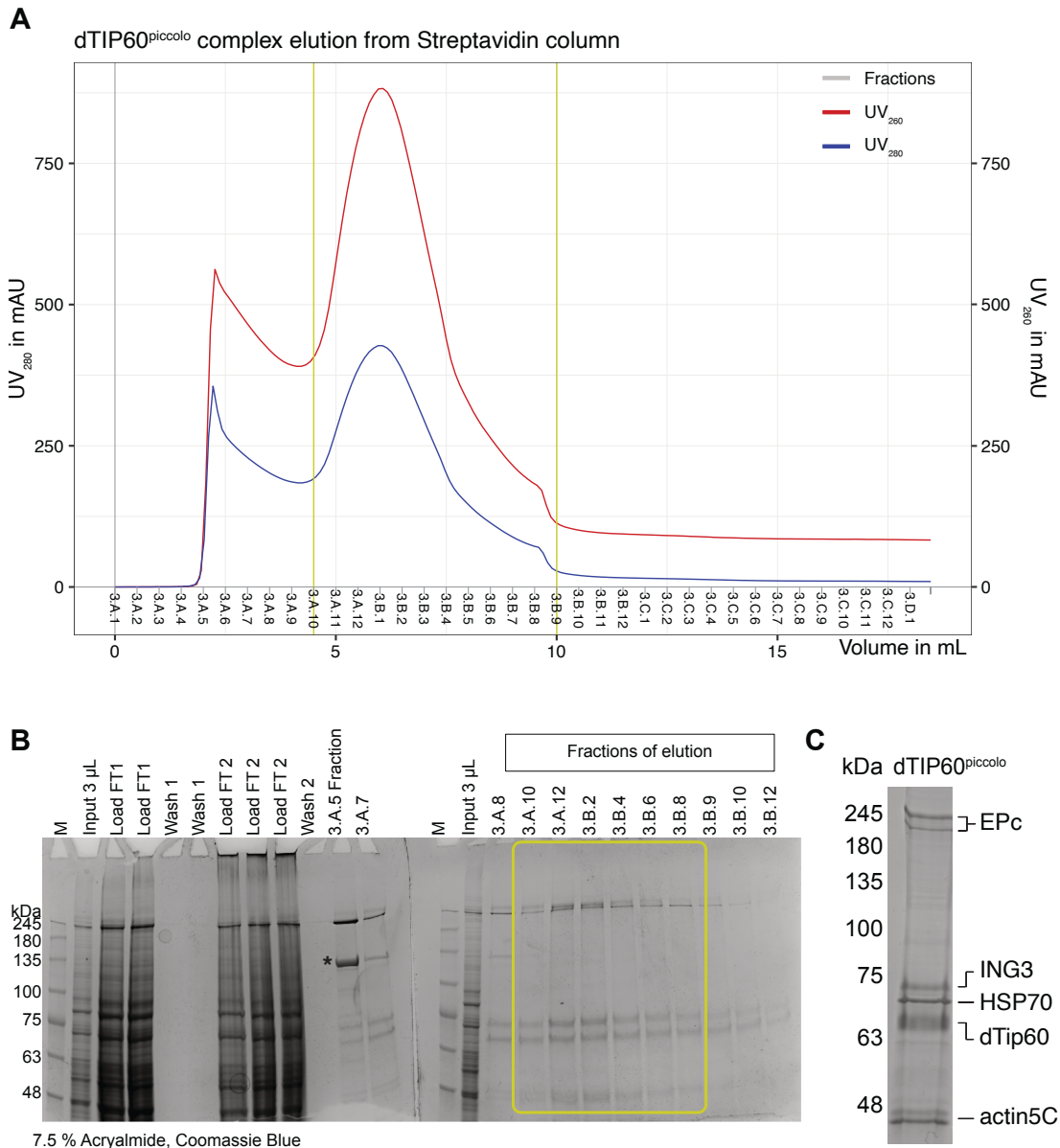
Figure 42: The dTIP60^{piccolo} complex can be separated from a carboxylase contaminant by Streptavidin chromatography. **A** Chromatogram of the streptavidin column at the Äkta, UV 260 nm shown in red and UV280 nm shown in blue. UV 260 nm has a higher absorption than the UV 280 nm, which indicates that there are substantial amounts of nucleic acids co-purified. Yellow vertical lines indicate pooled elution fractions, which were used for further experiments. **B** Coomassie blue stained SDS-PAGE gel of the purification steps of the dTIP60^{piccolo} complex and fractions of A. The asterisk marks the contamination, fractions from the first peak of A. The yellow box marks the pooled elution fractions (yellow vertical lines in A). **C** Pooled elution fractions of B with protein names identified by proteomics MS. Figure adapted from [216].

Actin is a known component of the NuA4 complex of yeast and its enrichment may be functionally important [196]. The *Drosophila melanogaster* and the *Spodopthera frugiperda* actin proteins are highly conserved, such that the protein is co-purified from the host expression system, albeit not cloned into the baculoviral expression vector.

The chromatogram revealed a high absorption at 260 nm wavelength indicative of nucleic acids (Figure 42A). This suggests the presence of RNA despite the fact that benzonase and RNase A are included in

the lysate buffer during the purification. Thus, the purified dTIP60^{piccolo} was further analyzed by RNase A and proteinase K digestion of the preparations (Figure 43A). As a control, 500 ng of roX2 RNA were treated the same way as the protein preparations tested. The smear-like staining was first observed for the dTIP60^{piccolo} samples and completely digested by incubation with RNase A. When the RNase A digestion was repeated in the presence of excess GFP RNA, the same result was observed (Figure 43B). In conclusion, the dTIP60^{piccolo} complex was purified using the streptavidin-biotin chromatography to a satisfying purity. The presence of all desired components was confirmed by MS and additionally actin-5C was discovered to be present in the preparations. Moreover, the preparations contained non-specific RNA, which suggests a high affinity of the dTIP60^{piccolo} complex for RNA.
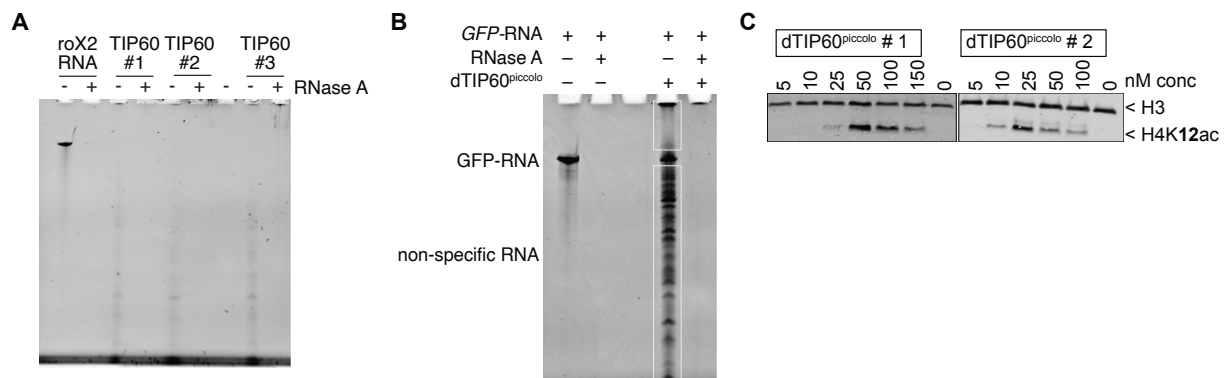
Figure 43: The histone acetyltransferase complex dTIP60^{piccolo} co-purified substantial amounts of non-specific RNA. **A** Midori green stained 8 M urea, 5% acrylamide gel of RNase A and proteinase K digestion of the three different dTIP60^{piccolo} preparations showing RNA contaminations. **B** Midori green stained 8 M urea, 5% acrylamide gel of additional GFP-RNA added to test efficient RNase A digestion of the non-specific and the GFP-RNA at the same time. **C** Western blot after 60 min HAT assay with dTIP60^{piccolo} complexes stained with anti-H4K12ac and anti-H3 antibody, respectively. The concentration of the dTIP60^{piccolo} is indicated, two representative enzyme preparations are shown. H4K12ac by Tip60 starts to be detectable between 10 and 25 nM. For all further HAT assays 50 nM of Tip60 were used. Figure adapted from [216].

7.9.2   The dTIP60^{piccolo} complex is active in vitro and acetylates H4K12 and K5

Next, I tested the activity of the purified dTIP60^{piccolo} complexes. Analogous to MOF, Tip60 alone can only acetylate free histones and requires association with EPc and Ing3 to acetylate nucleosomes [191, 197]. To compare the activities of the two complexes, I used the same experimental setup as for the MSL complex. The H4K12ac was observed in a 60 min HAT reaction with different dTIP60^{piccolo} concentrations, giving rise to a detectable H4K12ac signal between 10 and 25 nM (Figure 43C). However, the subsequent HAT assays coupled to MS analysis were performed at 50 nM dTIP60^{piccolo}. The reaction was incubated for different times from 5 to 60 min. 50 nM of dTIP60^{piccolo} were incubated

Results

with 200 nM nucleosome arrays and 10 μM acetyl-CoA. The levels of acetylation were determined by mass spectrometry with three independent dTIP60^{piccolo} preparations (Figure 44A).

A

t [min]	0	without RNase 5	15	30	60	with RNaseA 5	15	30	60
non-ac	100	98	96	92	89	99	97	90	86
H4K12ac		0	2	3	5	0	1	4	6
H4K5ac		1	1	2	2	0	1	2	2
H4K16ac		0	0	1	1	0	0	1	2
H4K8ac		0	0	1	1	0	0	1	1
di-ac		0		0	0		0	0	1
tri-ac				0	0			0	0
tetra-ac				0	0		0	0	0

%
100
75
50
25
0
n = 3

B
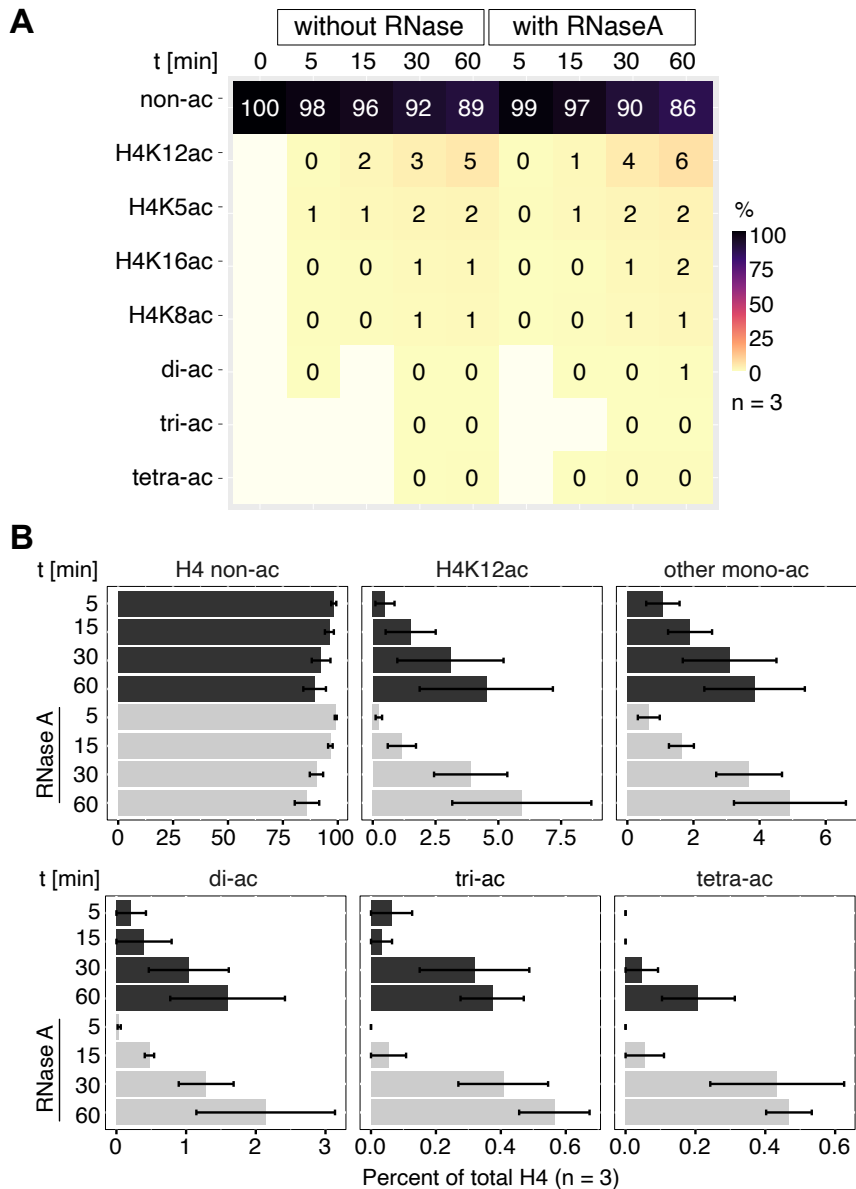
Figure 44: dTIP60^{piccolo} acetylates H4K12 and H4K5 *in vitro* in a time-dependent manner both in presence and absence of RNA. **A** Heatmap of the MS analysis of the acetylation patterns, mean of three independent protein complex preparations. Lower detection limit 0.003% indicated by white tiles, 0% indicates an acetylation < 0.5%. **C** Bar graph summarizing the MS analysis and displaying the error as standard error of the means of the three independent protein preparations [216].

After 60 min the H4K12 mono-acetylation reached on average 5%. As this appeared low compared to levels reached by the 4-MSL complex of 54% at 5 min, I considered if the high RNA content co-purified by the complex might inhibit the acetylation reaction. Therefore, I treated the HAT assay samples with RNase A to remove all RNA contamination. However, H4K12 mono-acetylation reached only 6% with RNase A treatment. This is not a significant difference considering the substantial error bar (Figure 44B). Besides H4K12ac, also H4K5 mono-acetylation was detected and increased over time up to 2% on average. Furthermore, H4K16ac and H4K8ac were detected at 2% and 1%, respectively, after 60 min incubation. The measured di-acetylations, which include H4K12K16ac and H4K5K8ac, reach only

0.5% and 1% after 60 min. The sum of all possible di-acetylations (including K5K12ac) reach around 2 to 3% (Figure 44B). To quantify the exact di-acetylation patterns, MS^3 mass spectrometry would have to be applied, which was not done as this was not the major scope of this project. Since in the mass spectrometric analysis no difference between the RNase A treated versus non-treated HAT assay samples was shown, dTIP60^{piccolo} does not appear to change its acetylation activity or its lysine specificity in presence of RNA. Tri- and tetra-acetylations were rare, around 0.5% at maximum incubation time. It seems that dTIP60^{piccolo} does not operate by a processive reaction mechanism along the H4 tail as the 4-MSL complex does. The measured HAT activity might be too low to resolve processive kinetics in this setup.

To confirm the lack of responsiveness of the dTIP60^{piccolo} to RNA the MS analysis was repeated with a 2-fold molar excess of GFP-470 RNA in absence or presence of RNase A. As the concentration and nature of the bound RNA to the dTIP60^{piccolo} complex is difficult to determine, the non-specific GFP-RNA was added in excess to saturate the complex with RNA. The MS analysis ascertained that it did not alter the acetylation activity of the dTIP60^{piccolo} complex (Figure 45).
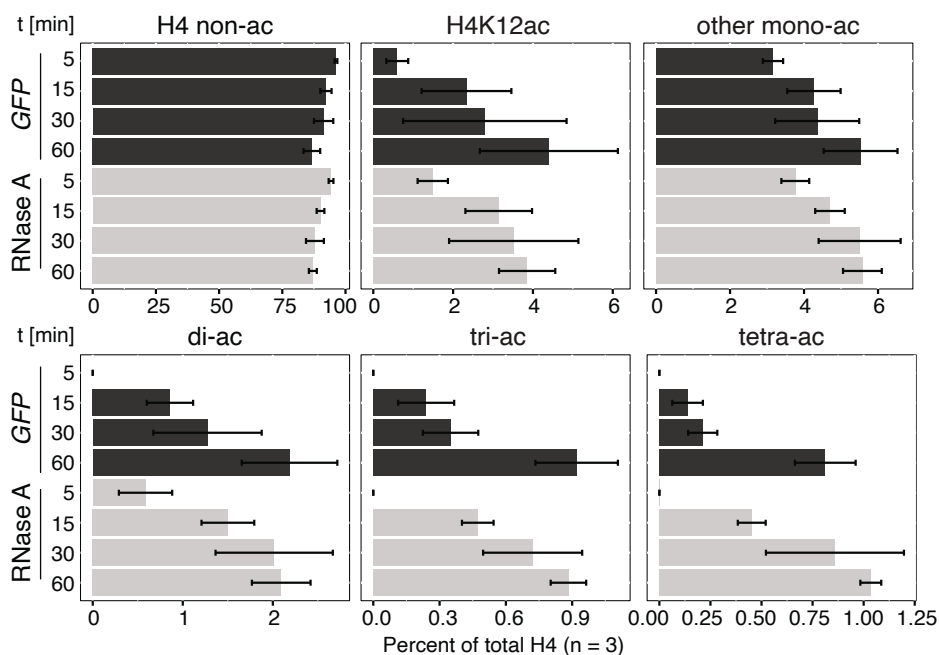
Figure 45: Additional GFP-RNA did not change the overall dTIP60^{piccolo} activity or specificity, as ascertained by MS analysis of the histone acetylation patterns. Bar plot summarizing the abundance of mono- and oligo-acetylated H4 tail motifs detected in (A). 'di-/tri-ac' represents the sum of all possible di- or tri-acetylated H4 tail motifs. 'other mono-ac' cumulates levels of K5ac, K8ac and K16ac. Standard error of the mean of 3 independent protein preparations is given. Figure adapted from [216].

All in all, dTIP60^{piccolo} acetylates H4K12 and K5, but the percentages are much lower than by the 4-MSL complex. There are no oligo-acetylation patterns detected even after long incubation times. The presence of RNA did not influence the acetylation reaction of dTIP60^{piccolo}.

8 Discussion

8.1 The dogma: the histone acetyltransferase MOF acetylates H4K16

The histone H4 acetyltransferase activity of MOF has been described more than two decades ago [95]. Ever since then, MOF was assumed to be specific for the H4K16ac, however even in that first study, slight H4K12ac was observed, but not further investigated. Later, MOF was found to be part of another protein complex, the non-specific lethal NSL complex [88, 89]. The NSL complex has a relaxed substrate specificity towards several lysines of the H4 tail, not specifically K16 [92]. This suggests that the subunits of the different complexes influence the specificity of MOF.

The acetylation activity of MOF on nucleosome arrays depends association with MSL1 and MSL3. However, the full MSL complex *in vivo* features additionally MSL2, at least one molecule of the lncRNA roX and MLE. This *in vitro* study delves into MOF in the MSL complex and its enzymatic activity and selectivity as a histone acetyltransferase (HAT) through detailed mass spectrometry analysis in contrast to traditional Western blotting as a readout.

8.2 H4K16ac and K12ac are part of oligo-acetylation patterns by the 4-MSL complex *in vitro*

When specific antibodies are used for detection by Western blot, indeed both H4K16ac and K12ac are detected after the HAT assay by the 4-MSL complex. Contrary to the expectation that incorporation of roX RNA would boost the acetylation reaction, the H4K12ac signal was suppressed by presence of long non-specific RNA. Moreover, by Western blot the H4K16ac signal appeared decreased, as well.

Detection and quantification of H4 acetylation patterns can be ascertained through mass spectrometry. Due to the exact mass determination and fragmentation of the peptides in MS-MS (or MS2), the exact position of the acetylation can be determined even for combinatorial patterns [201]. The combinatorial patterns are difficult to tease apart by Western blots, because the specific antibody would still recognize the K16ac within the K12K16 di-acetylation pattern, showcasing its excellent specificity. Oftentimes specific antibodies can detect the multiply acetylated peptides even better than the mono-acetylation [202, 203]. The accumulation of oligo-acetylation patterns explains the increased signal in both H4K16ac and K12ac in Western blots over time. In contrast, mass spectrometry resolves the different acetylation forms, such that over time K16ac is diminished at the expense of oligo-acetylation.

Oligo-acetylation patterns consistently included H4K16ac, which was ascertained as the initial acetylation at short incubation times (5 min). Subsequent patterns, at 15 and 30 min, involve di-acetylation of K12K16ac, followed by tri-acetylations and finally tetra-acetylation. This sequential pattern suggests a zipper-like outward progression of the MSL complex, processively acetylating lysine residues along the H4 tail (Figure 46, [216, 247]). This finding supports earlier observations of H4 tail

acetylation patterns in butyrate-treated HeLa cells, suggesting a model where acetylation begins at K16 and spreads outward towards the N-terminus [247]. Due to the experimental design in HeLa cells, it was postulated that 'inside-out' processivity a general feature of HATs, whereas HDACs would operate in the opposite direction 'outside-in' [247].
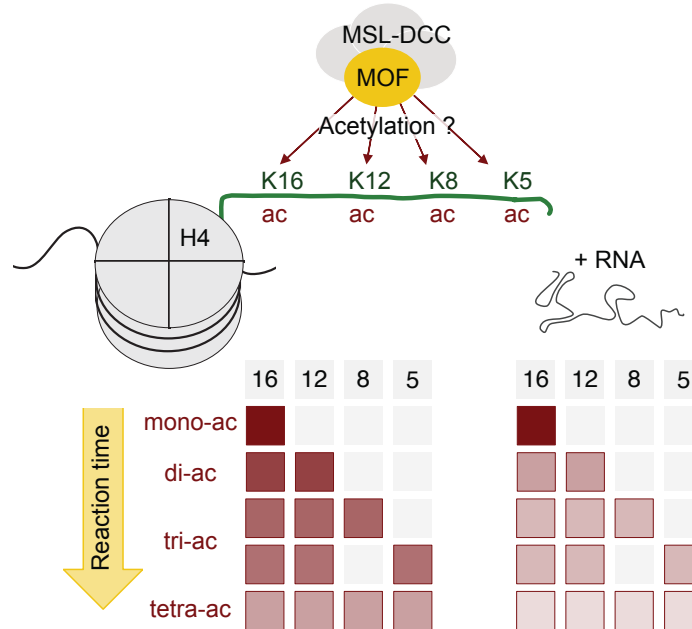
Figure 46: The MSL complex acetylates the H4 tail in a processive manner *in vitro*. The processivity along the H4 tail is suppressed by non-specific RNA. Consequently, the oligo-acetylation patterns are reduced in presence of RNA in favor of H4K16ac. Colored tiles represent acetylated lysine residue within the H4 tail. Figure adapted from Kiss *et al.*, 2024 [216].

