Structural Studies on the Dosage Compensation Complex from *Drosophila melanogaster*
# TABLE OF CONTENT

1. **INTRODUCTION**  
   1.1 Dosage Compensation  
   1.2 MSL Complex (Male Specific Lethal complex)  
   1.3 MSL Complex factors  
   1.4 Two fold up-regulation achieved by acetylation of histone tail (H4K16)  
   1.5 Targeting of the MSL complex to the X chromosome  
   1.5.1 Proposed mechanism for spreading of the MSL complex along the X chromosome  
   1.6 MLE Protein  
   1.7 Aim and objective of the project  

2. **MATERIALS AND METHODS**  
   2.1 Materials  
   2.1.2 DNA and Protein modifying enzymes  
   2.2 Molecular biology methods  
   2.2.1 Cloning of individual domains as well as full length protein  
   2.2.2 Amplification of desired product by polymerase chain reaction  
   2.2.3 Purification of PCR products by Qiagen purification kit  
   2.2.4 Restriction enzyme digestion of plasmid as well as desired PCR product  
   2.2.5 Amplified products were analyzed by agarose gel electrophoresis  
   2.2.6 Purification of restriction enzyme digested PCR products and plasmid DNA from agarose gels using Qiagen gel extraction kit  
   2.2.7 Ligation of DNA fragments  
   2.2.8 DNA sequencing  
   2.2.9 roX2 stem loop RNA in vitro production for binding studies  
   2.2.9.1 Production of RNA  
   2.2.9.2 Test-transcription of the plasmid DNA inserts  
   2.2.9.3 RNA purification  
   2.3 Microbiology methods  
   2.3.1 Transformation of E. coli cells with ligation mixtures  
   2.3.2 Colony screening  
   2.3.4 Plasmid DNA purification by Qiagen Miniprep Kit
2.3.5 Large scale expression of recombinant proteins
2.4 Recombinant protein purification and analysis
2.4.1 Preparation of cell lysate for protein purification
2.4.2 Immobilized metal affinity chromatography column for
purification of recombinant protein contains 6X His tag
2.4.3 Dialysis of Protein from Ni-chelating Sepharose column
2.4.4 Heparin Column for Nucleic acid removal
2.4.5 Gel filtration/Size exclusion chromatography
2.5 Baculo virus expression for full length MLE Protein
2.5.1 Blue/white screening
2.6 Limited proteolysis
2.7 Western blotting
2.8 Electrophoretic Mobility Shift Assay
2.9 Thermo Fluor Stability Assay
2.10 Isothermal Titration Calorimetry
2.11 Crystallization trials
2.11.1 Concentration of proteins
2.11.2 Setting up crystallization drops
3. RESULTS
3.1 MLE expression and purification
3.2 Purification of the construct containing two double stranded RNA binding domains
3.3 Expression and purification of constructs 2 (dsRBD1) and 3 (dsRBD2) with
His-GST in pETM30 vector
3.4 roX2 stem loop RNA preparation by *in-vitro* transcription
3.4.1 MLE binding of roX2 RNA
3.4.2 rox2 RNA binding by individual dsRBDs of MLE
3.4.3 Isothermal calorimetric (ITC) results
3.5 Mutation study
3.5.1 Double stranded RNA binding domains
3.5.2 Multiple sequence alignment of MLE dsRBDs
3.5.3 MLE mutagenesis
3.5.3.1 First round of mutagenesis 42
3.5.3.2 Second round of mutagenesis 45
3.6 MSL2 – a key player in Dosage Compensation process 47
3.7 Preliminary crystallization trials 51
4. DISCUSSION 53
5. ACKNOWLEDGEMENTS 54
6. REFERENCES 55
7. ABBREVIATIONS 62
1. INTRODUCTION

1.1 Dosage Compensation

The genetic control of sex determination is often associated with dimorphic sex chromosomes. In the XY system, females are homogametic (XX), whereas males are heterogametic (XY) (Charlesworth, 1996). While Y chromosomes are largely heterochromatic and contain few genes, the unequal distribution of X chromosomes would normally result in dramatic changes in gene dosage that could lead to developmental defects or death. Dosage compensation is a process that equalizes gene expression from the X chromosome between males and females (Straub and Becker, 2007). Furthermore, the equal distribution of gene expression between the X chromosome and the autosomes is ensured. The process of dosage compensation has been best studied in worms, flies and mammals, revealing three distinct strategies for equalizing the gene expression of X-chromosomal genes (Laverty, et al., 2010).