8.3 RNA non-specifically suppressed oligo-acetylation patterns by reducing prolonged binding to nucleosomes

Contrary to the initial hypothesis, roX RNA did not enhance acetylation activity but suppressed oligo-acetylation patterns (Figure 46, [216]). Moreover, not only roX RNA but also non-specific long GFP transcripts reduced oligo-acetylation patterns (Figure 47C, D, [201]). Curiously, *in vivo* chromatin is almost devoid of oligo-acetylation patterns. On the one hand this could be due to the impact of histone deacetylases (HDACs), which are not represented in the *in vitro* experiment. On the other hand, this is a possible role for the RNA as well *in vivo*. In nuclei RNA is abundant, for instance nascent hnRNA, especially at highly transcribed genes on the single male X chromosome (Figure 47D). However, roX RNA is bound by the MSL complex *in vivo*, thus roX RNA is enriched at binding sites of the MSL (Figure 47C, [38-40, 57]).
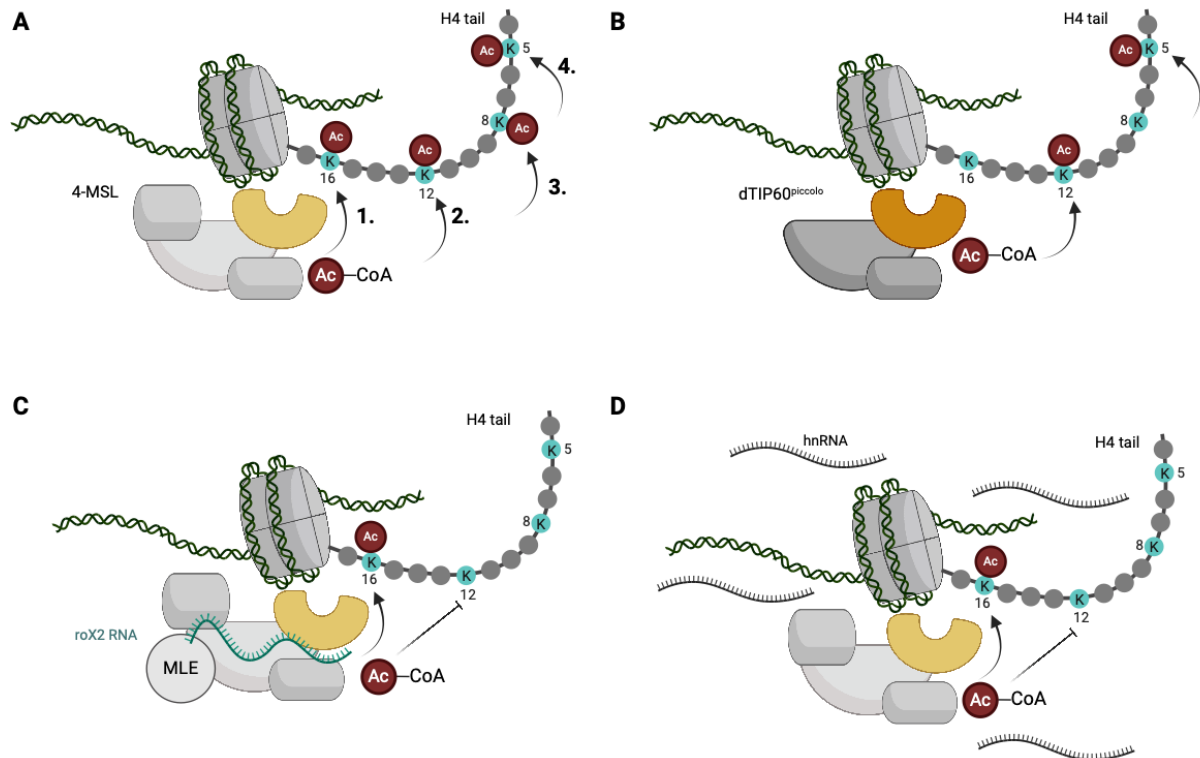
Figure 47: RNA suppresses oligo-acetylation patterns by the 4-MSL complex, but does not influence acetylation activity of dTIP60^{piccolo}. **A** Inside-out processivity of acetylation by the 4-MSL complex from H4K16 to K5 in absence of RNA. **B** dTIP60^{piccolo} solely acetylates H4K5 and K12 without processivity irrespective of RNA content. **C** *roX2* RNA inhibited the oligo-acetylation by the 4-MSL complex. **D** Long, nascent, non-specific RNA inhibited the oligo-acetylation and processivity, as well. Figure created with Biorender.

One supporting piece of data was obtained by collaboration with Dr. Marisa Müller, who performed electrophoretic mobility shift assays (EMSAs) with 4-MSL complexes on mono-nucleosome substrates (Figure 41). Intriguingly, long RNAs prevented the characteristic band shifts of a tight binding of the 4-MSL to the nucleosome, which indicates facilitated and faster dissociation of the MSL complex in presence of long RNA (Figure 48).

Discussion
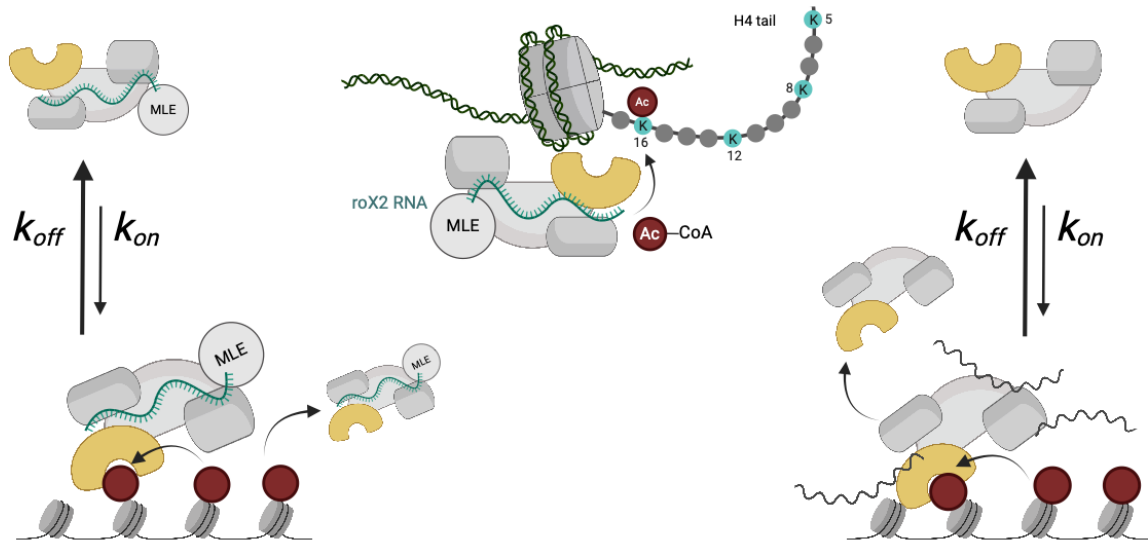
Figure 48: The dissociation rate of 4-MSL from the nucleosomes may be higher in presence of roX2 RNA (left) or non-specific RNA (right), thereby preventing the oligo-acetylation and facilitating 'spreading' to the next nucleosome. In male fly cells both mechanisms may contribute to the effect. Figure created with Biorender.

A higher dissociation rate may explain the 'spreading' mechanism, a long-standing concept in the field [29, 84]. The spreading hypothesis revolves around the idea that the MSL complex enters the chromatin of the male X chromosome at pioneering sites on the X (PionX) by sequence and shape recognition by MSL2 [40, 67, 68]. In the next step, the complex recognizes active genes marked by H3K36me3 and acetylates H4K16 in these regions [76]. It would be intriguing to show, whether there is a specific role for roX RNA in spreading as suggested by previous reports [248]. Conceivably, unbinding from nucleosomes and 'spreading' would be 'roX-specific' due to the high local concentration of incorporated roX in the MSL complex (Figure 48). Nevertheless, in this study I show that other non-specific RNAs may play a role as well, for example nascent RNA from the highly transcribed active genes (Figure 48). Certainly, *in vivo* both aspects may contribute to the mechanism of 'spreading' the MSL complex along the male X chromosome.

Future investigations should determine the dissociation constant in the absence and presence of RNA to validate this hypothesis. Furthermore, the characteristic features of RNA impacting the effect, like sequence or secondary structure should be ascertained. Moreover, *in vivo* studies using roX knock-out cells should explore the specific role for roX apart from other RNA in the nucleus. Moreover, nucleosome arrays do not reflect the complexity of chromatin and it is likely that other components may play a role. An elegant workaround could be the *in vitro* chromatin reconstitution system of chromatin from *Drosophila* embryo extracts (DREX) [68, 249]. These *in vitro* reconstitutions of complex chromatin approach physiological embryonic chromatin [68, 249, 250]. Nonetheless, these *Drosophila* histones from the DREX already carry histone marks, for example H4K12ac, which lead to a substantial background for the HAT assays [251].

All assays have been performed on non-modified, recombinant histone octamers, which do not carry, for example, H3K36 methylation. The H3K36me3 modification has been introduced as a recognition mechanism of the MSL complex to find its *in vivo* targets, the active genes on the male X chromosome

[75, 252-254]. However, my experiments show that H3K36me3 is not required to acetylate the H4 tail efficiently *in vitro*. In the future, one should compare the H3K36me3-modified nucleosome arrays to the unmodified substrate to see any impact on the acetylation selectivity or activity.

8.4 H4K16 is not required for the 4-MSL complex to recognize and acetylate the nucleosome

In collaboration with Dr. Dilan Pathirana and Prof. Dr. Jan Hasenauer (both University of Bonn) a mathematical model of simulated progressions of the MOF acetylation along the H4 tail was performed. This modeling and simulation lead to the conclusion that MOF is a processive enzyme and stays close to a nucleosome substrate after H4K16 monoacetylation.

Remarkably, the existence of K16 in the H4 tail was not strictly required. When an H4K16R mutation was introduced, which cannot be acetylated at the arginine replacing the lysine, the 4-MSL complex acetylated H4K12 and oligo-acetylation patterns to the same level with wild-type nucleosomes. While the 'inside-out' zipper model was suggested for HATs in HeLa cell experiments earlier [247], my data combined with the modelling provides proof for MOF as a HAT *in vitro [216]*.

8.5 Tip60 compared to MOF – two birds of a feather, each with its own song

Albeit having a HAT of the same enzymatic family as the catalytic subunit, the dTIP60[piccolo] complex and the 4-MSL complex seem to approach the nucleosome substrate in different ways. The MS analysis revealed a much lower activity for the dTIP60[piccolo] complex in the HAT assays compared to the 4-MSL complex. Moreover, the activity was limited to H4K12ac and K5ac mono-acetylations and did not proceed to considerable amounts oligo-acetylation patterns. Even though a processive reaction mechanism was suggested for Esa1, the yeast Tip60 homolog, there are as well data counterarguing such a 'tethered' reaction mechanism for Esa1 [246, 255]. Furthermore, the presence or absence of non-specific RNA did not influence the reaction specificity or the amount of acetylation.

Tip60 and MOF belong to the same MYST HAT family, which involves a conserved catalytic center. Nevertheless, the complexes' composition guides the substrate selectivity in both cases as observed for the MSL versus the NSL complex for instance [200]. On the same note, for Tip60 many substrates have been described and they seem to be selected not only by Tip60 itself but other members of the complex [244, 245, 256, 257]. Since the dTIP60[piccolo] complex is a small subcomplex of the 18 subunits complex of the DOMINO complex of *D. melanogaster [194, 258, 259]*, it is very probable, that there are subunits missing for optimal functionality *in vitro*. One candidate for boosting the acetylation activity might be Eaf6, which has been shown to stabilize the piccolo complex in yeast [189, 256, 260].

In this work, new evidence is provided, that RNA in the MSL complex can contribute to such substrate selectivity. In contrast, the presence of RNA did not interfere with the activity of the dTIP60[piccolo]. Interestingly, a lot of non-specific RNA was co-purified with the dTIP60[piccolo] complex, originating from

Discussion

the host expression cells Sf21. In the *Drosophila melanogaster* brain, Tip60 has been found to bind pre-mRNA and to be involved in alternative splicing [261]. Nevertheless, the specific Tip60 target RNAs unlikely to be present in Sf21 cells and to compensate for these missing components the complex might just associate with non-specific transcripts, suggesting a high affinity of the dTIP60^{piccolo} for RNA.

In the future it would be interesting to explore the Tip60 HAT activity further *in vitro*, for example with different subcomplexes, including more subunits, e.g., Eaf6, or the DOMINO holo-complex. Furthermore, the activity and specificity of Tip60 towards H2A.V could be further characterized. Moreover, it is well known, that not only Tip60, but also MOF acetylate further target proteins than histones *in vivo*. For instance, Tip60 performs autoacetylation, acetylation of cell-cycle and DNA-damage involved kinases, other proteins of the complex [72, 256, 262, 263]. In addition to autoacetylation and MSL protein acetylation MOF acetylates mitochondrial proteins [107, 108], is involved in metabolic pathways [109], can acetylate lamin [110], the histone demethylase LSD1 [111], the E3 ligase UHRF1 [112], the transcription factor YY1 [113], the immune response factor IRF3 [114], and the list is tentatively incomplete. Hence, it is intriguing to generate an acetylome for each of the HATs, which is a proteomic dataset, in which all potential acetylation sites by the acetyltransferase are evaluated [264].

9 References

1. Flemming, W. (1882). Zellsubstanz, Kern und Zelltheilung. Vogel.

2. Olins, D. E. & Olins, A. L. (2003). Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol. **4**, 809-814. https://doi.org/10.1038/nrm1225.

3. Avery, O. T., Macleod, C. M. & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med. **79**, 137-158. https://doi.org/10.1084/jem.79.2.137.

4. Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. **171**, 737-738. https://doi.org/10.1038/171737a0.

5. Franklin, R. E. & Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. Nature. **171**, 740-741. https://doi.org/10.1038/171740a0.

6. Wilkins, M. H., Stokes, A. R. & Wilson, H. R. (1953). Molecular structure of deoxypentose nucleic acids. Nature. **171**, 738-740. https://doi.org/10.1038/171738a0.

7. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible roles in the regulation of RNA synthesis. Proc Natl Acad Sci U S A. **51**, 786-794. https://doi.org/10.1073/pnas.51.5.786.

8. Allis, C. D., Jenuwein, T. & Reinberg, D. (2007). Epigenetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

9. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., Walter, P., Wilson, J. & Hunt, T. (2015). Molecular biology of the cell. Garland Science, New York, NY.

10. Workman, J. L. & Abmayr, S. M. (2013). Fundamentals of Chromatin. Springer New York.

11. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature. **389**, 251-260. https://doi.org/10.1038/38444.

12. Jenuwein, T. & Allis, C. D. (2001). Translating the histone code. Science. **293**, 1074-1080. https://doi.org/10.1126/science.1063127.

13. Soshnev, A. A., Josefowicz, S. Z. & Allis, C. D. (2016). Greater Than the Sum of Parts: Complexity of the Dynamic Epigenome. Mol Cell. **62**, 681-694. https://doi.org/10.1016/j.molcel.2016.05.004.

14. Allis, C. D., Berger, S. L., Côté, J., Dent, S., Jenuwein, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J. & Zhang, Y. (2007). New nomenclature for chromatin-modifying enzymes. Cell. **131**, 633-636. https://doi.org/10.1016/j.cell.2007.10.039.

15. Gowans, G. J., Bridgers, J. B., Zhang, J., Dronamraju, R., Burnetti, A., King, D. A., Thiengmany, A. V., Shinsky, S. A., Bhanu, N. V., Garcia, B. A., Buchler, N. E., Strahl, B. D. & Morrison, A. J. (2019). Recognition of Histone Crotonylation by Taf14 Links Metabolic State to Gene Expression. Mol Cell. **76**, 909-921 e903. https://doi.org/10.1016/j.molcel.2019.09.029.

16. Barnes, C. E., English, D. M. & Cowley, S. M. (2019). Acetylation & Co: an expanding repertoire of histone acylations regulates chromatin and transcription. Essays Biochem. **63**, 97-107. https://doi.org/10.1042/EBC20180061.

17. Lu-Culligan, W. J., Connor, L. J., Xie, Y., Ekundayo, B. E., Rose, B. T., Machyna, M., Pintado-Urbanc, A. P., Zimmer, J. T., Vock, I. W., Bhanu, N. V., King, M. C., Garcia, B. A., Bleichert, F. & Simon, M. D. (2023). Acetyl-methyllysine marks chromatin at active transcription start sites. Nature. **622**, 173-179. https://doi.org/10.1038/s41586-023-06565-9.

18. Kalashnikova, A. A., Porter-Goff, M. E., Muthurajan, U. M., Luger, K. & Hansen, J. C. (2013). The role of the nucleosome acidic patch in modulating higher order chromatin structure. J R Soc Interface. **10**, 20121022. https://doi.org/10.1098/rsif.2012.1022.

References

19. Dueva, R., Akopyan, K., Pederiva, C., Trevisan, D., Dhanjal, S., Lindqvist, A. & Farnebo, M. (2019). Neutralization of the Positive Charges on Histone Tails by RNA Promotes an Open Chromatin Structure. Cell Chem Biol. **26**, 1436-1449 e1435. https://doi.org/10.1016/j.chembiol.2019.08.002.

20. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J Mol Biol. **327**, 85-96. https://doi.org/10.1016/s0022-2836(03)00025-1.

21. Zhou, B. R., Feng, H., Ghirlando, R., Kato, H., Gruschus, J. & Bai, Y. (2012). Histone H4 K16Q mutation, an acetylation mimic, causes structural disorder of its N-terminal basic patch in the nucleosome. J Mol Biol. **421**, 30-37. https://doi.org/10.1016/j.jmb.2012.04.032.

22. Allis, C. D. & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat Rev Genet. **17**, 487-500. https://doi.org/10.1038/nrg.2016.59.

23. Kuroda, M. I., Hilfiker, A. & Lucchesi, J. C. (2016). Dosage Compensation in Drosophila-a Model for the Coordinate Regulation of Transcription. Genetics. **204**, 435-450. https://doi.org/10.1534/genetics.115.185108.

24. Graves, J. A. (2016). Evolution of vertebrate sex chromosomes and dosage compensation. Nat Rev Genet. **17**, 33-46. https://doi.org/10.1038/nrg.2015.2.

25. Boveri, T. (1904). Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns. Verlag von Gustav Fischer in Jena.

26. Morgan, T. H. (1910). Sex limited inheritance in *Drosophila*. Science. **32**, 120-122. https://doi.org/10.1126/science.32.812.120.

27. Disteche, C. M. (2016). Dosage compensation of the sex chromosomes and autosomes. Semin Cell Dev Biol. **56**, 9-18. https://doi.org/10.1016/j.semcdb.2016.04.013.

28. Lucchesi, J. C. (2018). Transcriptional modulation of entire chromosomes: dosage compensation. J Genet. **97**, 357-364. https://doi.org/10.1007/s12041-018-0919-7.

29. Samata, M. & Akhtar, A. (2018). Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. Annu Rev Biochem. **87**, 323-350. https://doi.org/10.1146/annurev-biochem-062917-011816.

30. Straub, T. & Becker, P. B. (2007). Dosage compensation: the beginning and end of generalization. Nat Rev Genet. **8**, 47-57. https://doi.org/10.1038/nrg2013.

31. Prestel, M., Feller, C. & Becker, P. B. (2010). Dosage compensation and the global re-balancing of aneuploid genomes. Genome Biol. **11**, 216. https://doi.org/10.1186/gb-2010-11-8-216.

32. Deng, X., Berletch, J. B., Nguyen, D. K. & Disteche, C. M. (2014). X chromosome regulation: diverse patterns in development, tissues and disease. Nat Rev Genet. **15**, 367-378. https://doi.org/10.1038/nrg3687.

33. Deng, X., Berletch, J. B., Ma, W., Nguyen, D. K., Hiatt, J. B., Noble, W. S., Shendure, J. & Disteche, C. M. (2013). Mammalian X upregulation is associated with enhanced transcription initiation, RNA half-life, and MOF-mediated H4K16 acetylation. Dev Cell. **25**, 55-68. https://doi.org/10.1016/j.devcel.2013.01.028.

34. Turner, B. M., Birley, A. J. & Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. Cell. **69**, 375-384. https://doi.org/10.1016/0092-8674(92)90417-b.

35. Cremer, T., Kurz, A., Zirbel, R., Dietzel, S., Rinke, B., Schröck, E., Speicher, M. R., Mathieu, U., Jauch, A., Emmerich, P., Scherthan, H., Ried, T., Cremer, C. & Lichter, P. (1993). Role of chromosome territories in the functional compartmentalization of the cell nucleus. Cold Spring Harb Symp Quant Biol. **58**, 777-792. https://doi.org/10.1101/sqb.1993.058.01.085.

36. Stadler, S., Schnapp, V., Mayer, R., Stein, S., Cremer, C., Bonifer, C., Cremer, T. & Dietzel, S. (2004). The architecture of chicken chromosome territories changes during differentiation. BMC Cell Biol. **5**, 44. https://doi.org/10.1186/1471-2121-5-44.

References

37. Cremer, T. & Cremer, C. (2006). Rise, fall and resurrection of chromosome territories: a historical perspective. Part II. Fall and resurrection of chromosome territories during the 1950s to 1980s. Part III. Chromosome territories and the functional nuclear architecture: experiments and models from the 1990s to the present. Eur J Histochem. **50**, 223-272. https://doi.org/10.4081/995.

38. Kelley, R. L., Meller, V. H., Gordadze, P. R., Roman, G., Davis, R. L. & Kuroda, M. I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell. **98**, 513-522. https://doi.org/10.1016/s0092-8674(00)81979-0.

39. Meller, V. H., Gordadze, P. R., Park, Y., Chu, X., Stuckenholz, C., Kelley, R. L. & Kuroda, M. I. (2000). Ordered assembly of roX RNAs into MSL complexes on the dosage-compensated X chromosome in Drosophila. Curr Biol. **10**, 136-143. https://doi.org/10.1016/s0960-9822(00)00311-0.

40. Villa, R., Schauer, T., Smialowski, P., Straub, T. & Becker, P. B. (2016). PionX sites mark the X chromosome for dosage compensation. Nature. **537**, 244-248. https://doi.org/10.1038/nature19338.

41. Tanaka, A., Fukunaga, A. & Oishi, K. (1976). Studies on the sex-specific lethals of Drosophila melanogaster. II. Further studies on a male-specific lethal gene, maleless. Genetics. **84**, 257-266. https://doi.org/10.1093/genetics/84.2.257.

42. Fukunaga, A., Tanaka, A. & Oishi, K. (1975). Maleless, a recessive autosomal mutant of Drosophila melanogaster that specifically kills male zygotes. Genetics. **81**, 135-141. https://doi.org/10.1093/genetics/81.1.135.

43. Belote, J. M. & Lucchesi, J. C. (1980). Control of X chromosome transcription by the maleless gene in Drosophila. Nature. **285**, 573-575. https://doi.org/10.1038/285573a0.

44. Belote, J. M. & Lucchesi, J. C. (1980). Male-specific lethal mutations of Drosophila melanogaster. Genetics. **96**, 165-186. https://doi.org/10.1093/genetics/96.1.165.

45. Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A. & Lucchesi, J. C. (1997). 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. **16**, 2054-2060. https://doi.org/10.1093/emboj/16.8.2054.

46. Akhtar, A., Zink, D. & Becker, P. B. (2000). Chromodomains are protein-RNA interaction modules. Nature. **407**, 405-409. https://doi.org/10.1038/35030169.

47. Franke, A. & Baker, B. S. (1999). The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila. Mol Cell. **4**, 117-122. https://doi.org/10.1016/s1097-2765(00)80193-8.

48. Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. & Davis, R. L. (1997). roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell. **88**, 445-457. https://doi.org/10.1016/s0092-8674(00)81885-1.

49. Hallacli, E., Lipp, M., Georgiev, P., Spielman, C., Cusack, S., Akhtar, A. & Kadlec, J. (2012). Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in Drosophila. Mol Cell. **48**, 587-600. https://doi.org/10.1016/j.molcel.2012.09.014.

50. Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S. & Akhtar, A. (2011). Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. Nat Struct Mol Biol. **18**, 142-149. https://doi.org/10.1038/nsmb.1960.

51. Piovesan, D., Del Conte, A., Clementel, D., Monzon, A. M., Bevilacqua, M., Aspromonte, M. C., Iserte, J. A., Orti, F. E., Marino-Buslje, C. & Tosatto, S. C. E. (2023). MobiDB: 10 years of intrinsically disordered proteins. Nucleic Acids Res. **51**, 438-444. https://doi.org/10.1093/nar/gkac1065.

52. Prabu, J. R., Müller, M., Thomae, A. W., Schüssler, S., Bonneau, F., Becker, P. B. & Conti, E. (2015). Structure of the RNA Helicase MLE Reveals the Molecular Mechanisms for Uridine Specificity and RNA-ATP Coupling. Mol Cell. **60**, 487-499. https://doi.org/10.1016/j.molcel.2015.10.011.

References

53. Kelley, R. L., Solovyeva, I., Lyman, L. M., Richman, R., Solovyev, V. & Kuroda, M. I. (1995). Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. Cell. **81**, 867-877. https://doi.org/10.1016/0092-8674(95)90007-1.

54. Bhadra, U., Pal-Bhadra, M. & Birchler, J. A. (2000). Histone acetylation and gene expression analysis of sex lethal mutants in Drosophila. Genetics. **155**, 753-763. https://doi.org/10.1093/genetics/155.2.753.

55. Gebauer, F., Merendino, L., Hentze, M. W. & Valcarcel, J. (1998). The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. RNA. **4**, 142-150. https://doi.org/n.a. PMC1369603.

56. Kelley, R. L., Wang, J., Bell, L. & Kuroda, M. I. (1997). Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. Nature. **387**, 195-199. https://doi.org/10.1038/387195a0.

57. Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A. & Becker, P. B. (2008). The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. PLoS Genet. **4**, e1000302. https://doi.org/10.1371/journal.pgen.1000302.

58. Straub, T., Zabel, A., Gilfillan, G. D., Feller, C. & Becker, P. B. (2013). Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. Genome Res. **23**, 473-485. https://doi.org/10.1101/gr.146407.112.

59. Villa, R., Forné, I., Müller, M., Imhof, A., Straub, T. & Becker, P. B. (2012). MSL2 combines sensor and effector functions in homeostatic control of the Drosophila dosage compensation machinery. Mol Cell. **48**, 647-654. https://doi.org/10.1016/j.molcel.2012.09.012.

60. Grimaud, C. & Becker, P. B. (2009). The dosage compensation complex shapes the conformation of the X chromosome in Drosophila. Genes Dev. **23**, 2490-2495. https://doi.org/10.1101/gad.539509.

61. Li, F., Schiemann, A. H. & Scott, M. J. (2008). Incorporation of the noncoding roX RNAs alters the chromatin-binding specificity of the Drosophila MSL1/MSL2 complex. Mol Cell Biol. **28**, 1252-1264. https://doi.org/10.1128/MCB.00910-07.

62. Oh, H., Park, Y. & Kuroda, M. I. (2003). Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev. **17**, 1334-1339. https://doi.org/10.1101/gad.1082003.

63. Schunter, S., Villa, R., Flynn, V., Heidelberger, J. B., Classen, A. K., Beli, P. & Becker, P. B. (2017). Ubiquitylation of the acetyltransferase MOF in Drosophila melanogaster. PLoS One. **12**, e0177408. https://doi.org/10.1371/journal.pone.0177408.

64. Fauth, T., Müller-Planitz, F., König, C., Straub, T. & Becker, P. B. (2010). The DNA binding CXC domain of MSL2 is required for faithful targeting the Dosage Compensation Complex to the X chromosome. Nucleic Acids Res. **38**, 3209-3221. https://doi.org/10.1093/nar/gkq026.

65. Zheng, S., Villa, R., Wang, J., Feng, Y., Wang, J., Becker, P. B. & Ye, K. (2014). Structural basis of X chromosome DNA recognition by the MSL2 CXC domain during Drosophila dosage compensation. Genes Dev. **28**, 2652-2662. https://doi.org/10.1101/gad.250936.114.

66. Zheng, S., Wang, J., Feng, Y., Wang, J. & Ye, K. (2012). Solution structure of MSL2 CXC domain reveals an unusual Zn3Cys9 cluster and similarity to pre-SET domains of histone lysine methyltransferases. PLoS One. **7**, e45437. https://doi.org/10.1371/journal.pone.0045437.

67. Albig, C., Tikhonova, E., Krause, S., Maksimenko, O., Regnard, C. & Becker, P. B. (2019). Factor cooperation for chromosome discrimination in Drosophila. Nucleic Acids Res. **47**, 1706-1724. https://doi.org/10.1093/nar/gky1238.

68. Eggers, N. & Becker, P. B. (2021). Cell-free genomics reveal intrinsic, cooperative and competitive determinants of chromatin interactions. Nucleic Acids Res. **49**, 7602-7617. https://doi.org/10.1093/nar/gkab558.

References

69. Eggers, N., Gkountromichos, F., Krause, S., Campos Sparr, A. & Becker, P. B. (2023). Physical interaction between MSL2 and CLAMP assures direct cooperativity and prevents competition at composite binding sites. Nucleic Acids Res. **51**, 9039-9054. https://doi.org/10.1093/nar/gkad680.

70. Valsecchi, C. I. K., Basilicata, M. F., Georgiev, P., Gaub, A., Seyfferth, J., Kulkarni, T., Panhale, A., Semplicio, G., Manjunath, V., Holz, H., Dasmeh, P. & Akhtar, A. (2021). RNA nucleation by MSL2 induces selective X chromosome compartmentalization. Nature. **589**, 137-142. https://doi.org/10.1038/s41586-020-2935-z.

71. Taipale, M., Rea, S., Richter, K., Vilar, A., Lichter, P., Imhof, A. & Akhtar, A. (2005). hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. Mol Cell Biol. **25**, 6798-6810. https://doi.org/10.1128/MCB.25.15.6798-6810.2005.

72. Gupta, A., Sharma, G. G., Young, C. S., Agarwal, M., Smith, E. R., Paull, T. T., Lucchesi, J. C., Khanna, K. K., Ludwig, T. & Pandita, T. K. (2005). Involvement of human MOF in ATM function. Mol Cell Biol. **25**, 5292-5305. https://doi.org/10.1128/MCB.25.12.5292-5305.2005.

73. Moore, S. A., Ferhatoglu, Y., Jia, Y., Al-Jiab, R. A. & Scott, M. J. (2010). Structural and biochemical studies on the chromo-barrel domain of male specific lethal 3 (MSL3) reveal a binding preference for mono- or dimethyllysine 20 on histone H4. J Biol Chem. **285**, 40879-40890. https://doi.org/10.1074/jbc.M110.134312.

74. Kim, D., Blus, B. J., Chandra, V., Huang, P., Rastinejad, F. & Khorasanizadeh, S. (2010). Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. Nat Struct Mol Biol. **17**, 1027-1029. https://doi.org/10.1038/nsmb.1856.

75. Larschan, E., Alekseyenko, A. A., Gortchakov, A. A., Peng, S., Li, B., Yang, P., Workman, J. L., Park, P. J. & Kuroda, M. I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell. **28**, 121-133. https://doi.org/10.1016/j.molcel.2007.08.011.

76. Sural, T. H., Peng, S., Li, B., Workman, J. L., Park, P. J. & Kuroda, M. I. (2008). The MSL3 chromodomain directs a key targeting step for dosage compensation of the Drosophila melanogaster X chromosome. Nat Struct Mol Biol. **15**, 1318-1325. https://doi.org/10.1038/nsmb.1520.

77. Stabell, M., Larsson, J., Aalen, R. B. & Lambertsson, A. (2007). Drosophila dSet2 functions in H3-K36 methylation and is required for development. Biochem Biophys Res Commun. **359**, 784-789. https://doi.org/10.1016/j.bbrc.2007.05.189.

78. Jayakrishnan, M., Havlová, M., Veverka, V., Regnard, C. & Becker, P. B. (2024). Genomic context-dependent histone H3K36 methylation by three *Drosophila* methyltransferases and implications for dedicated chromatin readers. bioRxiv. 2024.2002.2006.577191. https://doi.org/10.1101/2024.02.06.577191.

79. Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A. & Schübeler, D. (2008). 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. **28**, 3401-3409. https://doi.org/10.1128/MCB.00006-08.

80. Bell, O., Wirbelauer, C., Hild, M., Scharf, A. N., Schwaiger, M., MacAlpine, D. M., Zilbermann, F., van Leeuwen, F., Bell, S. P., Imhof, A., Garza, D., Peters, A. H. & Schübeler, D. (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in Drosophila. EMBO J. **26**, 4974-4984. https://doi.org/10.1038/sj.emboj.7601926.

81. Sharda, A. & Humphrey, T. C. (2022). The role of histone H3K36me3 writers, readers and erasers in maintaining genome stability. DNA Repair (Amst). **119**, 103407. https://doi.org/10.1016/j.dnarep.2022.103407.

82. Buscaino, A., Köcher, T., Kind, J. H., Holz, H., Taipale, M., Wagner, K., Wilm, M. & Akhtar, A. (2003). MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. Mol Cell. **11**, 1265-1277. https://doi.org/10.1016/s1097-2765(03)00140-0.

References

83. Lindehell, H., Glotov, A., Dorafshan, E., Schwartz, Y. B. & Larsson, J. (2021). The role of H3K36 methylation and associated methyltransferases in chromosome-specific gene regulation. Sci Adv. **7**, eabh4390. https://doi.org/10.1126/sciadv.abh4390.

84. Keller, C. I. & Akhtar, A. (2015). The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. Curr Opin Genet Dev. **31**, 1-11. https://doi.org/10.1016/j.gde.2015.03.007.

85. Prakash, S. K., Van den Veyver, I. B., Franco, B., Volta, M., Ballabio, A. & Zoghbi, H. Y. (1999). Characterization of a novel chromo domain gene in xp22.3 with homology to Drosophila msl-3. Genomics. **59**, 77-84. https://doi.org/10.1006/geno.1999.5844.

86. Basilicata, M. F., Bruel, A. L., Semplicio, G., Valsecchi, C. I. K., Aktaş, T., Duffourd, Y., Rumpf, T., Morton, J., Bache, I., Szymanski, W. G., Gilissen, C., Vanakker, O., Õunap, K., Mittler, G., van der Burgt, I., El Chehadeh, S., Cho, M. T., Pfundt, R., Tan, T. Y., ... & Study, D. (2018). De novo mutations in MSL3 cause an X-linked syndrome marked by impaired histone H4 lysine 16 acetylation. Nat Genet. **50**, 1442-1451. https://doi.org/10.1038/s41588-018-0220-y.

87. Brunet, T., McWalter, K., Mayerhanser, K., Anbouba, G. M., Armstrong-Javors, A., Bader, I., Baugh, E., Begtrup, A., Bupp, C. P., Callewaert, B. L., Cereda, A., Cousin, M. A., Del Rey Jimenez, J. C., Demmer, L., Dsouza, N. R., Fleischer, N., Gavrilova, R. H., Ghate, S., Graf, E., ... & Wagner, M. (2021). Defining the genotypic and phenotypic spectrum of X-linked MSL3-related disorder. Genet Med. **23**, 384-395. https://doi.org/10.1038/s41436-020-00993-y.

88. Cai, Y., Jin, J., Swanson, S. K., Cole, M. D., Choi, S. H., Florens, L., Washburn, M. P., Conaway, J. W. & Conaway, R. C. (2010). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J Biol Chem. **285**, 4268-4272. https://doi.org/10.1074/jbc.C109.087981.

89. Raja, S. J., Charapitsa, I., Conrad, T., Vaquerizas, J. M., Gebhardt, P., Holz, H., Kadlec, J., Fraterman, S., Luscombe, N. M. & Akhtar, A. (2010). The nonspecific lethal complex is a transcriptional regulator in Drosophila. Mol Cell. **38**, 827-841. https://doi.org/10.1016/j.molcel.2010.05.021.

90. Feller, C., Prestel, M., Hartmann, H., Straub, T., Söding, J. & Becker, P. B. (2012). The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. Nucleic Acids Res. **40**, 1509-1522. https://doi.org/10.1093/nar/gkr869.

91. Lam, K. C., Mühlpfordt, F., Vaquerizas, J. M., Raja, S. J., Holz, H., Luscombe, N. M., Manke, T. & Akhtar, A. (2012). The NSL complex regulates housekeeping genes in Drosophila. PLoS Genet. **8**, e1002736. https://doi.org/10.1371/journal.pgen.1002736.

92. Radzisheuskaya, A., Shliaha, P. V., Grinev, V. V., Shlyueva, D., Damhofer, H., Koche, R., Gorshkov, V., Kovalchuk, S., Zhan, Y., Rodriguez, K. L., Johnstone, A. L., Keogh, M. C., Hendrickson, R. C., Jensen, O. N. & Helin, K. (2021). Complex-dependent histone acetyltransferase activity of KAT8 determines its role in transcription and cellular homeostasis. Mol Cell. **81**, 1749-1765 e1748. https://doi.org/10.1016/j.molcel.2021.02.012.

93. Iyer, S. S., Sun, Y., Seyfferth, J., Manjunath, V., Samata, M., Alexiadis, A., Kulkarni, T., Gutierrez, N., Georgiev, P., Shvedunova, M. & Akhtar, A. (2023). The NSL complex is required for piRNA production from telomeric clusters. Life Science Alliance. **6**, e202302194. https://doi.org/10.26508/lsa.202302194.

94. Tsang, T. H., Wiese, M., Helmstädter, M., Stehle, T., Seyfferth, J., Shvedunova, M., Holz, H., Walz, G. & Akhtar, A. (2023). Transcriptional regulation by the NSL complex enables diversification of IFT functions in ciliated versus nonciliated cells. Sci Adv. **9**, eadh5598. https://doi.org/10.1126/sciadv.adh5598.

95. Akhtar, A. & Becker, P. B. (2000). Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol Cell. **5**, 367-375. https://doi.org/10.1016/s1097-2765(00)80431-1.

References

96. Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D. & Lucchesi, J. C. (2000). The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. Mol Cell Biol. **20**, 312-318. https://doi.org/10.1128/MCB.20.1.312-318.2000.

97. Gu, W., Szauter, P. & Lucchesi, J. C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster. Dev Genet. **22**, 56-64. https://doi.org/10.1002/(SICI)1520-6408(1998)22:1<56::AID-DVG6>3.0.CO;2-6.

98. Conrad, T., Cavalli, F. M., Holz, H., Hallacli, E., Kind, J., Ilik, I., Vaquerizas, J. M., Luscombe, N. M. & Akhtar, A. (2012). The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. Dev Cell. **22**, 610-624. https://doi.org/10.1016/j.devcel.2011.12.016.

99. Nielsen, P. R., Nietlispach, D., Buscaino, A., Warner, R. J., Akhtar, A., Murzin, A. G., Murzina, N. V. & Laue, E. D. (2005). Structure of the chromo barrel domain from the MOF acetyltransferase. J Biol Chem. **280**, 32326-32331. https://doi.org/10.1074/jbc.M501347200.

100. Jacobs, S. A. & Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science. **295**, 2080-2083. https://doi.org/10.1126/science.1069473.

101. Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V. & Laue, E. D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature. **416**, 103-107. https://doi.org/10.1038/nature722.

102. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389-3402. https://doi.org/10.1093/nar/25.17.3389.

103. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol. **215**, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.

104. Eisen, A., Utley, R. T., Nourani, A., Allard, S., Schmidt, P., Lane, W. S., Lucchesi, J. C. & Côté, J. (2001). The yeast NuA4 and *Drosophila* MSL complexes contain homologous subunits important for transcription regulation. J Biol Chem. **276**, 3484-3491. https://doi.org/10.1074/jbc.M008159200.

105. Doyon, Y. & Côté, J. (2004). The highly conserved and multifunctional NuA4 HAT complex. Curr Opin Genet Dev. **14**, 147-154. https://doi.org/10.1016/j.gde.2004.02.009.

106. Yang, C., Wu, J., Sinha, S. H., Neveu, J. M. & Zheng, Y. G. (2012). Autoacetylation of the MYST lysine acetyltransferase MOF protein. J Biol Chem. **287**, 34917-34926. https://doi.org/10.1074/jbc.M112.359356.

107. Chatterjee, A., Seyfferth, J., Lucci, J., Gilsbach, R., Preissl, S., Böttinger, L., Mårtensson, C. U., Panhale, A., Stehle, T., Kretz, O., Sahyoun, A. H., Avilov, S., Eimer, S., Hein, L., Pfanner, N., Becker, T. & Akhtar, A. (2016). MOF Acetyl Transferase Regulates Transcription and Respiration in Mitochondria. Cell. **167**, 722-738 e723. https://doi.org/10.1016/j.cell.2016.09.052.

108. Guhathakurta, S., Erdogdu, N. U., Hoffmann, J. J., Grzadzielewska, I., Schendzielorz, A., Seyfferth, J., Mårtensson, C. U., Corrado, M., Karoutas, A., Warscheid, B., Pfanner, N., Becker, T. & Akhtar, A. (2023). COX17 acetylation via MOF-KANSL complex promotes mitochondrial integrity and function. Nat Metab. https://doi.org/10.1038/s42255-023-00904-w.

109. Pessoa Rodrigues, C., Chatterjee, A., Wiese, M., Stehle, T., Szymanski, W., Shvedunova, M. & Akhtar, A. (2021). Histone H4 lysine 16 acetylation controls central carbon metabolism and diet-induced obesity in mice. Nat Commun. **12**, 6212. https://doi.org/10.1038/s41467-021-26277-w.

110. Karoutas, A., Szymanski, W., Rausch, T., Guhathakurta, S., Rog-Zielinska, E. A., Peyronnet, R., Seyfferth, J., Chen, H. R., de Leeuw, R., Herquel, B., Kimura, H., Mittler, G., Kohl, P., Medalia, O., Korbel, J. O. & Akhtar, A. (2019). The NSL complex maintains nuclear architecture stability via lamin A/C acetylation. Nat Cell Biol. **21**, 1248-1260. https://doi.org/10.1038/s41556-019-0397-z.

References

111. Luo, H., Shenoy, A. K., Li, X., Jin, Y., Jin, L., Cai, Q., Tang, M., Liu, Y., Chen, H., Reisman, D., Wu, L., Seto, E., Qiu, Y., Dou, Y., Casero, R. A. & Lu, J. (2016). MOF Acetylates the Histone Demethylase LSD1 to Suppress Epithelial-to-Mesenchymal Transition. Cell Rep. **15**, 2665-2678. https://doi.org/10.1016/j.celrep.2016.05.050.

112. Wang, L., Yang, X., Zhao, K., Huang, S., Qin, Y., Chen, Z., Hu, X., Jin, G. & Zhou, Z. (2024). MOF-mediated acetylation of UHRF1 enhances UHRF1 E3 ligase activity to facilitate DNA methylation maintenance. Cell Rep. **43**, 113908. https://doi.org/10.1016/j.celrep.2024.113908.

113. Wu, T., Zhao, B., Cai, C., Chen, Y., Miao, Y., Chu, J., Sui, Y., Li, F., Chen, W., Cai, Y., Wang, F. & Jin, J. (2023). The Males Absent on the First (MOF) Mediated Acetylation Alters the Protein Stability and Transcriptional Activity of YY1 in HCT116 Cells. Int J Mol Sci. **24**. https://doi.org/10.3390/ijms24108719.

114. Huai, W., Liu, X., Wang, C., Zhang, Y., Chen, X., Xu, S., Thomas, T., Li, N. & Cao, X. (2019). KAT8 selectively inhibits antiviral immunity by acetylating IRF3. J Exp Med. **216**, 772-785. https://doi.org/10.1084/jem.20181773.

115. Su, J., Wang, F., Cai, Y. & Jin, J. (2016). The Functional Analysis of Histone Acetyltransferase MOF in Tumorigenesis. Int J Mol Sci. **17**, 99. https://doi.org/10.3390/ijms17010099.

116. Kuroda, M. I., Kernan, M. J., Kreber, R., Ganetzky, B. & Baker, B. S. (1991). The maleless protein associates with the X chromosome to regulate dosage compensation in Drosophila. Cell. **66**, 935-947. https://doi.org/10.1016/0092-8674(91)90439-6.

117. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. **17**, 4713-4730. https://doi.org/10.1093/nar/17.12.4713.

118. Jagtap, P. K. A., Müller, M., Kiss, A. E., Thomae, A. W., Lapouge, K., Beck, M., Becker, P. B. & Hennig, J. (2023). Structural basis of RNA-induced autoregulation of the DExH-type RNA helicase maleless. Mol Cell. **83**, 4318-4333. https://doi.org/10.1016/j.molcel.2023.10.026.

119. Lee, C. G. & Hurwitz, J. (1993). Human RNA helicase A is homologous to the maleless protein of Drosophila. J Biol Chem. **268**, 16822-16830.

120. Zhang, S., Maacke, H. & Grosse, F. (1995). Molecular cloning of the gene encoding nuclear DNA helicase II. A bovine homologue of human RNA helicase A and Drosophila Mle protein. J Biol Chem. **270**, 16422-16427. https://doi.org/10.1074/jbc.270.27.16422.

121. Zhang, S. & Grosse, F. (1994). Nuclear DNA helicase II unwinds both DNA and RNA. Biochemistry. **33**, 3906-3912. https://doi.org/10.1021/bi00179a016.

122. Jain, A., Bacolla, A., Chakraborty, P., Grosse, F. & Vasquez, K. M. (2010). Human DHX9 helicase unwinds triple-helical DNA structures. Biochemistry. **49**, 6992-6999. https://doi.org/10.1021/bi100795m.

123. Lee, T. & Pelletier, J. (2016). The biology of DHX9 and its potential as a therapeutic target. Oncotarget. **7**, 42716-42739. https://doi.org/10.18632/oncotarget.8446.

124. Reenan, R. A., Hanrahan, C. J. & Ganetzky, B. (2000). 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. **25**, 139-149. https://doi.org/10.1016/s0896-6273(00)80878-8.

125. Bleichert, F. & Baserga, S. J. (2007). The long unwinding road of RNA helicases. Mol Cell. **27**, 339-352. https://doi.org/10.1016/j.molcel.2007.07.014.

126. Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem Sci. **36**, 19-29. https://doi.org/10.1016/j.tibs.2010.07.008.

127. Isken, O., Grassmann, C. W., Sarisky, R. T., Kann, M., Zhang, S., Grosse, F., Kao, P. N. & Behrens, S. E. (2003). Members of the NF90/NFAR protein group are involved in the life cycle of a positive-strand RNA virus. EMBO J. **22**, 5655-5665. https://doi.org/10.1093/emboj/cdg562.

References

128. Isken, O., Baroth, M., Grassmann, C. W., Weinlich, S., Ostareck, D. H., Ostareck-Lederer, A. & Behrens, S. E. (2007). Nuclear factors are involved in hepatitis C virus RNA replication. RNA. **13**, 1675-1692. https://doi.org/10.1261/rna.594207.

129. Tautz, N., Tews, B. A. & Meyers, G. (2015). The Molecular Biology of Pestiviruses. Adv Virus Res. **93**, 47-160. https://doi.org/10.1016/bs.aivir.2015.03.002.

130. Shi, R. Z., Pan, Y. Q. & Xing, L. (2021). RNA Helicase A Regulates the Replication of RNA Viruses. Viruses. **13**, 361. https://doi.org/10.3390/v13030361.

131. Cugusi, S., Kallappagoudar, S., Ling, H. & Lucchesi, J. C. (2015). The Drosophila Helicase Maleless (MLE) is Implicated in Functions Distinct From its Role in Dosage Compensation. Mol Cell Proteomics. **14**, 1478-1488. https://doi.org/10.1074/mcp.M114.040667.

132. Cugusi, S., Li, Y., Jin, P. & Lucchesi, J. C. (2016). The Drosophila Helicase MLE Targets Hairpin Structures in Genomic Transcripts. PLoS Genet. **12**, e1005761. https://doi.org/10.1371/journal.pgen.1005761.

133. Kernan, M. J., Kuroda, M. I., Kreber, R., Baker, B. S. & Ganetzky, B. (1991). napts, a mutation affecting sodium channel activity in Drosophila, is an allele of mle, a regulator of X chromosome transcription. Cell. **66**, 949-959. https://doi.org/10.1016/0092-8674(91)90440-a.

134. Wessels, H. H., Imami, K., Baltz, A. G., Kolinski, M., Beldovskaya, A., Selbach, M., Small, S., Ohler, U. & Landthaler, M. (2016). The mRNA-bound proteome of the early fly embryo. Genome Res. **26**, 1000-1009. https://doi.org/10.1101/gr.200386.115.

135. Jagtap, P. K. A., Müller, M., Masiewicz, P., von Bulow, S., Hollmann, N. M., Chen, P. C., Simon, B., Thomae, A. W., Becker, P. B. & Hennig, J. (2019). Structure, dynamics and roX2-lncRNA binding of tandem double-stranded RNA binding domains dsRBD1,2 of Drosophila helicase Maleless. Nucleic Acids Res. **47**, 4319-4333. https://doi.org/10.1093/nar/gkz125.

136. Lv, M., Yao, Y., Li, F., Xu, L., Yang, L., Gong, Q., Xu, Y. Z., Shi, Y., Fan, Y. J. & Tang, Y. (2019). Structural insights reveal the specific recognition of roX RNA by the dsRNA-binding domains of the RNA helicase MLE and its indispensable role in dosage compensation in Drosophila. Nucleic Acids Res. **47**, 3142-3157. https://doi.org/10.1093/nar/gky1308.

137. Maenner, S., Müller, M., Fröhlich, J., Langer, D. & Becker, P. B. (2013). ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. Mol Cell. **51**, 174-184. https://doi.org/10.1016/j.molcel.2013.06.011.