In *Drosophila*, dosage compensation (Figure: 1.1) is achieved by a twofold up-regulation of gene expression on the single X chromosome mediated by a dosage compensation complex (DCC). It has been shown that this complex is spread all over the *Drosophila* male X chromosome. As a consequence the expression level is equalized between both sexes and the balance between X chromosomes and autosomes is restored (Gelbart, et al., 2009).

By contrast, in worms and mammals dosage compensation is achieved by controlling the gene expression in the homogametic sex (XX). In mammals the balance in X-chromosomal gene expression is achieved by inactivation of one of the female X chromosomes. This inactivation is promoted by a 17 kb non-coding Xist RNA, which coats the inactive X chromosome to initiate silencing (Clemson, et al., 1996). Finally, in *Caenorhabditis elegans* dosage compensation takes place in homogametic hermaphrodite (XX) by the two fold down-regulation of the two X chromosomes (Amrein, 2000).
Figure: 1.1 Schematic diagrams of dosage compensation mechanisms showing how a non-compensated system is balanced by various mechanisms. In worms and mammals X-linked gene expression in the homogametic sex (XX) is twofold down-regulated by inactivating one X chromosome (mammals) or reducing the expression of both chromosomes by half (worms). Grey box in mammals stands for inactivation, whereas dark blue box indicates activation of the chromosomes. Flies use a male-specific mechanism, leading to a twofold up-regulation of the X-chromosomal gene expression. Female expression is kept unchanged. [Figure modified from Laverty, et al., 2010]

1.2 MSL Complex (Male Specific Lethal complex)

In *Drosophila* dosage compensation is mediated by a large ribonucleic protein complex called dosage compensation complex (DCC) or male specific lethal (MSL) complex (Figure: 1.2). This complex consists of five proteins MSL1, MSL2, MSL3, MOF (males absent on first) and MLE (maleless) and two non-coding RNAs, roX1 and roX2 (RNA on X chromosome) (Straub and Becker, 2007). These complexes spread to hundreds of discrete sites on male X chromosome (Lucchesi, et al., 2005). The key event in dosage compensation is acetylation of histone H4 at lysine 16 (H4K16) carried out by the histone acetyltransferase (HAT) (Smith, et al., 2000; Akhtar and Becker, 2001).
Figure: 1.2 Schematic diagram of the MSL complex in *Drosophila*. MSL1 interact with MSL2 with its RING finger domain. Chromo barrel domain (CHB) in MOF interacts with RNA. HAT domain in MOF acetylates histone 4 on lysine 16. Two RNA binding domains (RB1 and RB2) as well as the glycine rich repeat of MLE interact with RNA. MSL3 and MOF interact with the PEHE domain of MSL1 via MRG domain and ZnF domain, respectively. CRD domain in MSL3 interacts with nucleosome. (Green coil represents DNA in nucleosome, dark blue boxes represent histone proteins, and red coil stands for roX RNAs. Red ball represents acetylation site (H4K16), red flag represents tri-methylated lysine (H3K36me3). Light green indicates MSL2 protein, sky blue indicates MOF protein, yellow represents MLE protein, pink represents MSL3 protein, and lavender represents MSL1 protein) [Figure modified from Hallacli and Akhtar, 2009].

1.3 MSL complex factors

MSL1 protein consists of 1039 amino acid residues. The N-terminal residues of MSL1 are involved in chromatin entry site (CES) recognition (Scott, *et al.*, 2000 and Li, *et al.*, 2005). MSL1 is stabilized by interaction with MSL2 via its N-terminal coiled coil domain. MSL1 also interacts with MSL3 and MOF via a PEHE domain at its C-terminus. MSL2 protein has 773 amino acid residues, which interact with MSL1 via RING finger domain at its N-terminus. MSL2 is expressed only in *Drosophila* males. The DNA binding of the cysteine rich (CXC) domain is evolutionary conserved in human MSL2 ortholog (Lyman, *et al.*, 1997). CXC domain is involved in the proper X-chromosomal targeting (Fauth, *et al.*, 2010).