138. Izzo, A., Regnard, C., Morales, V., Kremmer, E. & Becker, P. B. (2008). Structure-function analysis of the RNA helicase maleless. Nucleic Acids Res. **36**, 950-962. https://doi.org/10.1093/nar/gkm1108.

139. Jarmoskaite, I. & Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. Annu Rev Biochem. **83**, 697-725. https://doi.org/10.1146/annurev-biochem-060713-035546.

140. Ozgur, S., Buchwald, G., Falk, S., Chakrabarti, S., Prabu, J. R. & Conti, E. (2015). The conformational plasticity of eukaryotic RNA-dependent ATPases. FEBS J. **282**, 850-863. https://doi.org/10.1111/febs.13198.

141. Gupta, A. & Gribskov, M. (2011). The role of RNA sequence and structure in RNA-protein interactions. J Mol Biol. **409**, 574-587. https://doi.org/10.1016/j.jmb.2011.04.007.

142. Chakraborty, P. & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA Repair (Amst). **10**, 654-665. https://doi.org/10.1016/j.dnarep.2011.04.013.

143. Müller, M., Schauer, T., Krause, S., Villa, R., Thomae, A. W. & Becker, P. B. (2020). Two-step mechanism for selective incorporation of lncRNA into a chromatin modifier. Nucleic Acids Res. **48**, 7483-7501. https://doi.org/10.1093/nar/gkaa492.

144. Ilik, I. A., Quinn, J. J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R. C., Luscombe, N., Backofen, R., Chang, H. Y. & Akhtar, A. (2013). Tandem stem-loops

References

in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. Mol Cell. **51**, 156-173. https://doi.org/10.1016/j.molcel.2013.07.001.

145. Statello, L., Guo, C. J., Chen, L. L. & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. **22**, 96-118. https://doi.org/10.1038/s41580-020-00315-9.

146. Rutenberg-Schoenberg, M., Sexton, A. N. & Simon, M. D. (2016). The Properties of Long Noncoding RNAs That Regulate Chromatin. Annu Rev Genomics Hum Genet. **17**, 69-94. https://doi.org/10.1146/annurev-genom-090314-024939.

147. Park, S. W., Kuroda, M. I. & Park, Y. (2008). Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in roX noncoding RNAs. Mol Cell Biol. **28**, 4952-4962. https://doi.org/10.1128/MCB.00219-08.

148. Amrein, H. & Axel, R. (1997). Genes expressed in neurons of adult male Drosophila. Cell. **88**, 459-469. https://doi.org/10.1016/s0092-8674(00)81886-3.

149. Park, Y., Oh, H., Meller, V. H. & Kuroda, M. I. (2005). Variable splicing of non-coding roX2 RNAs influences targeting of MSL dosage compensation complexes in Drosophila. RNA Biol. **2**, 157-164. https://doi.org/10.4161/rna.2.4.2473.

150. Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P. & Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. Cell Rep. **11**, 1110-1122. https://doi.org/10.1016/j.celrep.2015.04.023.

151. Guo, C. J., Ma, X. K., Xing, Y. H., Zheng, C. C., Xu, Y. F., Shan, L., Zhang, J., Wang, S., Wang, Y., Carmichael, G. G., Yang, L. & Chen, L. L. (2020). Distinct Processing of lncRNAs Contributes to Non-conserved Functions in Stem Cells. Cell. **181**, 621-636.e622. https://doi.org/10.1016/j.cell.2020.03.006.

152. Quinn, J. J., Zhang, Q. C., Georgiev, P., Ilik, I. A., Akhtar, A. & Chang, H. Y. (2016). Rapid evolutionary turnover underlies conserved lncRNA-genome interactions. Genes Dev. **30**, 191-207. https://doi.org/10.1101/gad.272187.115.

153. Quinn, J. J. & Chang, H. Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. **17**, 47-62. https://doi.org/10.1038/nrg.2015.10.

154. Moran, V. A., Perera, R. J. & Khalil, A. M. (2012). Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. Nucleic Acids Res. **40**, 6391-6400. https://doi.org/10.1093/nar/gks296.

155. Quinn, J. J., Ilik, I. A., Qu, K., Georgiev, P., Chu, C., Akhtar, A. & Chang, H. Y. (2014). Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification. Nat Biotechnol. **32**, 933-940. https://doi.org/10.1038/nbt.2943.

156. Ilik, I. A., Maticzka, D., Georgiev, P., Gutierrez, N. M., Backofen, R. & Akhtar, A. (2017). A mutually exclusive stem-loop arrangement in roX2 RNA is essential for X-chromosome regulation in. Genes Dev. **31**, 1973-1987. https://doi.org/10.1101/gad.304600.117.

157. Prayitno, K., Schauer, T., Regnard, C. & Becker, P. B. (2019). Progressive dosage compensation during Drosophila embryogenesis is reflected by gene arrangement. EMBO Rep. **20**, e48138. https://doi.org/10.15252/embr.201948138.

158. Lim, C. K. & Kelley, R. L. (2012). Autoregulation of the Drosophila Noncoding roX1 RNA Gene. PLoS Genet. **8**, e1002564. https://doi.org/10.1371/journal.pgen.1002564.

159. Kim, M., Faucillion, M. L. & Larsson, J. (2018). RNA-on-X 1 and 2 in Drosophila melanogaster fulfill separate functions in dosage compensation. PLoS Genet. **14**, e1007842. https://doi.org/10.1371/journal.pgen.1007842.

160. Rieder, L. E., Jordan, W. T. & Larschan, E. N. (2019). Targeting of the Dosage-Compensated Male X-Chromosome during Early Drosophila Development. Cell Rep. **29**, 4268-4275.e4262. https://doi.org/10.1016/j.celrep.2019.11.095.

References

161. Simon, M. D., Wang, C. I., Kharchenko, P. V., West, J. A., Chapman, B. A., Alekseyenko, A. A., Borowsky, M. L., Kuroda, M. I. & Kingston, R. E. (2011). The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A. **108**, 20497-20502. https://doi.org/10.1073/pnas.1113536108.

162. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell. **44**, 667-678. https://doi.org/10.1016/j.molcel.2011.08.027.

163. Villa, R., Jagtap, P. K. A., Thomae, A. W., Campos Sparr, A., Forné, I., Hennig, J., Straub, T. & Becker, P. B. (2021). Divergent evolution toward sex chromosome-specific gene regulation inDrosophila. Genes & Development. **35**, 1055-1070. https://doi.org/10.1101/gad.348411.121.

164. Graziadei, A. & Rappsilber, J. (2022). Leveraging crosslinking mass spectrometry in structural and cell biology. Structure. **30**, 37-54. https://doi.org/10.1016/j.str.2021.11.007.

165. O'Reilly, F. J. & Rappsilber, J. (2018). Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. Nat Struct Mol Biol. **25**, 1000-1008. https://doi.org/10.1038/s41594-018-0147-0.

166. Rappsilber, J. (2011). The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J Struct Biol. **173**, 530-540. https://doi.org/10.1016/j.jsb.2010.10.014.

167. Herzog, F. (2014). Measuring spatial restraints on native protein complexes using isotope-tagged chemical cross-linking and mass spectrometry. Methods Mol Biol. **1091**, 259-273. https://doi.org/10.1007/978-1-62703-691-7_19.

168. Iacobucci, C., Piotrowski, C., Aebersold, R., Amaral, B. C., Andrews, P., Bernfur, K., Borchers, C., Brodie, N. I., Bruce, J. E., Cao, Y., Chaignepain, S., Chavez, J. D., Claverol, S., Cox, J., Davis, T., Degliesposti, G., Dong, M. Q., Edinger, N., Emanuelsson, C., ... & Sinz, A. (2019). First Community-Wide, Comparative Cross-Linking Mass Spectrometry Study. Anal Chem. **91**, 6953-6961. https://doi.org/10.1021/acs.analchem.9b00658.

169. Forné, I., Ludwigsen, J., Imhof, A., Becker, P. B. & Müller-Planitz, F. (2012). Probing the conformation of the ISWI ATPase domain with genetically encoded photoreactive crosslinkers and mass spectrometry. Mol Cell Proteomics. **11**, M111 012088. https://doi.org/10.1074/mcp.M111.012088.

170. Harrer, N., Schindler, C. E. M., Brützel, L. K., Forné, I., Ludwigsen, J., Imhof, A., Zacharias, M., Lipfert, J. & Müller-Planitz, F. (2018). Structural Architecture of the Nucleosome Remodeler ISWI Determined from Cross-Linking, Mass Spectrometry, SAXS, and Modeling. Structure. **26**, 282-294 e286. https://doi.org/10.1016/j.str.2017.12.015.

171. Müller-Planitz, F. (2015). Crossfinder-assisted mapping of protein crosslinks formed by site-specifically incorporated crosslinkers. Bioinformatics. **31**, 2043-2045. https://doi.org/10.1093/bioinformatics/btv083.

172. Piersimoni, L., Kastritis, P. L., Arlt, C. & Sinz, A. (2022). Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions─A Method for All Seasons. Chem Rev. **122**, 7500-7531. https://doi.org/10.1021/acs.chemrev.1c00786.

173. Kleff, S., Andrulis, E. D., Anderson, C. W. & Sternglanz, R. (1995). Identification of a gene encoding a yeast histone H4 acetyltransferase. J Biol Chem. **270**, 24674-24677. https://doi.org/10.1074/jbc.270.42.24674.

174. Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y. & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell. **84**, 843-851. https://doi.org/10.1016/s0092-8674(00)81063-6.

175. Brownell, J. E. & Allis, C. D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. Curr Opin Genet Dev. **6**, 176-184. https://doi.org/10.1016/s0959-437x(96)80048-7.

References

176. Narita, T., Weinert, B. T. & Choudhary, C. (2019). Functions and mechanisms of non-histone protein acetylation. Nat Rev Mol Cell Biol. **20**, 156-174. https://doi.org/10.1038/s41580-018-0081-3.

177. Shang, S., Liu, J. & Hua, F. (2022). Protein acylation: mechanisms, biological functions and therapeutic targets. Signal Transduct Target Ther. **7**, 396. https://doi.org/10.1038/s41392-022-01245-y.

178. Sabari, B. R., Zhang, D., Allis, C. D. & Zhao, Y. (2017). Metabolic regulation of gene expression through histone acylations. Nat Rev Mol Cell Biol. **18**, 90-101. https://doi.org/10.1038/nrm.2016.140.

179. Chen, Y., Chen, W., Cobb, M. H. & Zhao, Y. (2009). PTMap-a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. Proc Natl Acad Sci U S A. **106**, 761-766. https://doi.org/10.1073/pnas.0811739106.

180. Xia, C., Tao, Y., Li, M., Che, T. & Qu, J. (2020). Protein acetylation and deacetylation: An important regulatory modification in gene transcription (Review). Exp Ther Med. **20**, 2923-2940. https://doi.org/10.3892/etm.2020.9073.

181. Marmorstein, R. & Zhou, M. M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb Perspect Biol. **6**, a018762. https://doi.org/10.1101/cshperspect.a018762.

182. Voss, A. K. & Thomas, T. (2018). Histone Lysine and Genomic Targets of Histone Acetyltransferases in Mammals. Bioessays. **40**, e1800078. https://doi.org/10.1002/bies.201800078.

183. Lee, K. K. & Workman, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat Rev Mol Cell Biol. **8**, 284-295. https://doi.org/10.1038/nrm2145.

184. Morales, V., Straub, T., Neumann, M. F., Mengus, G., Akhtar, A. & Becker, P. B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. EMBO J. **23**, 2258-2268. https://doi.org/10.1038/sj.emboj.7600235.

185. Wang, Q. & Goldstein, M. (2016). Small RNAs Recruit Chromatin-Modifying Enzymes MMSET and Tip60 to Reconfigure Damaged DNA upon Double-Strand Break and Facilitate Repair. Cancer Res. **76**, 1904-1915. https://doi.org/10.1158/0008-5472.CAN-15-2334.

186. Jeong, K. W., Kim, K., Situ, A. J., Ulmer, T. S., An, W. & Stallcup, M. R. (2011). Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. Nat Struct Mol Biol. **18**, 1358-1365. https://doi.org/10.1038/nsmb.2153.

187. Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q. & Ding, J. (2006). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res. **34**, 6621-6628. https://doi.org/10.1093/nar/gkl989.

188. Yuan, H., Rossetto, D., Mellert, H., Dang, W., Srinivasan, M., Johnson, J., Hodawadekar, S., Ding, E. C., Speicher, K., Abshiru, N., Perry, R., Wu, J., Yang, C., Zheng, Y. G., Speicher, D. W., Thibault, P., Verreault, A., Johnson, F. B., Berger, S. L., ... & Marmorstein, R. (2012). MYST protein acetyltransferase activity requires active site lysine autoacetylation. EMBO J. **31**, 58-70. https://doi.org/10.1038/emboj.2011.382.

189. Xu, P., Li, C., Chen, Z., Jiang, S., Fan, S., Wang, J., Dai, J., Zhu, P. & Chen, Z. (2016). The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism. Mol Cell. **63**, 965-975. https://doi.org/10.1016/j.molcel.2016.07.024.

190. Allard, S., Utley, R. T., Savard, J., Clarke, A., Grant, P., Brandl, C. J., Pillus, L., Workman, J. L. & Côté, J. (1999). NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. **18**, 5108-5119. https://doi.org/10.1093/emboj/18.18.5108.

191. Boudreault, A. A., Cronier, D., Selleck, W., Lacoste, N., Utley, R. T., Allard, S., Savard, J., Lane, W. S., Tan, S. & Côté, J. (2003). Yeast enhancer of polycomb defines global Esa1-dependent acetylation of chromatin. Genes Dev. **17**, 1415-1428. https://doi.org/10.1101/gad.1056603.

References

192. Doyon, Y., Cayrou, C., Ullah, M., Landry, A. J., Côté, V., Selleck, W., Lane, W. S., Tan, S., Yang, X. J. & Côté, J. (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. Mol Cell. **21**, 51-64. https://doi.org/10.1016/j.molcel.2005.12.007.

193. Kouzarides, T. (2007). Chromatin modifications and their function. Cell. **128**, 693-705. https://doi.org/10.1016/j.cell.2007.02.005.

194. Scacchetti, A., Schauer, T., Reim, A., Apostolou, Z., Campos Sparr, A., Krause, S., Heun, P., Wierer, M. & Becker, P. B. (2020). Drosophila SWR1 and NuA4 complexes are defined by DOMINO isoforms. Elife. **9**, e56325. https://doi.org/10.7554/eLife.56325.

195. Lu, P. Y. T., Kirlin, A. C., Aristizabal, M. J., Brewis, H. T., Levesque, N., Setiaputra, D. T., Avvakumov, N., Benschop, J. J., Groot Koerkamp, M., Holstege, F. C. P., Krogan, N. J., Yip, C. K., Côté, J. & Kobor, M. S. (2022). A balancing act: interactions within NuA4/TIP60 regulate picNuA4 function in Saccharomyces cerevisiae and humans. Genetics. **222**, iyac136. https://doi.org/10.1093/genetics/iyac136.

196. Wang, X., Ahmad, S., Zhang, Z., Côté, J. & Cai, G. (2018). Architecture of the Saccharomyces cerevisiae NuA4/TIP60 complex. Nat Commun. **9**, 1147. https://doi.org/10.1038/s41467-018-03504-5.

197. Doyon, Y., Selleck, W., Lane, W. S., Tan, S. & Côté, J. (2004). Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol. **24**, 1884-1896. https://doi.org/10.1128/MCB.24.5.1884-1896.2004.

198. Yan, Y., Harper, S., Speicher, D. W. & Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. Nat Struct Biol. **9**, 862-869. https://doi.org/10.1038/nsb849.

199. Kusch, T., Florens, L., Macdonald, W. H., Swanson, S. K., Glaser, R. L., Yates, J. R., Abmayr, S. M., Washburn, M. P. & Workman, J. L. (2004). Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. Science. **306**, 2084-2087. https://doi.org/10.1126/science.1103455.

200. Shvedunova, M. & Akhtar, A. (2022). Modulation of cellular processes by histone and non-histone protein acetylation. Nat Rev Mol Cell Biol. **23**, 329-349. https://doi.org/10.1038/s41580-021-00441-y.

201. Feller, C., Forné, I., Imhof, A. & Becker, P. B. (2015). Global and specific responses of the histone acetylome to systematic perturbation. Mol Cell. **57**, 559-571. https://doi.org/10.1016/j.molcel.2014.12.008.

202. Fuchs, S. M., Krajewski, K., Baker, R. W., Miller, V. L. & Strahl, B. D. (2011). Influence of combinatorial histone modifications on antibody and effector protein recognition. Curr Biol. **21**, 53-58. https://doi.org/10.1016/j.cub.2010.11.058.

203. Rothbart, S. B., Lin, S., Britton, L. M., Krajewski, K., Keogh, M. C., Garcia, B. A. & Strahl, B. D. (2012). Poly-acetylated chromatin signatures are preferred epitopes for site-specific histone H4 acetyl antibodies. Sci Rep. **2**, 489. https://doi.org/10.1038/srep00489.

204. Schirling, C., Heseding, C., Heise, F., Kesper, D., Klebes, A., Klein-Hitpass, L., Vortkamp, A., Hoffmann, D., Saumweber, H. & Ehrenhofer-Murray, A. E. (2010). Widespread regulation of gene expression in the Drosophila genome by the histone acetyltransferase dTip60. Chromosoma. **119**, 99-113. https://doi.org/10.1007/s00412-009-0247-z.

205. Kuo, Y. M., Henry, R. A. & Andrews, A. J. (2014). A quantitative multiplexed mass spectrometry assay for studying the kinetic of residue-specific histone acetylation. Methods. **70**, 127-133. https://doi.org/10.1016/j.ymeth.2014.08.003.

206. Kuo, Y. M., Henry, R. A., Tan, S., Côté, J. & Andrews, A. J. (2015). Site specificity analysis of Piccolo NuA4-mediated acetylation for different histone complexes. Biochem J. **472**, 239-248. https://doi.org/10.1042/BJ20150654.

References

207. Gilfillan, G. D., König, C., Dahlsveen, I. K., Prakoura, N., Straub, T., Lamm, R., Fauth, T. & Becker, P. B. (2007). Cumulative contributions of weak DNA determinants to targeting the Drosophila dosage compensation complex. Nucleic Acids Res. **35**, 3561-3572. https://doi.org/10.1093/nar/gkm282.

208. Straub, T., Neumann, M. F., Prestel, M., Kremmer, E., Kaether, C., Haass, C. & Becker, P. B. (2005). Stable chromosomal association of MSL2 defines a dosage-compensated nuclear compartment. Chromosoma. **114**, 352-364. https://doi.org/10.1007/s00412-005-0020-x.

209. Prestel, M., Feller, C., Straub, T., Mitlöhner, H. & Becker, P. B. (2010). The activation potential of MOF is constrained for dosage compensation. Mol Cell. **38**, 815-826. https://doi.org/10.1016/j.molcel.2010.05.022.

210. Weissmann, F., Petzold, G., VanderLinden, R., Huis In 't Veld, P. J., Brown, N. G., Lampert, F., Westermann, S., Stark, H., Schulman, B. A. & Peters, J. M. (2016). biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. Proc Natl Acad Sci U S A. **113**, E2564-2569. https://doi.org/10.1073/pnas.1604935113.

211. Klinker, H., Haas, C., Harrer, N., Becker, P. B. & Müller-Planitz, F. (2014). Rapid purification of recombinant histones. PLoS One. **9**, e104029. https://doi.org/10.1371/journal.pone.0104029.

212. Grimm, M., Zimniak, T., Kahraman, A. & Herzog, F. (2015). xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. Nucleic Acids Res. **43**, W362-369. https://doi.org/10.1093/nar/gkv463.

213. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A. & Vizcaíno, J. A. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. **50**, D543-D552. https://doi.org/10.1093/nar/gkab1038.

214. Rappsilber, J., Mann, M. & Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. **2**, 1896-1906. https://doi.org/10.1038/nprot.2007.261.

215. Pino, L. K., Searle, B. C., Bollinger, J. G., Nunn, B., MacLean, B. & MacCoss, M. J. (2020). The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. Mass Spectrom Rev. **39**, 229-244. https://doi.org/10.1002/mas.21540.

216. Kiss, A. E., Venkatasubramani, A. V., Pathirana, D., Krause, S., Campos Sparr, A., Hasenauer, J., Imhof, A., Müller, M. & Becker, P. B. (2024). Processivity and specificity of histone acetylation by the male-specific lethal complex. Nucleic Acids Res. gkae123. https://doi.org/10.1093/nar/gkae123.

217. Krietenstein, N., Wippo, C. J., Lieleg, C. & Korber, P. (2012). Genome-wide in vitro reconstitution of yeast chromatin with in vivo-like nucleosome positioning. Methods Enzymol. **513**, 205-232. https://doi.org/10.1016/B978-0-12-391938-0.00009-4.

218. Wippo, C. J. & Korber, P. (2012). In vitro reconstitution of in vivo-like nucleosome positioning on yeast DNA. Methods Mol Biol. **833**, 271-287. https://doi.org/10.1007/978-1-61779-477-3_17.

219. Luger, K., Rechsteiner, T. J. & Richmond, T. J. (1999). Preparation of nucleosome core particle from recombinant histones. Methods Enzymol. **304**, 3-19. https://doi.org/10.1016/s0076-6879(99)04003-3.

220. Klinker, H., Müller-Planitz, F., Yang, R., Forné, I., Liu, C. F., Nordenskiold, L. & Becker, P. B. (2014). ISWI remodelling of physiological chromatin fibres acetylated at lysine 16 of histone H4. PLoS One. **9**, e88411. https://doi.org/10.1371/journal.pone.0088411.

221. Routh, A., Sandin, S. & Rhodes, D. (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc Natl Acad Sci U S A. **105**, 8872-8877. https://doi.org/10.1073/pnas.0802336105.

References

222. Lowary, P. T. & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol. **276**, 19-42. https://doi.org/10.1006/jmbi.1997.1494.

223. Ludwigsen, J., Pfennig, S., Singh, A. K., Schindler, C., Harrer, N., Forné, I., Zacharias, M. & Müller-Planitz, F. (2017). Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. Elife. **6**, e21477. https://doi.org/10.7554/eLife.21477.

224. Rappsilber, J. & Mann, M. (2007). Analysis of the topology of protein complexes using cross-linking and mass spectrometry. CSH Protoc. **2007**, pdb.prot4594. https://doi.org/10.1101/pdb.prot4594.

225. Iacobucci, C. & Sinz, A. (2017). To Be or Not to Be? Five Guidelines to Avoid Misassignments in Cross-Linking/Mass Spectrometry. Anal Chem. **89**, 7832-7835. https://doi.org/10.1021/acs.analchem.7b02316.

226. Combe, C. W., Fischer, L. & Rappsilber, J. (2015). xiNET: cross-link network maps with residue resolution. Mol Cell Proteomics. **14**, 1137-1147. https://doi.org/10.1074/mcp.O114.042259.

227. Hagemann, G., Solis-Mezarino, V., Singh, S., Potocnjak, M., Kumar, C. & Herzog, F. (2022). Quantitative crosslinking and mass spectrometry determine binding interfaces and affinities mediating kinetochore stabilization. bioRxiv. 2022.2003.2031.486303. https://doi.org/10.1101/2022.03.31.486303.

228. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature. **596**, 583-589. https://doi.org/10.1038/s41586-021-03819-2.

229. Kosinski, J., von Appen, A., Ori, A., Karius, K., Müller, C. W. & Beck, M. (2015). Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. J Struct Biol. **189**, 177-183. https://doi.org/10.1016/j.jsb.2015.01.014.

230. Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H. & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. **27**, 14-25. https://doi.org/10.1002/pro.3235.

231. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H. & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. **30**, 70-82. https://doi.org/10.1002/pro.3943.

232. Morra, R., Yokoyama, R., Ling, H. & Lucchesi, J. C. (2011). Role of the ATPase/helicase maleless (MLE) in the assembly, targeting, spreading and function of the male-specific lethal (MSL) complex of Drosophila. Epigenetics Chromatin. **4**, 6. https://doi.org/10.1186/1756-8935-4-6.

233. He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D. C., Wilusz, J. E., Garcia, B. A. & Bonasio, R. (2016). High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. Mol Cell. **64**, 416-430. https://doi.org/10.1016/j.molcel.2016.09.034.

234. Consortium, U. (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. **51**, D523-D531. https://doi.org/10.1093/nar/gkac1052.

235. Oberbeckmann, E., Niebauer, V., Watanabe, S., Farnung, L., Moldt, M., Schmid, A., Cramer, P., Peterson, C. L., Eustermann, S., Hopfner, K. P. & Korber, P. (2021). Ruler elements in chromatin remodelers set nucleosome array spacing and phasing. Nat Commun. **12**, 3232. https://doi.org/10.1038/s41467-021-23015-0.

236. Strahl, B. D. & Allis, C. D. (2000). The language of covalent histone modifications. Nature. **403**, 41-45. https://doi.org/10.1038/47412.

237. Viita, T. & Côté, J. (2022). The MOZ-BRPF1 acetyltransferase complex in epigenetic crosstalk linked to gene regulation, development, and human diseases. Front Cell Dev Biol. **10**, 1115903. https://doi.org/10.3389/fcell.2022.1115903.

References

238. Beaver, M., Karisetty, B. C., Zhang, H., Bhatnagar, A., Armour, E., Parmar, V., Brown, R., Xiang, M. & Elefant, F. (2022). Chromatin and transcriptomic profiling uncover dysregulation of the Tip60 HAT/HDAC2 epigenomic landscape in the neurodegenerative brain. Epigenetics. **17**, 786-807. https://doi.org/10.1080/15592294.2021.1959742.

239. Turner, B. M. (2000). Histone acetylation and an epigenetic code. Bioessays. **22**, 836-845. https://doi.org/10.1002/1521-1878(200009)22:9<836::AID-BIES9>3.0.CO;2-X.

240. Verdin, E. & Ott, M. (2015). 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. Nat Rev Mol Cell Biol. **16**, 258-264. https://doi.org/10.1038/nrm3931.

241. Bannister, A. J. & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. **21**, 381-395. https://doi.org/10.1038/cr.2011.22.

242. Ahmad, K., Henikoff, S. & Ramachandran, S. (2022). Managing the Steady State Chromatin Landscape by Nucleosome Dynamics. Annu Rev Biochem. **91**, 183-195. https://doi.org/10.1146/annurev-biochem-032620-104508.

243. Blasi, T., Feller, C., Feigelman, J., Hasenauer, J., Imhof, A., Theis, F. J., Becker, P. B. & Marr, C. (2016). Combinatorial Histone Acetylation Patterns Are Generated by Motif-Specific Reactions. Cell Syst. **2**, 49-58. https://doi.org/10.1016/j.cels.2016.01.002.

244. Devoucoux, M., Roques, C., Lachance, C., Lashgari, A., Joly-Beauparlant, C., Jacquet, K., Alerasool, N., Prudente, A., Taipale, M., Droit, A., Lambert, J. P., Hussein, S. M. I. & Côté, J. (2022). MRG Proteins Are Shared by Multiple Protein Complexes With Distinct Functions. Mol Cell Proteomics. **21**, 100253. https://doi.org/10.1016/j.mcpro.2022.100253.

245. Procida, T., Friedrich, T., Jack, A. P. M., Peritore, M., Bönisch, C., Eberl, H. C., Daus, N., Kletenkov, K., Nist, A., Stiewe, T., Borggrefe, T., Mann, M., Bartkuhn, M. & Hake, S. B. (2021). JAZF1, A Novel p400/TIP60/NuA4 Complex Member, Regulates H2A.Z Acetylation at Regulatory Regions. Int J Mol Sci. **22**. https://doi.org/10.3390/ijms22020678.

246. Berndsen, C. E., Selleck, W., McBryant, S. J., Hansen, J. C., Tan, S. & Denu, J. M. (2007). Nucleosome recognition by the Piccolo NuA4 histone acetyltransferase complex. Biochemistry. **46**, 2091-2099. https://doi.org/10.1021/bi602366n.

247. Zhang, K., Williams, K. E., Huang, L., Yau, P., Siino, J. S., Bradbury, E. M., Jones, P. R., Minch, M. J. & Burlingame, A. L. (2002). Histone acetylation and deacetylation: identification of acetylation and methylation sites of HeLa histone H4 by mass spectrometry. Mol Cell Proteomics. **1**, 500-508. https://doi.org/10.1074/mcp.m200031-mcp200.

248. Figueiredo, M. L., Kim, M., Philip, P., Allgardsson, A., Stenberg, P. & Larsson, J. (2014). Non-coding roX RNAs prevent the binding of the MSL-complex to heterochromatic regions. PLoS Genet. **10**, e1004865. https://doi.org/10.1371/journal.pgen.1004865.

249. Becker, P. B. & Wu, C. (1992). Cell-free system for assembly of transcriptionally repressed chromatin from Drosophila embryos. Mol Cell Biol. **12**, 2241-2249. https://doi.org/10.1128/mcb.12.5.2241-2249.1992.

250. Völker-Albert, M. C., Pusch, M. C., Fedisch, A., Schilcher, P., Schmidt, A. & Imhof, A. (2016). A Quantitative Proteomic Analysis of In Vitro Assembled Chromatin. Mol Cell Proteomics. **15**, 945-959. https://doi.org/10.1074/mcp.M115.053553.

251. Scharf, A. N., Meier, K., Seitz, V., Kremmer, E., Brehm, A. & Imhof, A. (2009). Monomethylation of lysine 20 on histone H4 facilitates chromatin maturation. Mol Cell Biol. **29**, 57-67. https://doi.org/10.1128/MCB.00989-08.

252. Larschan, E., Soruco, M. M., Lee, O. K., Peng, S., Bishop, E., Chery, J., Goebel, K., Feng, J., Park, P. J. & Kuroda, M. I. (2012). Identification of chromatin-associated regulators of MSL complex targeting in Drosophila dosage compensation. PLoS Genet. **8**, e1002830. https://doi.org/10.1371/journal.pgen.1002830.

References

253. Larschan, E., Bishop, E. P., Kharchenko, P. V., Core, L. J., Lis, J. T., Park, P. J. & Kuroda, M. I. (2011). X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature. **471**, 115-118. https://doi.org/10.1038/nature09757.

254. Gelbart, M. E., Larschan, E., Peng, S., Park, P. J. & Kuroda, M. I. (2009). Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat Struct Mol Biol. **16**, 825-832. https://doi.org/10.1038/nsmb.1644.

255. Arnold, K. M., Lee, S. & Denu, J. M. (2011). Processing mechanism and substrate selectivity of the core NuA4 histone acetyltransferase complex. Biochemistry. **50**, 727-737. https://doi.org/10.1021/bi101355a.

256. Fréchard, A., Faux, C., Hexnerova, R., Crucifix, C., Papai, G., Smirnova, E., McKeon, C., Ping, F. L. Y., Helmlinger, D., Schultz, P. & Ben-Shem, A. (2023). The structure of the NuA4-Tip60 complex reveals the mechanism and importance of long-range chromatin modification. Nat Struct Mol Biol. **30**, 1337-1345. https://doi.org/10.1038/s41594-023-01056-x.

257. Zhang, H., Devoucoux, M., Song, X., Li, L., Ayaz, G., Cheng, H., Tempel, W., Dong, C., Loppnau, P., Côté, J. & Min, J. (2020). Structural Basis for EPC1-Mediated Recruitment of MBTD1 into the NuA4/TIP60 Acetyltransferase Complex. Cell Rep. **30**, 3996-4002 e3994. https://doi.org/10.1016/j.celrep.2020.03.003.

258. Prozzillo, Y., Cuticone, S., Ferreri, D., Fattorini, G., Messina, G. & Dimitri, P. (2021). In Vivo Silencing of Genes Coding for dTip60 Chromatin Remodeling Complex Subunits Affects Polytene Chromosome Organization and Proper Development in Drosophila melanogaster. Int J Mol Sci. **22**, 4525. https://doi.org/10.3390/ijms22094525.

259. Prozzillo, Y., Fattorini, G., Ferreri, D., Leo, M., Dimitri, P. & Messina, G. (2023). Knockdown of DOM/Tip60 Complex Subunits Impairs Male Meiosis of Drosophila melanogaster. Cells. **12**, 1348. https://doi.org/10.3390/cells12101348.

260. Auger, A., Galarneau, L., Altaf, M., Nourani, A., Doyon, Y., Utley, R. T., Cronier, D., Allard, S. & Côté, J. (2008). Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants. Mol Cell Biol. **28**, 2257-2270. https://doi.org/10.1128/MCB.01755-07.

261. Bhatnagar, A., Krick, K., Karisetty, B. C., Armour, E. M., Heller, E. A. & Elefant, F. (2023). Tip60's Novel RNA-Binding Function Modulates Alternative Splicing of Pre-mRNA Targets Implicated in Alzheimer's Disease. J Neurosci. **43**, 2398-2423. https://doi.org/10.1523/JNEUROSCI.2331-22.2023.

262. Cai, Y., Jin, J., Tomomori-Sato, C., Sato, S., Sorokina, I., Parmely, T. J., Conaway, R. C. & Conaway, J. W. (2003). Identification of new subunits of the multiprotein mammalian TRRAP/TIP60-containing histone acetyltransferase complex. J Biol Chem. **278**, 42733-42736. https://doi.org/10.1074/jbc.C300389200.

263. Squatrito, M., Gorrini, C. & Amati, B. (2006). Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends Cell Biol. **16**, 433-442. https://doi.org/10.1016/j.tcb.2006.07.007.

264. Lashgari, A., Lambert, J. P. & Côté, J. (2019). Measurement and Analysis of Lysine Acetylation by KAT Complexes In Vitro and In Vivo. Methods Mol Biol. **1983**, 57-77. https://doi.org/10.1007/978-1-4939-9434-2_5.

10 Table of Abbreviations

Abbreviation	Explanation
4sU	4-thio-uridine
Å	Ångström
aa	Amino acid
ab	Antibody
ac	acetylation
ACN	Acetonitrile
ADP	Adenosinediphosphate
AF	AlphaFold
AI	Artificial intelligence
Amp	Ampicillin
ATP	Adenosinetriphosphate
AU/a.u.	Arbitrary units
B	Beads
bp	Base pair(s)
BLAST	Basic logical alignment search tool
BS3	Bis(sulfosuccinimidyl)suberate, crosslinker
BSA	Bovine serum albumin
c	Concentration
CC	Coiled coil
CF	Cytoplasmic fraction
CHART	Capture hybridization analysis of RNA targets (method)
ChIP	Chromatin immunoprecipitation (method)
ChIRP	Chromatin isolation by RNA purification (method)
Chl	Chloramphenicol
CLAMP	Chromatin-linked adapter for MSL proteins
Ctrl	Control reaction
DCC	Dosage compensation complex
ddH$_2$O	Doubly distilled water
DDR	DNA damage response
DIP	DNA immunoprecipitation
DNA	Deoxyribonucleic acid
DREX	*Drosophila* embryo extract (preblastoderm)
DROME	*Drosophila melanogaster*
Dm	*Drosophila melanogaster*
ds	Double-stranded
Eaf6	Esa1-associated factor 6
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylenglycol-bis(2-aminoethyl)-N,N,N',N'-tetraacetic acid
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay (method)
EPc	Enhancer of polycomb
FDR	False discovery rate, 5% in this dissertation (unless otherwise stated)
FLAG	FLAG® epitope tag; short, hydrophilic, eight-amino acid peptide (DYKDDDDK tag)
FRAP	Fluorescence recovery after photobleaching (method)
FRET	Förster resonance energy transfer (method)
FT	Flow through
GFP	Green fluorescent protein
gp	Guinnea pig
IDR	Intrinsically disordered region
IN	Input
ING3	Inhibitor of growth protein 3
IP	Immunoprecipitation
IPTG	Isopropyl-D-thiogalactopyranoside

Table of Abbreviations

ivt	*In vitro* transcribed
H3K36me3	Histone 3 lysine 36 trimethylation
H4K16ac	Histone 4 lysine 16 acetylation
HAS	High affinity site(s)
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEPES	N-(2-hydroxyethyl)piperazine-H´-2- ethanesulfonic acid
hnRNA	heterogenous nuclear RNA
kb	Kilobases
Kc	*Drosophila melanogaster* cells of female karotype
kDa	Kilo-Dalton (mass unit for biological molecules)
lnc	Long non-coding
M	Marker
MD	Molecular dynamics
me	methylation
MLE	Maleless
MNase	Micrococcal nuclease
mo	mouse
MOF	Males-absent-on-the-first
MS	Mass spectrometry
MSL	Male-specific lethal
MRE	MSL recognition element
MRG	MORF4-related gene family (MRG) domain
mRNA	Messenger RNA
MW	Molecular weight
napts	no action potential, temperature-sensitive, an MLE mutation
NCBI	National Center for Biotechnology Information
NE	Nuclear extract
NMR	Nuclear magnetic resonance
nt	Nucleotide(s)
NuA4	Nucleosome acetyltransferase of H4
OB-fold	Oligosaccharide-binding fold, domain of MLE helicase
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis (method)
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data base
PionX	Pioneering-sites-on-the-X
Poly-dA	Poly-deoxy-adenylic acid
PTM	Post-translational modification
R	Rest (in a chemical structure)
rb	rabbit
RB	RNA binding domain
Ref	Reference number for MS samples
RHA	RNA helicase A
RING	Really Interesting New Gene, protein domain, zinc finger
RIP	RNA immunoprecipitation
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
roX	RNA on the X (chromosome)
rt	rat
RT	Room temperature
S2	Schneider cells, *Drosophila melanogaster* cells of male karyotype
SDS	Sodium dodecyl sulfate
SEM	Standard Error of the Mean
Seq	sequencing
Sf	*Spodoptera frugiperda*
SGD	Salt gradient dialysis
SHAPE	Selective 2′-hydroxyl acylation analyzed by primer extension (method)

Table of Abbreviations

SL	Stem-loop
ss	Single-stranded
STAGE	stop-and-go-extraction (method)
TF	Transcription factor
TFA	Trifluoroacetic acid
Tip60	Tat-interactive protein 60-kDa
UNR	Upstream of N-ras
UTR	Untranslated region of an mRNA
UV	Ultraviolet
vitRIP	*In vitro* RNA immunoprecipitation (method)
V/V	Volume per volume
WB	Western blot
WT	Wild-type
w/V	Weight per volume
XCI	X chromosome inactivation
Xist	X-inactive specific transcript
XL	Crosslinking
ZGA	Zygotic genome activation

11 Acknowledgement

First and foremost, I would like to thank Prof. Dr. Peter B. Becker my supervisor and "Doktorvater". He accepted me to pursue this PhD project in his research group and supported me with his ever-lasting patience, his long-standing experience in the field and strategic words of guidance.

Secondly, but not less importantly, I thank Dr. Marisa Müller. In many ways I was "her" PhD student more than Peter's. I learned countless important lessons from her. She contributed several missing puzzle pieces of data to the project and the paper. She spent a lot of time to improve my writing and graphical design skills for the better.

Thirdly, and in my opinion, they deserve to be named up top, I want to thank the technicians of the lab, Silke Krause and Aline Campos Sparr. Not only they contributed as co-authors to my publication, they also made everyday lab life pleasant and less cumbersome than other labs, who do not thrive from their constant support and advice on technicalities. Special thanks to Aline Campos-Sparr, who helped with her golden hands with SGD chromatin assemblies, MNase assays and HAT assays. I am very proud that you are now an expert in these techniques. Thanks for teaching me about Brazil and Brazilian culture, even though I have never managed to go there during my PhD time.

I acknowledge the Protein Analytics Unit at the Biomedical Center, Ludwig-Maximilians University Munich, for providing equipment, materials, services, expertise and assistance with data analysis. First and foremost, Dr. Ignasi Forné, who went out of his way to help with any kind of problem.

Thanks as well to Anuroop Venkateswaran Venkatasubramani, who quantified and performed the primary analysis of the acetylation data. He also taught me the art of histone sample preparation for MS, which I then mastered and optimized on my own.

My gratitude goes to Prof. Dr. Axel Imhof, important collaborator on all my projects and member of my TAC. Additionally, thanks to Prof. Dr. Elena Conti, TAC member, inspiring woman in science and invaluable input on my experimental designs.

Dr. Sebastian Eustermann for AF modelling and collaboration on the structural XL-MS project. May my data be useful for the future project. I think we're both sorry, we never managed to reach publication stage – well, I am.

My collaborators in Bonn Dr. Dilan Pathirana and Prof. Dr. Jan Hasenauer for their great input on the mathematical modelling of the acetylation pathway.

Thanks to my "bench buddy" Dr. Nikolas Eggers, who most importantly taught me, that even in the busiest and most stressful times, to remember that there is still time. Vera Kleene, who was a great "PhD twin", perfect partner in crime to plan and organize events, journal clubs and international dinners. I am thankful for the help and advice from Dr. Petra Vizjak, who taught me the art of histone purification and octamer reconstitution. Dr. Elisa Oberbeckmann, who taught me the skills of SGD. I thank the coffee team for keeping me well caffeinated, Janet, Jay, Gizem, Fotios, Viola, Lara, Beyza, Namisha, Chondamma, Lorenz and Drin.

Thanks to the IRTG1064 (Elizabeth Schröder-Reiter) and the QBM (Markus Hohle) for fun activities, workshops and retreats.

I acknowledge Dr. Sophia Reindl, who initiated the idea of doing a PhD in Munich and provided the contact. I think I would have neither applied nor been accepted without her help.

Thanks to my husband Christoph Müller-Hermes, without whom I would have quit after few months in or also at later stages of the PhD.

12 Appendix

12.1 Python script to convert crosslinking data into .csv format for the visualization CrossFinderToXvis.py

```python
import re
import sys

if __name__ == "__main__":
    d = open(sys.argv[1], "r")
    lines = "".join(d.readlines()).replace("\n","")
    #lines.replace(" \t","\t")
    lines = lines.split("\r")

    print "Protein1,Protein2,AbsPos1,AbsPos2"

    if len(lines) > 2:
        header = re.sub("\s*\t", "\t", lines[1].strip()).split("\t") #lines[1].strip().replace(" \t", "\t").split("\t")
        p1 = header.index("protein1")
        p2 = header.index("protein2")
        a1 = header.index("XLpos_abs1")
        a2 = header.index("XLpos_abs2")

        for i in range(2, len(lines)):
            content = re.sub("\s*\t", "\t", lines[i].strip()).split("\t")
            if len(content) >= max(p1,p2,a1,a2):
                pl1 = content[p1].split(", ")
                pl2 = content[p2].split(", ")
                al1 = content[a1].split(", ")
                al2 = content[a2].split(", ")

                if len(pl1) == len(al1) and len(pl2) == len(al2):
                    for j in range(len(pl1)):
                        for k in range(len(pl2)):
                            print pl1[j] + "," + pl2[k] + "," + al1[j] + "," + al2[k]
                #print content[p1] + "," + content[p2] + "," + content[a1] + "," + content[a2]

    d.close()
```

Appendix

12.2 Mass spectrometry XL-MS data

12.2.1 Ref1844 (2-MSL)

Table 1: MSL1ΔC and MSL2 were crosslinked with BS3 (0.5, 1.0 and 1.5 mM). Summary of detected crosslink positions with FDR <5%. Duplicate values have been removed to eliminate redundancy, resulting in a cleaner overview. The values are ordered by 'Protein 1' position of the crosslink. Data of Ref. 1844 shown.

Protein1	Protein2	Protein1	Protein2	Protein1	Protein2
MSL1 071	MSL2 436	MSL1 592	MSL1 579	MSL1 703	MSL1 823
MSL1 102	MSL1 558	MSL1 592	MSL1 582	MSL1 704	MSL1 628
MSL1 102	MSL1 562	MSL1 592	MSL1 621	MSL1 704	MSL1 663
MSL1 102	MSL1 605	MSL1 592	MSL1 628	MSL1 704	MSL1 666
MSL1 102	MSL2 436	MSL1 592	MSL1 652	MSL1 704	MSL1 697
MSL1 141	MSL1 146	MSL1 592	MSL1 666	MSL1 704	MSL1 719
MSL1 146	MSL1 582	MSL1 599	MSL1 484	MSL1 704	MSL1 776
MSL1 146	MSL1 628	MSL1 599	MSL1 485	MSL1 718	MSL1 628
MSL1 153	MSL1 141	MSL1 599	MSL1 574	MSL1 718	MSL1 652
MSL1 236	MSL1 146	MSL1 599	MSL1 579	MSL1 718	MSL1 697
MSL1 268	MSL1 260	MSL1 599	MSL1 582	MSL1 718	MSL1 703
MSL1 269	MSL1 260	MSL1 599	MSL1 614	MSL1 718	MSL1 704
MSL1 449	MSL1 558	MSL1 599	MSL1 628	MSL1 718	MSL1 719
MSL1 459	MSL1 449	MSL1 605	MSL1 146	MSL1 719	MSL1 697
MSL1 459	MSL1 484	MSL1 605	MSL1 592	MSL1 719	MSL1 703
MSL1 459	MSL1 485	MSL1 605	MSL1 621	MSL1 719	MSL1 704
MSL1 459	MSL1 582	MSL1 605	MSL1 628	MSL1 778	MSL2 473
MSL1 464	MSL1 449	MSL1 614	MSL1 582	MSL2 088	MSL2 055
MSL1 464	MSL1 484	MSL1 621	MSL1 628	MSL2 088	MSL2 057
MSL1 464	MSL1 485	MSL1 621	MSL1 650	MSL2 135	MSL1 269
MSL1 464	MSL1 579	MSL1 621	MSL1 652	MSL2 135	MSL2 72
MSL1 479	MSL1 488	MSL1 621	MSL1 666	MSL2 442	MSL2 453
MSL1 484	MSL1 133	MSL1 628	MSL1 652	MSL2 442	MSL2 481
MSL1 484	MSL1 488	MSL1 628	MSL1 666	MSL2 442	MSL2 496
MSL1 484	MSL2 514	MSL1 652	MSL1 146	MSL2 449	MSL2 481
MSL1 485	MSL1 131	MSL1 652	MSL1 268	MSL2 453	MSL2 473
MSL1 485	MSL1 133	MSL1 652	MSL1 574	MSL2 453	MSL2 481
MSL1 485	MSL1 486	MSL1 652	MSL1 628	MSL2 464	MSL2 473
MSL1 485	MSL1 488	MSL1 652	MSL1 666	MSL2 464	MSL2 481
MSL1 485	MSL2 512	MSL1 652	MSL1 668	MSL2 473	MSL2 451
MSL1 485	MSL2 514	MSL1 652	MSL1 675	MSL2 473	MSL2 481
MSL1 488	MSL1 260	MSL1 652	MSL1 719	MSL2 481	MSL2 498
MSL1 488	MSL1 484	MSL1 663	MSL1 628	MSL2 487	MSL2 442
MSL1 488	MSL1 485	MSL1 663	MSL1 668	MSL2 487	MSL2 453

MSL1 528	MSL1 534	MSL1 666	MSL1 776	MSL2 487	MSL2 473
MSL1 530	MSL1 534	MSL1 675	MSL1 628	MSL2 487	MSL2 498
MSL1 558	MSL1 582	MSL1 675	MSL1 652	MSL2 487	MSL2 503
MSL1 558	MSL1 628	MSL1 675	MSL1 666	MSL2 487	MSL2 716
MSL1 562	MSL1 555	MSL1 675	MSL1 697	MSL2 496	MSL2 481
MSL1 574	MSL1 555	MSL1 676	MSL1 663	MSL2 496	MSL2 503
MSL1 574	MSL1 558	MSL1 676	MSL1 666	MSL2 496	MSL2 514
MSL1 574	MSL1 562	MSL1 676	MSL1 668	MSL2 496	MSL2 716
MSL1 574	MSL1 582	MSL1 676	MSL1 697	MSL2 498	MSL2 481
MSL1 579	MSL1 558	MSL1 697	MSL1 628	MSL2 498	MSL2 514
MSL1 592	MSL1 558	MSL1 703	MSL1 628	MSL2 514	MSL2 512
MSL1 592	MSL1 562	MSL1 703	MSL1 719	MSL2 519	MSL2 517
MSL1 592	MSL1 574				

12.2.2 Ref2201 (MSL1 alone)

Table 2: MSL1 was crosslinked with BS3 (0.5, 0.75 and 1.0 mM). Summary of detected crosslink positions with FDR <5%. Duplicate values have been removed to eliminate redundancy, resulting in a cleaner overview. Values are ordered by the 'Protein 1' column. Data of Ref. 2201 shown.