MLE is an ATPase and RNA/DNA helicase with multiple RNA-interaction domains, including its N-terminal double-stranded RNA-binding domains and its C-terminal glycine-rich region. It has been shown that the helicase activity of MLE is required for the spreading along the X chromosome (Lee, *et al.*, 1997; Izzo *et al.*, 2008). The roX RNAs are non-coding RNA molecules necessary for the *Drosophila* dosage compensation. The two non-coding RNAs (roX1 and roX2) are functionally redundant despite their different size and primary sequence.
All *Drosophila* MSL proteins have their counterparts in human; orthologs for the roX RNA have not been identified. Compared to the *Drosophila* MSL proteins, the human orthologs are shorter due to the lack of unstructured linkers between the stable domains. The domain architecture is essentially the same as in the *Drosophila* proteins. MLE has the human ortholog of RNA helicaseA (hRHA), which is involved in the loading of small interfering RNAs into RISC (RNA-induced silencing complex) (Robb and Rana, 2007).

### 1.4 Two fold up-regulation achieved by acetylation of histone tail (H4K16)

Dosage compensation is believed to result from a global change in chromatin structure. At the molecular level the male X chromosome is marked by acetylation of H4K16 at sites where dosage compensation complexes bind to the chromosome (Turner, *et al.*, 1992). This specific chromatin modification correlates with increased transcriptional level of X-linked genes and has been proposed to be the key event for dosage compensation. However, it remains unclear how the final twofold up-regulation of X-chromosomal genes is achieved. Recently, it has been proposed that the MSL complex promotes progression and processivity of RNAP II across the bodies of active X-linked genes (Larschan, *et al.*, 2011).

### 1.5 Targeting of the MSL complex to the X-chromosome

The main question regarding the recruitment of the MSL complex to the X chromosome is how the MSL complex can distinguish X chromosomes from autosomes. In *Drosophila* female’s dosage compensation is prevented by the repression of MSL2. It has been shown that dosage compensation takes place in females when MSL2 is expressed ectopically. Thus *Drosophila* females carry all the information necessary for MSL targeting (Kelley, *et al.*, 1995), indicating that specific sequence elements are associated with the X chromosome that distinguishes it from other chromosomes.

A sub-complex of MSL1 and MSL2 was detected at a set of about 35-70 sites on the X chromosome in absence of the other MSL proteins (Palmer, *et al.*, 1994; Lyman, *et al.*, 1997), suggesting that there are high-affinity MSL binding sites at the X-chromosome. These high-affinity sites have been identified as chromatin entry sites (CESs) (Lyman, *et al.*, 1997). Using high resolution techniques (ChIP-on-chip) about 150 CESs were found distributed along the *Drosophila* X-chromosome. A 21 bp GA-rich motif was identified called MSL recognition element (MRE), present in 91 % of all CESs (Aleksyenko, *et al.*, 2008). From these chromatin entry sites the MSL complex spreads over the X-chromosome. The exact mechanism of targeting is not known. Thus it is likely that other factors might be required for the CES binding (Gelbart,*et al.*, 2009). Also the role of the roX RNAs that are incorporated in the MSL complex co-transcriptionally is not understood. Interestingly, some of CESs are present within the roX RNAs genes (Ilik, *et al.*, 2009). It remains unclear how the X-chromosome targeting works in mammals, where no corresponding non-coding RNA components have been identified.
1.5.1 Proposed mechanism for spreading of the MSL complex along the X chromosome

After the binding of DCC to the chromatin entry sites the complex is spread along the X chromosome. The identification of over 150 CESs (Alekseyenko, et al., 2008) and the finding that the majority of MSL complex bound to the X chromosome is within 5 - 10 kb of a CES (Sural, et al., 2008) suggest that spreading is a local phenomenon. There are two models for the spreading of MSL complex along the X chromosome, an affinities model where DNA elements play a specific role in recruiting the MSL complex (Gilfillan, et al., 2007; Buscaino, et al., 2006) and a transcription model postulating that the MSL complex localizes along the X chromosome by recognizing actively transcribed genes (Figure: 1.5.1).