Protein1	Protein2	Protein1	Protein2	Protein1	Protein2
MSL1 0048	MSL1 0133	MSL1 0574	MSL1 0675	MSL1 0697	MSL1 0703
MSL1 0048	MSL1 0882	MSL1 0574	MSL1 0676	MSL1 0697	MSL1 0704
MSL1 0048	MSL1 0916	MSL1 0574	MSL1 0697	MSL1 0697	MSL1 0823
MSL1 0071	MSL1 0006	MSL1 0574	MSL1 0719	MSL1 0703	MSL1 0173
MSL1 0071	MSL1 0146	MSL1 0574	MSL1 0776	MSL1 0703	MSL1 0582
MSL1 0071	MSL1 0652	MSL1 0574	MSL1 0778	MSL1 0703	MSL1 0628
MSL1 0071	MSL1 0997	MSL1 0574	MSL1 0916	MSL1 0703	MSL1 0652
MSL1 0102	MSL1 0501	MSL1 0579	MSL1 0449	MSL1 0703	MSL1 0666
MSL1 0102	MSL1 0882	MSL1 0579	MSL1 0534	MSL1 0703	MSL1 0668
MSL1 0102	MSL1 1010	MSL1 0579	MSL1 0558	MSL1 0703	MSL1 0719
MSL1 0131	MSL1 0133	MSL1 0579	MSL1 0628	MSL1 0703	MSL1 0823
MSL1 0133	MSL1 0628	MSL1 0579	MSL1 0652	MSL1 0703	MSL1 0908
MSL1 0133	MSL1 0652	MSL1 0579	MSL1 0719	MSL1 0703	MSL1 0916
MSL1 0133	MSL1 0719	MSL1 0582	MSL1 0131	MSL1 0704	MSL1 0146
MSL1 0133	MSL1 0908	MSL1 0582	MSL1 0486	MSL1 0704	MSL1 0173
MSL1 0141	MSL1 0146	MSL1 0582	MSL1 0555	MSL1 0704	MSL1 0592
MSL1 0141	MSL1 0916	MSL1 0582	MSL1 0562	MSL1 0704	MSL1 0621
MSL1 0146	MSL1 0006	MSL1 0582	MSL1 0579	MSL1 0704	MSL1 0628
MSL1 0146	MSL1 0133	MSL1 0582	MSL1 0650	MSL1 0704	MSL1 0650
MSL1 0146	MSL1 0141	MSL1 0582	MSL1 0652	MSL1 0704	MSL1 0652
MSL1 0146	MSL1 0173	MSL1 0592	MSL1 0146	MSL1 0704	MSL1 0663
MSL1 0146	MSL1 0260	MSL1 0592	MSL1 0269	MSL1 0704	MSL1 0666
MSL1 0146	MSL1 0269	MSL1 0592	MSL1 0555	MSL1 0704	MSL1 0668

Appendix

MSL1 0146	MSL1 0449	MSL1 0592	MSL1 0558	MSL1 0704	MSL1 0675
MSL1 0146	MSL1 0574	MSL1 0592	MSL1 0570	MSL1 0704	MSL1 0676
MSL1 0146	MSL1 0582	MSL1 0592	MSL1 0574	MSL1 0704	MSL1 0697
MSL1 0146	MSL1 0614	MSL1 0592	MSL1 0579	MSL1 0704	MSL1 0703
MSL1 0146	MSL1 0621	MSL1 0592	MSL1 0582	MSL1 0704	MSL1 0719
MSL1 0146	MSL1 0628	MSL1 0592	MSL1 0614	MSL1 0704	MSL1 0908
MSL1 0146	MSL1 0666	MSL1 0592	MSL1 0621	MSL1 0704	MSL1 0916
MSL1 0146	MSL1 0675	MSL1 0592	MSL1 0628	MSL1 0704	MSL1 0997
MSL1 0146	MSL1 0676	MSL1 0592	MSL1 0650	MSL1 0718	MSL1 0146
MSL1 0146	MSL1 0719	MSL1 0592	MSL1 0652	MSL1 0718	MSL1 0449
MSL1 0146	MSL1 0823	MSL1 0592	MSL1 0666	MSL1 0718	MSL1 0592
MSL1 0146	MSL1 0899	MSL1 0592	MSL1 0719	MSL1 0718	MSL1 0614
MSL1 0146	MSL1 0908	MSL1 0592	MSL1 0823	MSL1 0718	MSL1 0621
MSL1 0146	MSL1 0916	MSL1 0599	MSL1 0146	MSL1 0718	MSL1 0628
MSL1 0146	MSL1 1010	MSL1 0599	MSL1 0555	MSL1 0718	MSL1 0652
MSL1 0153	MSL1 0141	MSL1 0599	MSL1 0558	MSL1 0718	MSL1 0663
MSL1 0153	MSL1 0269	MSL1 0599	MSL1 0574	MSL1 0718	MSL1 0666
MSL1 0168	MSL1 0173	MSL1 0599	MSL1 0579	MSL1 0718	MSL1 0668
MSL1 0168	MSL1 0260	MSL1 0599	MSL1 0582	MSL1 0718	MSL1 0675
MSL1 0168	MSL1 0628	MSL1 0599	MSL1 0614	MSL1 0718	MSL1 0676
MSL1 0168	MSL1 0719	MSL1 0599	MSL1 0621	MSL1 0718	MSL1 0697
MSL1 0197	MSL1 0131	MSL1 0599	MSL1 0628	MSL1 0718	MSL1 0703
MSL1 0236	MSL1 0146	MSL1 0599	MSL1 0666	MSL1 0718	MSL1 0704
MSL1 0236	MSL1 0168	MSL1 0599	MSL1 0668	MSL1 0718	MSL1 0719
MSL1 0236	MSL1 0260	MSL1 0605	MSL1 0146	MSL1 0718	MSL1 0776
MSL1 0236	MSL1 0269	MSL1 0605	MSL1 0562	MSL1 0718	MSL1 0778
MSL1 0236	MSL1 0628	MSL1 0605	MSL1 0574	MSL1 0718	MSL1 0804
MSL1 0255	MSL1 0146	MSL1 0605	MSL1 0579	MSL1 0718	MSL1 0823
MSL1 0255	MSL1 0168	MSL1 0605	MSL1 0582	MSL1 0718	MSL1 0902
MSL1 0255	MSL1 0260	MSL1 0605	MSL1 0592	MSL1 0718	MSL1 0908
MSL1 0255	MSL1 0268	MSL1 0605	MSL1 0599	MSL1 0718	MSL1 0916
MSL1 0255	MSL1 0269	MSL1 0605	MSL1 0614	MSL1 0718	MSL1 0997
MSL1 0255	MSL1 0614	MSL1 0605	MSL1 0621	MSL1 0718	MSL1 1009
MSL1 0255	MSL1 0628	MSL1 0605	MSL1 0628	MSL1 0718	MSL1 1010
MSL1 0255	MSL1 0997	MSL1 0605	MSL1 0650	MSL1 0718	MSL1 1031
MSL1 0260	MSL1 0579	MSL1 0605	MSL1 0652	MSL1 0719	MSL1 0146
MSL1 0260	MSL1 0666	MSL1 0605	MSL1 0663	MSL1 0719	MSL1 0449
MSL1 0268	MSL1 0173	MSL1 0605	MSL1 0666	MSL1 0719	MSL1 0592
MSL1 0268	MSL1 0260	MSL1 0605	MSL1 0704	MSL1 0719	MSL1 0614
MSL1 0268	MSL1 0908	MSL1 0605	MSL1 0719	MSL1 0719	MSL1 0621
MSL1 0269	MSL1 0260	MSL1 0605	MSL1 1009	MSL1 0719	MSL1 0652
MSL1 0269	MSL1 0268	MSL1 0605	MSL1 1010	MSL1 0719	MSL1 0666

Appendix

MSL1 0269	MSL1 0558	MSL1 0605	MSL1 1031	MSL1 0719	MSL1 0676
MSL1 0269	MSL1 0703	MSL1 0614	MSL1 0146	MSL1 0719	MSL1 0697
MSL1 0269	MSL1 0916	MSL1 0614	MSL1 0260	MSL1 0719	MSL1 0703
MSL1 0425	MSL1 0449	MSL1 0614	MSL1 0449	MSL1 0719	MSL1 0704
MSL1 0425	MSL1 0628	MSL1 0614	MSL1 0534	MSL1 0719	MSL1 0776
MSL1 0449	MSL1 0558	MSL1 0614	MSL1 0555	MSL1 0719	MSL1 0804
MSL1 0449	MSL1 0562	MSL1 0614	MSL1 0558	MSL1 0719	MSL1 0823
MSL1 0449	MSL1 0628	MSL1 0614	MSL1 0562	MSL1 0719	MSL1 0902
MSL1 0449	MSL1 0697	MSL1 0614	MSL1 0574	MSL1 0719	MSL1 0908
MSL1 0449	MSL1 0703	MSL1 0614	MSL1 0579	MSL1 0719	MSL1 0916
MSL1 0459	MSL1 0146	MSL1 0614	MSL1 0582	MSL1 0719	MSL1 0997
MSL1 0459	MSL1 0269	MSL1 0614	MSL1 0592	MSL1 0719	MSL1 1009
MSL1 0459	MSL1 0425	MSL1 0614	MSL1 0599	MSL1 0719	MSL1 1010
MSL1 0459	MSL1 0449	MSL1 0614	MSL1 0628	MSL1 0719	MSL1 1031
MSL1 0459	MSL1 0484	MSL1 0614	MSL1 0650	MSL1 0778	MSL1 0663
MSL1 0459	MSL1 0485	MSL1 0614	MSL1 0652	MSL1 0778	MSL1 0666
MSL1 0459	MSL1 0488	MSL1 0614	MSL1 0663	MSL1 0804	MSL1 0146
MSL1 0459	MSL1 0524	MSL1 0614	MSL1 0666	MSL1 0804	MSL1 0574
MSL1 0459	MSL1 0528	MSL1 0614	MSL1 0668	MSL1 0804	MSL1 0666
MSL1 0459	MSL1 0530	MSL1 0614	MSL1 0675	MSL1 0804	MSL1 0668
MSL1 0459	MSL1 0555	MSL1 0614	MSL1 0676	MSL1 0804	MSL1 0776
MSL1 0459	MSL1 0574	MSL1 0614	MSL1 0697	MSL1 0804	MSL1 0902
MSL1 0459	MSL1 0582	MSL1 0614	MSL1 0703	MSL1 0804	MSL1 0916
MSL1 0459	MSL1 0592	MSL1 0614	MSL1 0704	MSL1 0804	MSL1 1010
MSL1 0459	MSL1 0605	MSL1 0614	MSL1 0718	MSL1 0816	MSL1 0652
MSL1 0459	MSL1 0621	MSL1 0614	MSL1 0719	MSL1 0816	MSL1 0823
MSL1 0459	MSL1 0628	MSL1 0614	MSL1 0776	MSL1 0816	MSL1 0902
MSL1 0459	MSL1 0663	MSL1 0614	MSL1 0823	MSL1 0816	MSL1 0916
MSL1 0459	MSL1 0675	MSL1 0614	MSL1 0899	MSL1 0816	MSL1 0997
MSL1 0459	MSL1 0676	MSL1 0614	MSL1 0997	MSL1 0816	MSL1 1009
MSL1 0459	MSL1 0703	MSL1 0614	MSL1 1009	MSL1 0816	MSL1 1010
MSL1 0459	MSL1 0704	MSL1 0614	MSL1 1010	MSL1 0823	MSL1 0006
MSL1 0459	MSL1 0916	MSL1 0614	MSL1 1031	MSL1 0823	MSL1 0776
MSL1 0459	MSL1 0997	MSL1 0621	MSL1 0449	MSL1 0823	MSL1 0899
MSL1 0459	MSL1 1009	MSL1 0621	MSL1 0555	MSL1 0823	MSL1 0916
MSL1 0459	MSL1 1010	MSL1 0621	MSL1 0558	MSL1 0870	MSL1 1010
MSL1 0464	MSL1 0146	MSL1 0621	MSL1 0574	MSL1 0882	MSL1 0006
MSL1 0464	MSL1 0153	MSL1 0621	MSL1 0579	MSL1 0882	MSL1 0146
MSL1 0464	MSL1 0269	MSL1 0621	MSL1 0582	MSL1 0882	MSL1 0816
MSL1 0464	MSL1 0425	MSL1 0621	MSL1 0614	MSL1 0882	MSL1 0823
MSL1 0464	MSL1 0449	MSL1 0621	MSL1 0628	MSL1 0882	MSL1 0902
MSL1 0464	MSL1 0484	MSL1 0621	MSL1 0650	MSL1 0882	MSL1 0916

Appendix

MSL1 0464	MSL1 0485	MSL1 0621	MSL1 0652	MSL1 0882	MSL1 1009
MSL1 0464	MSL1 0488	MSL1 0621	MSL1 0666	MSL1 0882	MSL1 1010
MSL1 0464	MSL1 0524	MSL1 0621	MSL1 0703	MSL1 0899	MSL1 0908
MSL1 0464	MSL1 0528	MSL1 0621	MSL1 0704	MSL1 0902	MSL1 0131
MSL1 0464	MSL1 0530	MSL1 0621	MSL1 0776	MSL1 0902	MSL1 0133
MSL1 0464	MSL1 0558	MSL1 0621	MSL1 0916	MSL1 0902	MSL1 0486
MSL1 0464	MSL1 0574	MSL1 0628	MSL1 0555	MSL1 0902	MSL1 0488
MSL1 0464	MSL1 0579	MSL1 0628	MSL1 0652	MSL1 0902	MSL1 0628
MSL1 0464	MSL1 0582	MSL1 0628	MSL1 0666	MSL1 0902	MSL1 0776
MSL1 0464	MSL1 0592	MSL1 0628	MSL1 0776	MSL1 0902	MSL1 0899
MSL1 0464	MSL1 0605	MSL1 0652	MSL1 0449	MSL1 0902	MSL1 0908
MSL1 0464	MSL1 0614	MSL1 0652	MSL1 0555	MSL1 0902	MSL1 0916
MSL1 0464	MSL1 0621	MSL1 0652	MSL1 0570	MSL1 0916	MSL1 0006
MSL1 0464	MSL1 0628	MSL1 0652	MSL1 0574	MSL1 0916	MSL1 0558
MSL1 0464	MSL1 0663	MSL1 0652	MSL1 0579	MSL1 0916	MSL1 0628
MSL1 0464	MSL1 0703	MSL1 0652	MSL1 0582	MSL1 0916	MSL1 0650
MSL1 0464	MSL1 0704	MSL1 0652	MSL1 0614	MSL1 0916	MSL1 0652
MSL1 0464	MSL1 0823	MSL1 0652	MSL1 0621	MSL1 0916	MSL1 0776
MSL1 0464	MSL1 0916	MSL1 0652	MSL1 0628	MSL1 0916	MSL1 0899
MSL1 0464	MSL1 0997	MSL1 0652	MSL1 0650	MSL1 0916	MSL1 0916
MSL1 0464	MSL1 1010	MSL1 0652	MSL1 0666	MSL1 0997	MSL1 0006
MSL1 0479	MSL1 0131	MSL1 0652	MSL1 0668	MSL1 0997	MSL1 0133
MSL1 0479	MSL1 0484	MSL1 0652	MSL1 0675	MSL1 0997	MSL1 0146
MSL1 0479	MSL1 0485	MSL1 0652	MSL1 0676	MSL1 0997	MSL1 0574
MSL1 0479	MSL1 0486	MSL1 0652	MSL1 0697	MSL1 0997	MSL1 0579
MSL1 0479	MSL1 0488	MSL1 0652	MSL1 0703	MSL1 0997	MSL1 0582
MSL1 0479	MSL1 0501	MSL1 0652	MSL1 0704	MSL1 0997	MSL1 0592
MSL1 0479	MSL1 0524	MSL1 0652	MSL1 0719	MSL1 0997	MSL1 0614
MSL1 0479	MSL1 0558	MSL1 0652	MSL1 0776	MSL1 0997	MSL1 0621
MSL1 0479	MSL1 0599	MSL1 0652	MSL1 0778	MSL1 0997	MSL1 0628
MSL1 0479	MSL1 0605	MSL1 0652	MSL1 0823	MSL1 0997	MSL1 0675
MSL1 0479	MSL1 0621	MSL1 0652	MSL1 0908	MSL1 0997	MSL1 0676
MSL1 0479	MSL1 0652	MSL1 0652	MSL1 0916	MSL1 0997	MSL1 0703
MSL1 0479	MSL1 0703	MSL1 0663	MSL1 0146	MSL1 0997	MSL1 0704
MSL1 0479	MSL1 0704	MSL1 0663	MSL1 0269	MSL1 0997	MSL1 0719
MSL1 0484	MSL1 0133	MSL1 0663	MSL1 0555	MSL1 0997	MSL1 0804
MSL1 0484	MSL1 0488	MSL1 0663	MSL1 0562	MSL1 0997	MSL1 0823
MSL1 0485	MSL1 0131	MSL1 0663	MSL1 0574	MSL1 0997	MSL1 0899
MSL1 0485	MSL1 0133	MSL1 0663	MSL1 0579	MSL1 0997	MSL1 0908
MSL1 0485	MSL1 0486	MSL1 0663	MSL1 0582	MSL1 0997	MSL1 0916
MSL1 0485	MSL1 0488	MSL1 0663	MSL1 0592	MSL1 0997	MSL1 1009
MSL1 0488	MSL1 0260	MSL1 0663	MSL1 0599	MSL1 0997	MSL1 1010

Appendix

MSL1 0488	MSL1 0484	MSL1 0663	MSL1 0614	MSL1 0997	MSL1 1031
MSL1 0488	MSL1 0485	MSL1 0663	MSL1 0621	MSL1 1009	MSL1 0006
MSL1 0488	MSL1 0558	MSL1 0663	MSL1 0628	MSL1 1009	MSL1 0146
MSL1 0488	MSL1 0574	MSL1 0663	MSL1 0668	MSL1 1009	MSL1 0269
MSL1 0488	MSL1 0579	MSL1 0663	MSL1 0675	MSL1 1009	MSL1 0558
MSL1 0488	MSL1 0621	MSL1 0663	MSL1 0676	MSL1 1009	MSL1 0574
MSL1 0488	MSL1 0628	MSL1 0663	MSL1 0697	MSL1 1009	MSL1 0621
MSL1 0488	MSL1 0718	MSL1 0663	MSL1 0703	MSL1 1009	MSL1 0628
MSL1 0517	MSL1 0501	MSL1 0663	MSL1 0704	MSL1 1009	MSL1 0666
MSL1 0517	MSL1 0558	MSL1 0663	MSL1 0719	MSL1 1009	MSL1 0804
MSL1 0524	MSL1 0534	MSL1 0663	MSL1 0804	MSL1 1009	MSL1 0823
MSL1 0524	MSL1 0558	MSL1 0663	MSL1 0908	MSL1 1009	MSL1 0899
MSL1 0524	MSL1 0574	MSL1 0663	MSL1 0916	MSL1 1009	MSL1 0908
MSL1 0524	MSL1 0579	MSL1 0663	MSL1 0997	MSL1 1009	MSL1 0916
MSL1 0524	MSL1 0614	MSL1 0663	MSL1 1009	MSL1 1009	MSL1 1010
MSL1 0524	MSL1 0628	MSL1 0663	MSL1 1010	MSL1 1009	MSL1 1031
MSL1 0528	MSL1 0484	MSL1 0666	MSL1 0555	MSL1 1010	MSL1 0006
MSL1 0528	MSL1 0485	MSL1 0666	MSL1 0579	MSL1 1010	MSL1 0146
MSL1 0528	MSL1 0534	MSL1 0666	MSL1 0628	MSL1 1010	MSL1 0269
MSL1 0528	MSL1 0555	MSL1 0666	MSL1 0675	MSL1 1010	MSL1 0558
MSL1 0528	MSL1 0558	MSL1 0666	MSL1 0676	MSL1 1010	MSL1 0574
MSL1 0528	MSL1 0574	MSL1 0666	MSL1 0719	MSL1 1010	MSL1 0614
MSL1 0528	MSL1 0628	MSL1 0666	MSL1 0916	MSL1 1010	MSL1 0621
MSL1 0530	MSL1 0534	MSL1 0666	MSL1 1009	MSL1 1010	MSL1 0628
MSL1 0530	MSL1 0574	MSL1 0666	MSL1 1010	MSL1 1010	MSL1 0650
MSL1 0534	MSL1 0555	MSL1 0668	MSL1 0614	MSL1 1010	MSL1 0652
MSL1 0534	MSL1 0558	MSL1 0668	MSL1 0628	MSL1 1010	MSL1 0666
MSL1 0534	MSL1 0628	MSL1 0668	MSL1 0652	MSL1 1010	MSL1 0776
MSL1 0558	MSL1 0484	MSL1 0668	MSL1 0703	MSL1 1010	MSL1 0804
MSL1 0558	MSL1 0485	MSL1 0668	MSL1 0704	MSL1 1010	MSL1 0823
MSL1 0558	MSL1 0582	MSL1 0668	MSL1 0719	MSL1 1010	MSL1 0899
MSL1 0558	MSL1 0628	MSL1 0675	MSL1 0582	MSL1 1010	MSL1 0916
MSL1 0558	MSL1 0908	MSL1 0675	MSL1 0628	MSL1 1010	MSL1 1010
MSL1 0562	MSL1 0555	MSL1 0675	MSL1 0666	MSL1 1010	MSL1 1031
MSL1 0570	MSL1 0555	MSL1 0675	MSL1 0668	MSL1 1031	MSL1 0131
MSL1 0570	MSL1 0574	MSL1 0675	MSL1 0697	MSL1 1031	MSL1 0146
MSL1 0570	MSL1 0579	MSL1 0675	MSL1 0703	MSL1 1031	MSL1 0486
MSL1 0570	MSL1 0582	MSL1 0675	MSL1 0704	MSL1 1031	MSL1 0574
MSL1 0570	MSL1 0628	MSL1 0675	MSL1 0719	MSL1 1031	MSL1 0614
MSL1 0574	MSL1 0260	MSL1 0676	MSL1 0621	MSL1 1031	MSL1 0628
MSL1 0574	MSL1 0449	MSL1 0676	MSL1 0628	MSL1 1031	MSL1 0666
MSL1 0574	MSL1 0555	MSL1 0676	MSL1 0663	MSL1 1031	MSL1 0697

MSL1 0574	MSL1 0558	MSL1 0676	MSL1 0666	MSL1 1031	MSL1 0719
MSL1 0574	MSL1 0562	MSL1 0676	MSL1 0668	MSL1 1031	MSL1 0776
MSL1 0574	MSL1 0582	MSL1 0676	MSL1 0675	MSL1 1031	MSL1 0823
MSL1 0574	MSL1 0614	MSL1 0676	MSL1 0697	MSL1 1031	MSL1 0902
MSL1 0574	MSL1 0628	MSL1 0676	MSL1 0703	MSL1 1031	MSL1 0916
MSL1 0574	MSL1 0652	MSL1 0676	MSL1 0719	MSL1 1031	MSL1 1009
MSL1 0574	MSL1 0666	MSL1 0676	MSL1 0776	MSL1 1031	MSL1 1010
MSL1 0574	MSL1 0668	MSL1 0697	MSL1 0628		

12.2.3 Ref2270 (MSL2 alone)

Table 3: MSL2 was crosslinked with BS3 (0.25, 0.5, 0.75 and 1.0 mM). Summary of detected crosslink positions with FDR <5%. Duplicate values have been removed to eliminate redundancy, resulting in a cleaner overview. Data of Ref. 2270 shown.

Protein1	Protein2	Protein1	Protein2	Protein1	Protein2
MSL2 008	MSL2 496	MSL2 464	MSL2 449	MSL2 496	MSL2 512
MSL2 436	MSL2 442	MSL2 464	MSL2 451	MSL2 496	MSL2 514
MSL2 436	MSL2 449	MSL2 464	MSL2 473	MSL2 496	MSL2 519
MSL2 436	MSL2 451	MSL2 464	MSL2 481	MSL2 496	MSL2 716
MSL2 436	MSL2 481	MSL2 473	MSL2 451	MSL2 498	MSL2 473
MSL2 436	MSL2 498	MSL2 473	MSL2 481	MSL2 498	MSL2 481
MSL2 442	MSL2 451	MSL2 473	MSL2 498	MSL2 498	MSL2 487
MSL2 442	MSL2 453	MSL2 481	MSL2 498	MSL2 498	MSL2 514
MSL2 442	MSL2 481	MSL2 481	MSL2 503	MSL2 498	MSL2 716
MSL2 442	MSL2 487	MSL2 481	MSL2 514	MSL2 503	MSL2 473
MSL2 442	MSL2 496	MSL2 481	MSL2 716	MSL2 503	MSL2 481
MSL2 442	MSL2 498	MSL2 487	MSL2 449	MSL2 503	MSL2 512
MSL2 442	MSL2 503	MSL2 487	MSL2 453	MSL2 503	MSL2 514
MSL2 449	MSL2 451	MSL2 487	MSL2 473	MSL2 514	MSL2 057
MSL2 449	MSL2 453	MSL2 487	MSL2 481	MSL2 514	MSL2 512
MSL2 449	MSL2 481	MSL2 487	MSL2 496	MSL2 519	MSL2 514
MSL2 449	MSL2 498	MSL2 487	MSL2 498	MSL2 519	MSL2 517
MSL2 449	MSL2 716	MSL2 487	MSL2 503	MSL2 522	MSL2 517
MSL2 453	MSL2 449	MSL2 487	MSL2 716	MSL2 524	MSL2 514
MSL2 453	MSL2 451	MSL2 496	MSL2 473	MSL2 640	MSL2 498
MSL2 453	MSL2 473	MSL2 496	MSL2 481	MSL2 653	MSL2 481
MSL2 453	MSL2 481	MSL2 496	MSL2 498	MSL2 653	MSL2 503
MSL2 453	MSL2 716	MSL2 496	MSL2 503	MSL2 715	MSL2 716

12.2.4 Ref2126 (3-MSL)

Table 4: 3-MSL (MSL1-MSL3-MOF) was crosslinked with BS3 (1 mM, two MS runs). Summary of the data given. Crosslink positions with FDR <5%. Duplicate values have been removed to eliminate redundancy, resulting in a cleaner overview. The values are ordered by 'Protein 1' position of the crosslink. Data of Ref. 2126 shown.

Protein1	Protein2	Protein1	Protein2	Protein1	Protein2
MOF 0483	MOF 0501	MSL1 0599	MSL1 0574	MSL3 0034	MSL3 0110
MOF 0483	MSL1 0916	MSL1 0599	MSL1 0579	MSL3 0034	MSL3 0111
MOF 0483	MSL3 0406	MSL1 0599	MSL1 0582	MSL3 0034	MSL3 0224
MOF 0483	MSL3 0420	MSL1 0599	MSL1 0614	MSL3 0053	MSL1 0823
MOF 0518	MSL1 0916	MSL1 0599	MSL1 0621	MSL3 0053	MSL1 0916
MOF 0532	MOF 0518	MSL1 0599	MSL3 0224	MSL3 0053	MSL3 0110
MOF 0532	MSL1 0916	MSL1 0605	MSL1 0579	MSL3 0053	MSL3 0111
MOF 0539	MOF 0478	MSL1 0605	MSL1 0582	MSL3 0053	MSL3 0170
MOF 0539	MOF 0501	MSL1 0605	MSL1 0592	MSL3 0053	MSL3 0224
MOF 0539	MOF 0507	MSL1 0605	MSL1 0621	MSL3 0091	MSL1 0823
MOF 0539	MSL1 0899	MSL1 0605	MSL1 0628	MSL3 0091	MSL1 0916
MOF 0567	MOF 0539	MSL1 0605	MSL3 0224	MSL3 0091	MSL3 0106
MOF 0567	MOF 0541	MSL1 0614	MSL1 0558	MSL3 0091	MSL3 0110
MOF 0618	MSL1 0916	MSL1 0614	MSL1 0574	MSL3 0091	MSL3 0111
MOF 0671	MOF 0706	MSL1 0614	MSL1 0582	MSL3 0091	MSL3 0224
MOF 0671	MSL1 0916	MSL1 0614	MSL1 0592	MSL3 0101	MOF 0776
MOF 0694	MOF 0798	MSL1 0614	MSL1 0599	MSL3 0101	MSL3 0034
MOF 0694	MOF 0801	MSL1 0614	MSL1 0628	MSL3 0101	MSL3 0091
MOF 0776	MSL1 0675	MSL1 0621	MSL1 0628	MSL3 0101	MSL3 0170
MOF 0776	MSL1 0676	MSL1 0628	MOF 0170	MSL3 0101	MSL3 0224
MOF 0776	MSL1 0816	MSL1 0628	MOF 0172	MSL3 0116	MOF 0776
MOF 0776	MSL1 0823	MSL1 0628	MSL1 0650	MSL3 0116	MSL3 0034
MOF 0776	MSL3 0026	MSL1 0628	MSL1 0652	MSL3 0116	MSL3 0091
MOF 0776	MSL3 0106	MSL1 0628	MSL1 0666	MSL3 0159	MSL3 0157
MOF 0776	MSL3 0159	MSL1 0652	MSL1 0666	MSL3 0161	MOF 0501
MOF 0776	MSL3 0161	MSL1 0663	MSL1 0628	MSL3 0161	MOF 0607
MOF 0776	MSL3 0170	MSL1 0663	MSL1 0668	MSL3 0161	MOF 0690
MOF 0776	MSL3 0224	MSL1 0663	MSL1 0675	MSL3 0161	MOF 0706
MOF 0798	MOF 0738	MSL1 0663	MSL1 0676	MSL3 0161	MSL1 0071
MOF 0801	MOF 0483	MSL1 0668	MSL1 0628	MSL3 0161	MSL1 0173
MOF 0801	MOF 0738	MSL1 0675	MSL1 0628	MSL3 0161	MSL1 0449
MOF 0801	MOF 0798	MSL1 0675	MSL1 0666	MSL3 0161	MSL1 0555
MSL1 0102	MSL3 0024	MSL1 0676	MSL1 0663	MSL3 0161	MSL1 0650
MSL1 0141	MSL1 0146	MSL1 0676	MSL1 0666	MSL3 0161	MSL1 0697
MSL1 0146	MSL1 0260	MSL1 0676	MSL1 0668	MSL3 0161	MSL1 0804
MSL1 0153	MOF 0798	MSL1 0703	MSL1 0628	MSL3 0161	MSL1 0908

Appendix

MSL1 0236	MSL1 0146	MSL1 0703	MSL1 0719	MSL3 0161	MSL1 1034
MSL1 0236	MSL1 0269	MSL1 0704	MSL1 0628	MSL3 0161	MSL3 0157
MSL1 0255	MSL1 0146	MSL1 0704	MSL1 0697	MSL3 0161	MSL3 0159
MSL1 0255	MSL1 0269	MSL1 0704	MSL1 0719	MSL3 0161	MSL3 0170
MSL1 0268	MSL1 0260	MSL1 0718	MSL1 0628	MSL3 0224	MOF 0145
MSL1 0269	MSL1 0260	MSL1 0718	MSL1 0703	MSL3 0224	MOF 0172
MSL1 0269	MSL1 0628	MSL1 0718	MSL1 0704	MSL3 0224	MOF 0671
MSL1 0449	MSL1 0484	MSL1 0718	MSL1 0719	MSL3 0224	MSL1 0011
MSL1 0449	MSL1 0628	MSL1 0718	MSL3 0224	MSL3 0224	MSL1 0269
MSL1 0459	MSL1 0449	MSL1 0719	MSL1 0628	MSL3 0224	MSL1 0449
MSL1 0459	MSL1 0484	MSL1 0719	MSL1 0703	MSL3 0224	MSL1 0484
MSL1 0459	MSL1 0485	MSL1 0719	MSL3 0224	MSL3 0224	MSL1 0485
MSL1 0464	MSL1 0449	MSL1 0778	MOF 0776	MSL3 0224	MSL1 0486
MSL1 0479	MSL1 0488	MSL1 0778	MSL1 0804	MSL3 0224	MSL1 0555
MSL1 0479	MSL3 0224	MSL1 0778	MSL1 0823	MSL3 0224	MSL1 0574
MSL1 0484	MSL1 0133	MSL1 0778	MSL3 0224	MSL3 0224	MSL1 0579
MSL1 0484	MSL1 0488	MSL1 0804	MOF 0776	MSL3 0224	MSL1 0582
MSL1 0485	MSL1 0131	MSL1 0804	MSL1 0823	MSL3 0224	MSL1 0614
MSL1 0485	MSL1 0133	MSL1 0804	MSL3 0034	MSL3 0224	MSL1 0621
MSL1 0485	MSL1 0486	MSL1 0804	MSL3 0170	MSL3 0224	MSL1 0628
MSL1 0485	MSL1 0488	MSL1 0804	MSL3 0224	MSL3 0224	MSL1 0652
MSL1 0488	MSL1 0484	MSL1 0816	MOF 0776	MSL3 0224	MSL1 0666
MSL1 0488	MSL1 0485	MSL1 0816	MSL1 0997	MSL3 0224	MSL1 0675
MSL1 0528	MSL1 0534	MSL1 0816	MSL3 0170	MSL3 0224	MSL1 0676
MSL1 0530	MSL1 0534	MSL1 0916	MOF 0501	MSL3 0224	MSL1 0703
MSL1 0558	MSL1 0582	MSL1 0916	MOF 0507	MSL3 0224	MSL1 0704
MSL1 0558	MSL1 0628	MSL1 0997	MSL1 0628	MSL3 0224	MSL1 0719
MSL1 0562	MSL1 0555	MSL1 0997	MSL1 0804	MSL3 0224	MSL1 0776
MSL1 0574	MSL1 0558	MSL1 0997	MSL1 0823	MSL3 0224	MSL1 0916
MSL1 0574	MSL1 0562	MSL1 0997	MSL3 0034	MSL3 0224	MSL1 1010
MSL1 0574	MSL1 0582	MSL1 0997	MSL3 0170	MSL3 0224	MSL3 0110
MSL1 0574	MSL1 0628	MSL1 1009	MSL3 0034	MSL3 0224	MSL3 0111
MSL1 0579	MSL1 0558	MSL1 1009	MSL3 0170	MSL3 0224	MSL3 0112
MSL1 0582	MSL1 0484	MSL1 1010	MSL3 0034	MSL3 0224	MSL3 0170
MSL1 0582	MSL1 0562	MSL1 1031	MSL1 1010	MSL3 0224	MSL3 0224
MSL1 0582	MSL1 0579	MSL3 0034	MOF 0776	MSL3 0224	MSL3 0461
MSL1 0582	MSL1 0628	MSL3 0034	MSL1 0449	MSL3 0243	MSL3 0224
MSL1 0592	MSL1 0574	MSL3 0034	MSL1 0628	MSL3 0406	MSL3 0389
MSL1 0592	MSL1 0579	MSL3 0034	MSL1 0666	MSL3 0420	MOF 0501
MSL1 0592	MSL1 0582	MSL3 0034	MSL1 0703	MSL3 0447	MSL3 0220
MSL1 0592	MSL1 0614	MSL3 0034	MSL1 0704	MSL3 0460	MSL1 1031
MSL1 0592	MSL1 0621	MSL3 0034	MSL1 0823	MSL3 0461	MSL1 1031

| MSL1 0592 | MSL1 0628 | MSL3 0034 | MSL1 0916 | MSL3 0461 | MSL3 0170 |
| MSL1 0592 | MSL3 0224 | MSL3 0034 | MSL3 0074 | | |

12.2.5 Ref2185 (4-MSL)

Table 5: 4-MSL (MSL1-MSL2-MSL3-MOF) was crosslinked with BS3 (0.75, 1.0 and 1.5 mM). Summary of the data given. Crosslink positions with FDR <5%, sorted by 'Protein 1'. Duplicate values have been removed to eliminate redundancy, resulting in a cleaner overview. Data of Ref. 2185 shown.

Protein1	Protein2	Protein1	Protein2	Protein1	Protein2
MOF 0567	MOF 0539	MSL1 0663	MSL1 0628	MSL2 0496	MSL2 0481
MOF 0567	MOF 0541	MSL1 0663	MSL1 0675	MSL2 0496	MSL2 0503
MOF 0671	MSL1 0916	MSL1 0663	MSL1 0676	MSL2 0496	MSL2 0514
MOF 0770	MOF 0706	MSL1 0666	MSL1 0628	MSL2 0498	MSL2 0503
MOF 0773	MOF 0706	MSL1 0675	MSL1 0666	MSL2 0498	MSL2 0514
MOF 0776	MSL2 0823	MSL1 0676	MSL1 0666	MSL2 0519	MSL2 0514
MOF 0801	MOF 0798	MSL1 0676	MSL1 0668	MSL2 0519	MSL2 0517
MOF 0801	MSL2 0168	MSL1 0704	MSL1 0628	MSL3 0091	MSL3 0110
MOF 0801	MSL3 0460	MSL1 0704	MSL1 0697	MSL3 0091	MSL3 0111
MSL1 0141	MSL1 0146	MSL1 0704	MSL1 0719	MSL3 0091	MSL3 0170
MSL1 0146	MSL3 0224	MSL1 0997	MSL1 0823	MSL3 0091	MSL3 0224
MSL1 0153	MSL1 0141	MSL1 1009	MOF 0086	MSL3 0101	MSL3 0026
MSL1 0168	MSL1 0173	MSL1 1009	MOF 0478	MSL3 0101	MSL3 0034
MSL1 0268	MSL1 0260	MSL1 1009	MOF 0507	MSL3 0101	MSL3 0091
MSL1 0269	MSL1 0260	MSL1 1009	MSL1 0003	MSL3 0106	MSL3 0026
MSL1 0440	MSL1 0449	MSL1 1009	MSL1 0011	MSL3 0116	MSL3 0101
MSL1 0459	MSL1 0449	MSL1 1009	MSL1 0255	MSL3 0159	MSL3 0157
MSL1 0464	MSL1 0484	MSL1 1009	MSL1 0562	MSL3 0161	MOF 0501
MSL1 0464	MSL1 0485	MSL1 1009	MSL1 1031	MSL3 0161	MOF 0607
MSL1 0479	MSL1 0488	MSL1 1009	MSL1 1034	MSL3 0161	MOF 0690
MSL1 0484	MSL1 0133	MSL1 1009	MSL2 0022	MSL3 0161	MOF 0706
MSL1 0484	MSL1 0488	MSL1 1009	MSL2 0057	MSL3 0161	MSL1 0071
MSL1 0484	MSL2 0514	MSL1 1009	MSL2 0473	MSL3 0161	MSL1 0173
MSL1 0488	MSL1 0484	MSL1 1009	MSL2 0475	MSL3 0161	MSL1 0449
MSL1 0488	MSL1 0485	MSL1 1009	MSL3 0112	MSL3 0161	MSL1 0555
MSL1 0528	MSL1 0534	MSL1 1009	MSL3 0157	MSL3 0161	MSL1 0650
MSL1 0530	MSL1 0534	MSL2 0135	MSL1 0269	MSL3 0161	MSL1 0697
MSL1 0570	MSL1 0574	MSL2 0135	MSL2 0072	MSL3 0161	MSL1 0804
MSL1 0574	MSL1 0558	MSL2 0442	MSL2 0453	MSL3 0161	MSL1 0908
MSL1 0574	MSL1 0582	MSL2 0449	MSL2 0481	MSL3 0161	MSL1 1034
MSL1 0592	MSL1 0579	MSL2 0464	MSL1 0459	MSL3 0161	MSL2 0475
MSL1 0592	MSL1 0582	MSL2 0464	MSL2 0473	MSL3 0161	MSL3 0159
MSL1 0605	MSL1 0621	MSL2 0464	MSL2 0481	MSL3 0161	MSL3 0170

MSL1 0621	MOF 0170	MSL2 0473	MSL2 0481	MSL3 0224	MSL2 0498
MSL1 0621	MSL1 0628	MSL2 0481	MSL2 0498	MSL3 0254	MSL3 0406
MSL1 0621	MSL1 0650	MSL2 0481	MSL2 0514	MSL3 0447	MOF 0738
MSL1 0652	MSL1 0628	MSL2 0487	MSL2 0498	MSL3 0447	MSL1 0804
MSL1 0652	MSL3 0224	MSL2 0487	MSL2 0503	MSL3 0447	MSL3 0220

12.2.6 Ref5290 (2-MSL with MLE and roX2)

Table 6: 2-MSL (MSL1ΔC-MSL2) and MLE (equimolar) were crosslinked with BS3 (1.0 mM). roX2 RNA was added in indicated data subsets (0 = no RNA added, 1 = equimolar roX2 added, 10 = 10-fold excess roX2 added). Summary of the data given, ordered by 'Protein 1'. Crosslink positions with FDR <20%. Duplicate values have been removed to eliminate redundancy, resulting in a cleaner overview. Data of Ref. 5290 shown.

roX2	Protein1	Protein2	roX2	Protein1	Protein2	roX2	Protein1	Protein2
0	MLE 0019	MLE 0028	1	MLE 0053	MLE 0054	10	MLE 0053	MLE 0054
0	MLE 0053	MLE 0054	1	MLE 0053	MSL1 0048	10	MLE 0053	MSL1 0048
0	MLE 0188	MLE 0178	1	MLE 0053	MSL1 0255	10	MLE 0053	MSL1 0255
0	MLE 0280	MLE 0274	1	MLE 0261	MSL2 0481	10	MLE 0054	MLE 0058
0	MLE 0701	MSL2 0487	1	MLE 0280	MLE 0274	10	MLE 0073	MLE 0107
0	MSL1 0048	MSL1 0053	1	MLE 0319	MSL1 0449	10	MSL1 0048	MLE 0053
0	MSL1 0139	MSL1 0496	1	MLE 0622	MLE 0618	10	MSL1 0153	MSL1 0146
0	MSL1 0268	MLE 0260	1	MLE 0701	MSL2 0487	10	MSL1 0236	MSL1 0269
0	MSL1 0269	MSL1 0268	1	MLE 1149	MSL1 0621	10	MSL1 0268	MLE 0260
0	MSL1 0464	MSL1 0473	1	MSL1 0048	MLE 0053	10	MSL1 0269	MSL1 0268
0	MSL1 0487	MLE 0473	1	MSL1 0102	MSL2 0481	10	MSL1 0464	MSL1 0459
0	MSL1 0487	MLE 0481	1	MSL1 0168	MSL1 0146	10	MSL1 0570	MSL1 0562
0	MSL1 0487	MSL1 0253	1	MSL1 0168	MSL1 0153	10	MSL1 0574	MLE 0383
0	MSL1 0487	MSL1 0562	1	MSL1 0236	MSL1 0269	10	MSL1 0574	MSL1 0570
0	MSL1 0487	MSL1 0666	1	MSL1 0255	MSL1 0260	10	MSL1 0574	MSL1 0582
0	MSL1 0487	MSL1 0718	1	MSL1 0255	MSL1 0269	10	MSL1 0582	MSL1 0579
0	MSL1 0487	MSL1 0810	1	MSL1 0268	MSL1 0260	10	MSL1 0599	MSL1 0592
0	MSL1 0496	MLE 0498	1	MSL1 0269	MSL1 0268	10	MSL1 0605	MSL1 0592
0	MSL1 0496	MLE 0503	1	MSL1 0464	MSL1 0459	10	MSL1 0605	MSL1 0599
0	MSL1 0496	MSL1 0481	1	MSL1 0464	MSL2 0473	10	MSL1 0614	MSL1 0605
0	MSL1 0496	MSL1 0668	1	MSL1 0464	MSL2 0548	10	MSL1 0621	MSL1 0614
0	MSL1 0498	MSL1 0503	1	MSL1 0570	MSL1 0562	10	MSL1 0621	MSL1 0628
0	MSL1 0514	MSL1 0512	1	MSL1 0574	MSL1 0570	10	MSL1 0628	MSL1 0666
0	MSL1 0574	MSL1 0570	1	MSL1 0574	MSL1 0582	10	MSL1 0652	MSL1 0650
0	MSL1 0574	MSL1 0582	1	MSL1 0582	MSL1 0579	10	MSL1 0663	MLE 0224
0	MSL1 0582	MSL1 0579	1	MSL1 0599	MSL1 0592	10	MSL1 0663	MSL1 0666
0	MSL1 0605	MSL1 0918	1	MSL1 0605	MLE 0918	10	MSL1 0663	MSL1 0703
0	MSL1 0621	MSL1 0936	1	MSL1 0605	MSL1 0592	10	MSL1 0666	MLE 0224
0	MSL1 0663	MLE 0666	1	MSL1 0605	MSL1 0599	10	MSL1 0668	MLE 0253

0	MSL1 0663	MSL1 0178	1	MSL1 0605	MSL1 0614	10	MSL1 0668	MLE 0810
0	MSL1 0668	MSL1 0912	1	MSL1 0614	MLE 0936	10	MSL1 0668	MSL1 0562
0	MSL1 0718	MSL1 1055	1	MSL1 0614	MSL1 0599	10	MSL1 0668	MSL1 0666
0	MSL2 0464	MSL2 0459	1	MSL1 0614	MSL1 0605	10	MSL1 0668	MSL1 0718
0	MSL2 0570	MSL2 0562	1	MSL1 0621	MSL1 0614	10	MSL1 0668	MSL2 0473
0	MSL2 0574	MSL1 0579	1	MSL1 0621	MSL1 0628	10	MSL1 0668	MSL2 0481
0	MSL2 0599	MSL2 0592	1	MSL1 0650	MSL2 0512	10	MSL1 0675	MSL1 0666
0	MSL2 0605	MSL2 0599	1	MSL1 0652	MSL1 0650	10	MSL1 0676	MSL1 0666
0	MSL2 0621	MLE 0628	1	MSL1 0663	MSL1 0666	10	MSL1 0676	MSL1 0675
0	MSL2 0621	MSL2 0614	1	MSL1 0663	MSL1 0703	10	MSL1 0704	MSL1 0697
0	MSL2 0652	MSL2 0650	1	MSL1 0666	MSL1 0704	10	MSL1 0704	MSL1 0703
0	MSL2 0663	MSL2 0703	1	MSL1 0668	MLE 0253	10	MSL1 0704	MSL1 0704
0	MSL2 0676	MLE 0675	1	MSL1 0668	MLE 0810	10	MSL1 0718	MLE 1055
0	MSL2 0704	MSL1 0704	1	MSL1 0668	MSL1 0562	10	MSL1 0718	MSL1 0628
0	MSL2 0704	MSL2 0675	1	MSL1 0668	MSL1 0666	10	MSL1 0718	MSL1 0719
0	MSL2 0704	MSL2 0703	1	MSL1 0668	MSL1 0718	10	MSL1 0804	MSL1 0582
0	MSL2 0718	MSL1 0628	1	MSL1 0668	MSL2 0473	10	MSL1 0804	MSL1 0823
0	MSL2 0718	MSL1 0719	1	MSL1 0668	MSL2 0481	10	MSL1 0816	MSL1 0823
			1	MSL1 0676	MSL1 0668	10	MSL2 0464	MSL2 0473
			1	MSL1 0676	MSL1 0675	10	MSL2 0473	MSL1 0449
			1	MSL1 0704	MSL1 0675	10	MSL2 0487	MLE 0253
			1	MSL1 0704	MSL1 0697	10	MSL2 0487	MLE 0766
			1	MSL1 0704	MSL1 0703	10	MSL2 0487	MLE 0810
			1	MSL1 0704	MSL1 0704	10	MSL2 0487	MSL1 0260
			1	MSL1 0718	MLE 1055	10	MSL2 0487	MSL1 0562
			1	MSL1 0718	MSL1 0628	10	MSL2 0487	MSL1 0666
			1	MSL1 0718	MSL1 0719	10	MSL2 0487	MSL1 0718
			1	MSL1 0804	MSL1 0823	10	MSL2 0487	MSL2 0473
			1	MSL1 0823	MSL1 0776	10	MSL2 0487	MSL2 0481
			1	MSL2 0139	MSL2 0487	10	MSL2 0496	MLE 1081
			1	MSL2 0139	MSL2 0496	10	MSL2 0496	MSL2 0481
			1	MSL2 0487	MLE 0253	10	MSL2 0496	MSL2 0487
			1	MSL2 0487	MLE 0810	10	MSL2 0496	MSL2 0498
			1	MSL2 0487	MSL1 0562	10	MSL2 0498	MSL2 0503
			1	MSL2 0487	MSL1 0666	10	MSL2 0503	MLE 0546
			1	MSL2 0487	MSL1 0718	10	MSL2 0503	MLE 0712
			1	MSL2 0487	MSL2 0473	10	MSL2 0503	MLE 0918
			1	MSL2 0487	MSL2 0481	10	MSL2 0503	MSL1 0582
			1	MSL2 0496	MSL2 0481	10	MSL2 0503	MSL1 0614
			1	MSL2 0496	MSL2 0487	10	MSL2 0503	MSL1 0663
			1	MSL2 0496	MSL2 0498	10	MSL2 0503	MSL2 0285
			1	MSL2 0498	MSL2 0487	10	MSL2 0503	MSL2 0418

Appendix

			1	MSL2 0498	MSL2 0498	10	MSL2 0503	MSL2 0498
			1	MSL2 0498	MSL2 0503	10	MSL2 0781	MLE 1301
			1	MSL2 0514	MSL2 0512			

12.3 Chromatogram of the H4K16R mutant octamers and WT octamers

A Äkta run of wild-type octamers

Octamer wild-type purification of Superdex 200 16/600 prep grade
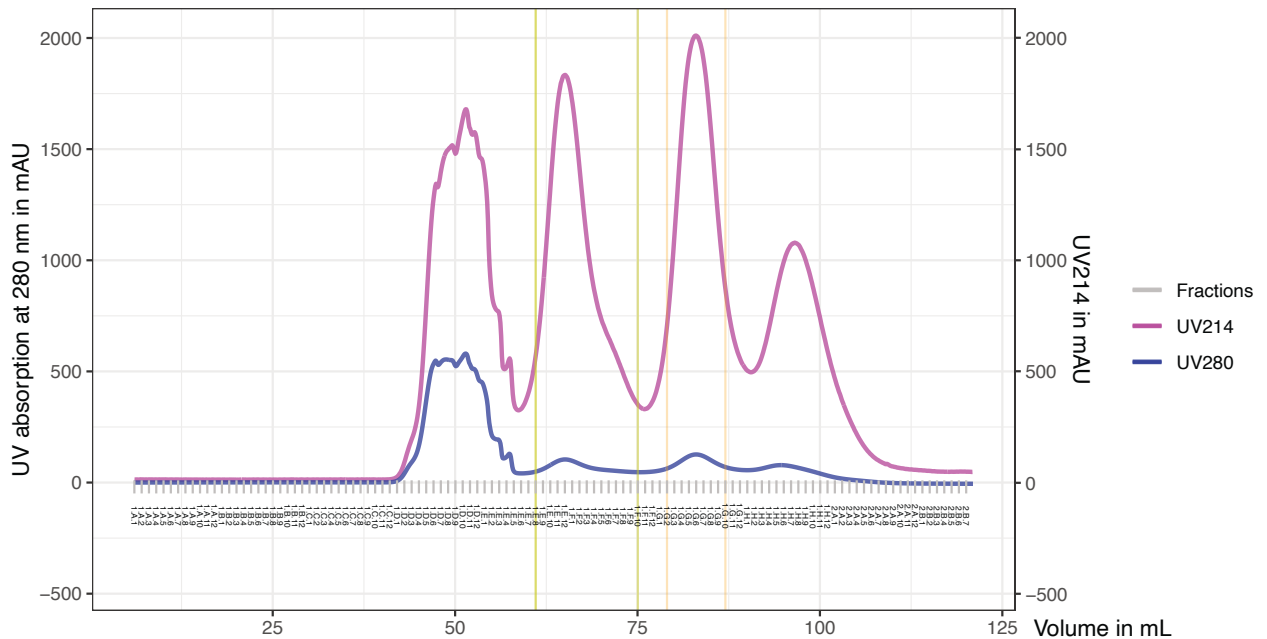
B Äkta run of mutant octamers

Octamer H4K16R purification of Superdex 200 16/600 prep grade
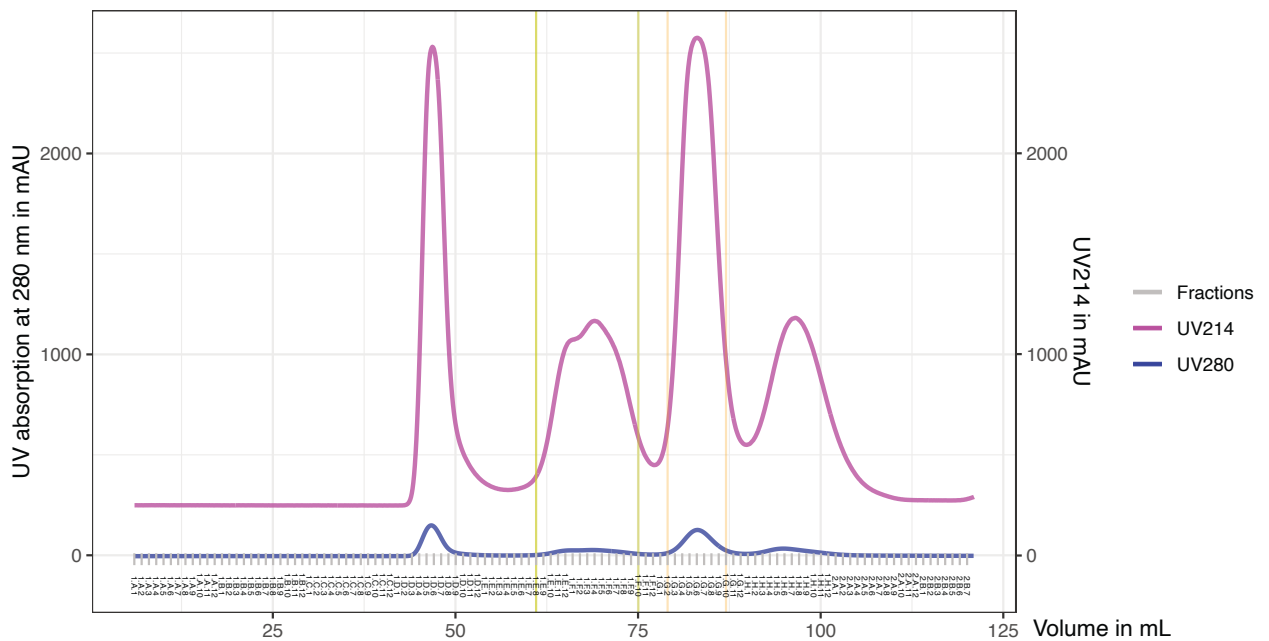
Figure 49: Size-exclusion chromatogram of the wild-type octamers and H4K16R mutant octamers demonstrate equal behavior on the size exclusion column. **A** Wild-type chromatogram with UV214 and UV280 nm absorptions. **B** H4K16R octamers chromatogram with UV214 and UV280 nm absorptions. Both octamer types eluted at 60-75 mL (yellow vertical lines) and H2A-H2B dimers at 77-85 mL (orange vertical lines).

12.4 Script to smooth curves of the Äkta chromatograms in R

```
UVplot <- function(UV1) {
```

Appendix

```
m <- read.table(UV1);          # dataframe m
mn = nrow(m);                                          # number of rows in m
r =((round(m[mn,1]))*11-15);                    # number of rows in output file
c = ncol(m);                                            # number of columns is the same
# create output matrix A, filled with 0 for the start
library(plotrix);                            # for twoord plot
library(ggplot2);

A = matrix(0.00, nrow=r, ncol=c);
# matrix a counter
a = 1;
counter=0;
UV280 = 0;

# mL counter
mL = round(m[1,1]);

for(i in 1:nrow(m)){
  if(i==nrow(m) && (A[r,c] ==0 || A[r,2] ==0 )){
    A[r,c] = UV280/counter; # Once you reach the end, write the last value in the last cell of matrix A
    A[r,1] = m[mn,1];      # Once you reach the end, write the last value in the last cell of matrix A
  }
  else if(mL + 0.1 >= m[i,1]){                        # in an interval of 0.1 mL sum all UV values and count them
    counter = counter +1;
    UV280 = UV280 + m[i,c];
  }
  else if(mL + 0.1 < m[i,1]){                         # when you reach the mL interval border
    A[a, c] = (A[a, c] + UV280/counter);
    A[a, 1] = A[a, 1] + mL                          # write the entries in matrix A
    counter=1;                                      # put the counter back to 1
    mL = m[i,1];                              # put the mL counter on the new value
    UV280 = m[i,c];                      # put the UV counter on the new value
    a=a+1;                                          # and the matrix counter on the next row
  }
}
# Once you have finished writing this table you could use it in excel:
write.table(A, file = paste(UV1, ".txt", sep = ""), sep = "\t");

vec_mL = c(A[,1]);
print(vec_mL);
print(A[,2]);
}
```

12.5 Script to plot curves of the Äkta chromatograms in R

```
UVdraw <- function(UV1, ConcB, Fracs, date){
  library ('ggplot2')
# we need 3 colors, these are the names for the labels
colors = c("UV260", "UV280", "Fractions");

UV1.df <- as.data.frame(as.matrix(read.table(UV1)))
ConcB.df <- as.data.frame(as.matrix(read.table(ConcB)))

# This will be for the fractionation
Frac.df <- as.data.frame(as.matrix(read.table(Fracs)))

#To plot chromatogram (matrix A):
pdf(paste(date, ".pdf", sep =""), width=8.5,height=5.0); # you could also plot .png or .jpeg if you need it
gg <- ggplot(UV1.df, aes(x=UV1.df[,1])) +
  geom_line(aes(x=ConcB.df[,1], y=ConcB.df[,2], color=colors[1]), size = 1.0, alpha=0.8) +
  geom_line(aes(y=UV1.df[,2], color=colors[2]), size = 1.0, alpha=0.8) +
  scale_y_continuous("UV absorption at 280 nm in mAU", sec.axis = sec_axis(~ ., name = "UV260 in mAU")) +
```

```
labs(title= paste("Äkta run of ", date, sep=""),
    subtitle="Piccolo-NuA4 complex purification on StrepTrap column 2 mL sandwich",
    x= "Volume in mL") +
scale_fill_discrete(name = "") +
theme_bw() +
geom_vline(xintercept = as.numeric(as.character(Frac.df$V1[Frac.df$V2 == "3.A.10"]))-as.numeric(as.character(Frac.df$V1[1])),
color = 'yellow3', alpha = 0.7) +
geom_vline(xintercept = as.numeric(as.character(Frac.df$V1[Frac.df$V2 == "3.B.9"]))-as.numeric(as.character(Frac.df$V1[1])),
color = 'yellow3', alpha = 0.5) +
#geom_vline(xintercept = as.numeric(as.character(Frac.df$V1[Frac.df$V2 == "1.G.2"])), color = 'orange', alpha = 0.3) +
#xlim(c(6,121)) + # limit the plotted range to a sensible mL range
geom_segment(aes(x  =  as.numeric(as.character(Frac.df$V1))-as.numeric(as.character(Frac.df$V1[1])), y = -13, xend =
as.numeric(as.character(Frac.df$V1))-as.numeric(as.character(Frac.df$V1[1])), yend = 1, colour = "Fractions"), data = Frac.df) +
#Fractions
annotate("text", x = as.numeric(as.character(Frac.df$V1))-as.numeric(as.character(Frac.df$V1[1])), y = -39, size = 2.5, angle =
270, label = as.character(Frac.df$V2)) +
scale_colour_manual(name='', values=c('Fractions'='grey', 'UV260'='red3', 'UV280'='blue3'))#actual color definition


plot(gg)
dev.off();
return(plot(gg))
}
```

12.6 Mass detection and analysis by Skyline of acetylation patterns of H4

Figure 50: The 4-MSL complex oligo-acetylates H4 tail lysines *in vitro*. **A** MS1 spectra illustrating the variation in relative abundance of mono-acetylation (Mono-ac), di-acetylation (Di-ac), tri-acetylation (Tri-ac), and tetra-acetylation (Tetra-ac) on Histone 4 (H4) over different time intervals for the wild-type sample. Only precursor ion m/z values are depicted. Data corresponds to MS Ref8241 (replicate 3 displayed). **B** MS1 spectra demonstrating time-dependent alterations in the relative abundance of mono-acetylation (Mono-ac) and di-acetylation (Di-ac) for the H4K16R ... distinctly observed. Data corresponds to MS Ref8241 (replicate 3 di... Venkatasubramani [216].

Appendix

12.7 Quantification of Western blots script

```r
quantitateWesternBlotte <- function(Input, SampleNames, AcetylationType, g, OutputName){
# Input <- "~/Documents/Data/WesternBlot.txt"
# SampleNames <- c("4-MSL no RNA control", "roX2 1x 280 ng", "roX2 2x 560 ng", "roX2 4x 1120 ng", "GFP-470 1x 280 ng",
"GFP-470 2x 560 ng", "GFP-470 4x 1120 ng", "roX123 1x 280 ng", "roX123 2x 560 ng", "roX123 4x 1120 ng", "GFP-76 1x 280
ng", "GFP-76 2x 560 ng", "GFP-76 4x 1120 ng", "No Enzyme")
# AcetylationType <- "H4K16ac"
# g <- position of the control in the table/ in the Western blot. If you don't want to normalize further, type g =0
# Output <- "~/Documents/Data/WesternBlotOutput"

  library(ggplot2)
  library('dplyr')
  library('tidyr')
  library("viridis")
  library(ggrepel)
  library("wesanderson")
  library("stringr")

# Read the values from the table of exported data from ImageStudioLite
  #intensities <- as_tibble(read.table(Input, sep ='\t', header = TRUE, dec = "."))
  intensities <- Input

# Define a couple of vectors c will be the normalized values, d is to check if the areas, i.e. the boxes, match.
# The normalizedSignals will be for the final result and the plot.
  c <- vector()
  d <- vector()
  e <- vector()
  normalizedSignals <- vector()
  normalizedSignals2 <- vector()
  e                      <-                      rep(seq(1,length(unique(intensities$Image.Name))),                      each
=length(intensities$Image.Name)/length(unique(intensities$Image.Name))/2)
  f <- rep(AcetylationType, times=nrow(intensities)/2)

# Calculate all normalized signals, irrespective whether they make sense
  for(i in 2:length(intensities$Name)) c[i-1] <- (intensities$Signal[i]/intensities$Signal[i-1])
# Calculate the area ratios
  for(i in 2:length(intensities$Name)) d[i-1] <- intensities$Area[i-1]/intensities$Area[i]
  print(d)

# Only if the area ratios match, i.e. ratio equals 1, the normalized signal will be written in the vector
  for(j in 1:length(d)){if(d[j]==1) normalizedSignals[j] <- c[j]}
  normalizedSignals <- normalizedSignals[!is.na(normalizedSignals)]
  print(normalizedSignals)
if(g == 0){
  normalizedSignals2 <- normalizedSignals
}
else{
  for(j in 1:length(normalizedSignals)){
      normalizedSignals2[j] <- normalizedSignals[j]/normalizedSignals[(g+length(SampleNames)*(e[j]-1))]
      }
  }
# Create a dataframe for the output / the plot. The SamplesNames given in the input will be used (use correct order!).
  x = data.frame(v1 = SampleNames, v2 = normalizedSignals2, v3 = e, v4 = f, v5 = paste(SampleNames, e, sep ="") )
  print(x)
# This causes the order to be exactly the order of the samples in the input table
  x$v1 <- factor(x$v1, levels = SampleNames)
  data1 <- x %>% select(v1, v2, v3)
  my_sum1 <- data1 %>%
      group_by(v1) %>%
      summarise(
          n=n(),
          mean=mean(v2),
```

```
        sd=sd(v2)
    ) %>%
   mutate( se=sd/sqrt(n))  %>%
    mutate( ic=se * qt((1-0.05)/2 + .5, n-1))


 # Define two shades of grey for the two replicates, repeated and alternated
 replicate_colors <- rep(c("#333333", "#666666"), length.out = nrow(x))

# Plotting time!
 p<-ggplot(data=x, aes(x=v1, y=v2, fill=v5)) +
     geom_bar(stat="identity",                    show.legend              =              FALSE,
position="dodge")+theme_minimal()+#scale_fill_grey()+#scale_fill_viridis(discrete = TRUE,option = "mako") +
     scale_fill_manual(values = replicate_colors) +
     theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
     xlab("HAT assay samples") +
     ylab(paste(AcetylationType, " relative acetylation", sep =""))

 p1<-ggplot(x, aes(x=as.factor(str_sub(v1, 1, nchar(as.character(v1)))), y=v2, shape = as.factor(v3), size = 4, fill="grey20")) +
#color = substr(v1,1,5))) +
     geom_point()+
     #geom_text_repel(aes(label = round(v2, digits = 3)), size = 4)+
     #scale_colour_grey()+
     theme_minimal()+ theme(legend.position="none", axis.text.x = element_text(angle = 90, hjust = 1)) +
     xlab("HAT assay samples")+
     ylab(paste(AcetylationType, " relative acetylation", sep =""))


 p2 <- ggplot(my_sum1) +
   geom_bar( aes(x=v1, y=mean), stat="identity", fill="grey20", alpha=0.9, width = 0.80) +
   geom_errorbar( aes(x=v1, ymin=mean-se, ymax=mean+se), width=0.25, colour="black", alpha=0.9, size=0.5) +
   theme_minimal()+ theme(legend.position="none", axis.text.x = element_text(angle = 90, hjust = 1))+
   ggtitle(paste(AcetylationType, " relative acetylation", sep =""))

 pdf(paste(OutputName, ".pdf", sep =""), width=7.0,height=4.5)
 plot(p2, label_size = 48)
 dev.off();

# Return the dataframe or the plot to check, if correct.
 return(list(p, p1, p2))
}
```

12.8 Histone acetylation by mass spectrometry heatmap and bar graph script

```
meansPlotte <- function(Input,SampleNames, OutputName){
 # SampleNames as such:
 # s1 ModellerSampleNames <- c("Modification", "SGD only 1",
 # "SGD +MLE 1",
 # "4-MSL 1",
 # "4-MSL + MLE 1",
 # "4-MSL + MLE + roX2 half 1",
 # "4-MSL + MLE + roX2 equal 1",
 # "4-MSL + MLE + roX2 double 1",
 # "4-MSl + MLE + tRNA 1",
 # "4-MSL + MLE + poly-dA 1") #etc
 # s2 <- c("Modification", "SGD 1", "SGD 2", "me0 1", "me0 DCC roX2 1","me0 DCC as-roX2 1", "me0 2", "me0 DCC roX2
2" , "me0 DCC as-roX2 2" )
 # if this is in effect colnames(Ref8241_percent)[1] <- "Modification", it is not necessary to have SampleNames.

 library("ggplot2")
 library("stringr")
 library('dplyr')
 library('tidyr')
```

Appendix

```r
  library("viridis")
  library("ggsci")
  library("wesanderson")
  library("RColorBrewer")
  library("tidyverse")

percentages <- Input
#percentages <- as_tibble(read.table(Input, sep =',', header = TRUE, dec = "."))
colnames(percentages)[1] <- "Modification"
y_positions <- as.factor(percentages$Modification) # Resolves the y_position nonsense above

#SampleNames <- colnames(percentages)
colnames(percentages) <- SampleNames
SuperNames <- substr(SampleNames[2:length(SampleNames)],1,nchar(SampleNames[2:length(SampleNames)])-2)
SuperNames <- gsub("_", ".", SuperNames) #' This was for a special case to combine 2 datasets'

c_name <- as.character(pull(percentages, Modification)) # this stops this whole c_name naming business

percentages$Modification <- as.character(percentages$Modification)
data_p <-gather(percentages,
        key = "Sample",
        value = "Percent", -Modification)
dada_p <- data_p %>% drop_na(Percent)
niceName <- substr(dada_p$Sample,1,nchar(dada_p$Sample)-2)
niceName <- gsub("_", ".", niceName) #' This was for a special case to combine 2 datasets'
dada_p <- cbind(dada_p, niceName) #dada_p will now have better names for the samples, too
dada_p <- dada_p[!grepl("Cells",as.character(dada_p$niceName)),] # To remove the 00m (0 min) time points/control from the
plot
dada_p <- dada_p[!grepl("poly",as.character(dada_p$Modification)),] # To remove the poly-dA or tRNA
dada_p2 <- dada_p %>%
  group_by(Modification, niceName) %>%
  summarise(mean_percent = mean(Percent), sd_percent = sd(Percent), sem_percent = sd(Percent)/sqrt(length(Percent))) #Tidy
the data and calculate the mean,
# standard deviation (sd) and standard error of the mean (sem).
dada_p2$Modification <- factor(dada_p2$Modification, levels = y_positions) # to achieve the desired order of Modification
panels
#dada_p2     <-     dada_p2     %>%     filter(niceName     %in%     c("4.MSL",     "4.MSL.MLE",
"4.MSL.MLE.roX2.half","4.MSL.MLE.roX2.equal","4.MSL.MLE.roX2.double","4.MSL.MLE.GFP.half","4.MSL.MLE.GFP.equal","
4.MSL.MLE.GFP.double","4.MSL.roX2")) # to remove some sample from the plot

z1 <- spread(dada_p, Modification, Percent)
z1[is.na(z1)] <- 0
z3 <- aggregate(z1[, 3:(ncol(z1))], list(z1$niceName), mean)
#z3 <- subset(z3, !endsWith(as.character(Group.1),"roX2"),)

z4 <- gather(z3,
        key = "Modification",
        value = "Percent", -Group.1)
z4$Modification <- as.character(z4$Modification)

z4 <- subset(z4, !endsWith(as.character(Modification),"R"),)
#z4 <- subset(z4, !endsWith(as.character(z4$Group.1),"A"),)
z4$Modification <- factor(z4$Modification, levels = rev(y_positions))
#z4 <- z4 %>%  filter(Modification %in% c("H4_otherMonoAc","H4K16ac","H4_diAc", "H4_triAc","H4_tetraAc", "H4_noPTM"
)) # to select some samples for the plot

# For heatmap
p2 <- ggplot(z4, aes(x = Group.1, y = reorder(Modification, Percent), fill = Percent))+
  geom_tile()+geom_text(aes(label=ifelse((Percent > 0), round(Percent, digits = 0), ifelse(Percent < 0, "", " "))))+
  geom_tile(data = subset(z4, Percent == 0), fill = "ivory")+
  scale_fill_viridis_c(option = "magma", direction = -1, limits= c(0,100)) + #if it doesn't reach 100 % as in Christian Feller's 2015
data
  scale_y_discrete(limits = rev(y_positions))+
  theme(axis.text.x = element_text(angle =90, hjust = 1))+
  coord_fixed() + #squared tiles
```

Appendix

```
  xlab("RNAi Feller et al., 2015") +
  ylab("H4 Acetylation")

p3        <-        ggplot(data=subset(dada_p2,        !is.na(Modification)),        aes(factor(fct_relevel(dada_p2$niceName,
sort(unique(dada_p2$niceName), decreasing = T))), mean_percent, fill=substr(dada_p2$niceName,1,20))) +
  geom_bar(stat="identity", width = 0.80) + labs(x = "RNAi, Feller et al., 2015", y = "Percent of total H4")+
  scale_fill_grey()+
  coord_flip()+
  scale_x_discrete(guide = guide_axis(angle = 0), labels = abbreviate)+#, limits = unique(niceName)[!grepl("90m",
unique(niceName))]) +
  facet_wrap(~ Modification, scales="free", nrow = 3) +
  theme(panel.background     =     element_rect(fill    ="white",     color    =    "black"),legend.position="none",
strip.background=element_rect(colour="white",fill="white"))+
  geom_errorbar(aes(ymin=mean_percent-sem_percent, ymax=mean_percent+sem_percent), width=.2)

pdf(paste(OutputName, ".pdf", sep =""), width=4.0,height=7.0)
plot(p2, label_size = 48)
dev.off();
return(list(p2, p3))
}
```