Several recent studies support rather the transcription model. It has been shown that the MSL complex is located to actively transcribed genes on the X chromosome suggesting that the MSL complex is spreading from CESs to sites of active transcription (Alekseyenko, et al., 2006). The MSL complex can recognize certain patterns of active genes, e.g. H3K36me3, a histone modification carried out by SET2 (K-methyltransferase 3) (Larschan, et al., 2007; Bell, et al., 2008). The N-terminal chromo barrel of MSL3 is required for the interaction with H3K36me3 modified nucleosomes in vivo (Sural, et al., 2008). When the chromo barrel is mutated the spreading beyond 1 kb from CESs is disrupted in vivo suggesting that MSL3 promotes spreading through the recognition of H3K35me3 (Figure: 1.5.1). It remains unclear whether there are more features of active genes besides the H3K36me3 that can be recognized by the MSL complex.
1.5.1 Two-step model for MSL complex targeting. (A) The MSL complex assembles at the roX genes co-transcriptionally (blue) and recognizes chromatin entry sites (CESs, red) that are located exclusively at the X chromosomes. (B) The MSL complex recognizes features of actively transcribed genes, such as H3K36me3 (stars). There also might be putative sequence elements on the X-chromosome recognized by MSL complex (green lines). Spreading of the MSL complex along the X-chromosome seems to be a local event, the mechanism of spreading remains unclear; it is possible that the MSL complex scans along the X-chromosome searching for active genes. The MSL complex skips inactive genes (Figure modified from Gelbart, et al., 2009).

1.6 MLE Protein

MLE consists of 1293 amino acid residues. It contains seven short segments which are defining the super family of DNA/RNA helicases. MLE is an ATP-dependent helicase enzyme and involved in the melting of DNA/RNA substrates. Deletion and mutation (loss of function) in ATPase domain affects the helicase activity and spreading which leads to the male specific lethality due to their failure in dosage compensation process. ATPase activity is sufficient for transcriptional activation but helicase activity is important for spreading (Izzo, et al., 2008; Kuroda, et al., 1991; Lucchesi, et al., 2000). MLE has a potential to bind RNA as it contains several RNA binding domains, including two dsRBD domain in its N terminus. The N-terminal as well as C-terminal glycine rich repeat were shown to bind RNA in vitro (Izzo, et al., 2008).
Recently, it has been found that MLE is an ortholog of RNA helicase A (RHA) in vertebrates, which is involved in the loading of small interfering RNAs into RISC (RNA-induced silencing complex) (Robb and Rana, 2007). RHA recruits the pre-initiation complex via minimal transactivation domain (MTAD). The amino acid sequence of MTAD is conserved in MLE. In vitro binding study of MLE revealed the function of MLE protein in recruitment of RNA polymerase II via predicted MTAD (325-376) in Drosophila MLE protein (Aratani, et al., 2008).

1.7 Aim and objective of the project

The mechanism of dosage compensation is an excellent model for studying chromatin modification-mediated transcription regulation and the role of non-coding RNA in regulation of gene expression. This complex promotes acetylation of lysine 16 at histone H4 on the X chromosome resulting in two fold up-regulation of most of the X-linked genes. Details of this process still remain unclear, in particular how these complexes are targeted and spread along the male X chromosome and what is the role of individual proteins and RNAs in the assembly.

Particularly in this project, the main focus was on MLE protein. MLE is one of the components of the MSL complex. In this project, the role of MLE and the role of roX RNAs within the dosage compensation mechanism was analyzed. The aim of the project was to produce recombinant fragments of Drosophila MLE protein (full length as well as individual domains) and purify sufficient amount of protein for crystallization. Purified proteins were further studied for RNA binding with roX2 RNA. We prepared specific structure-based MLE mutants that are unable to bind RNA which will now be tested in transgenic flies to better understand the role of RNA binding by MLE in the dosage compensation.
2. MATERIALS AND METHODS

2.1 Materials:

I have used *E-coli* as a host for individual domains of MLE protein expression. Full length MLE protein was expressed in baculovirus expression system. The bacterial strains used are shown in Table: 2.1.1 (a) and the expression vectors in Table: 2.1.1 (b). Maps of the vectors are shown in Figure: 2.1.1.

Table: 2.1.1 (a) Bacterial strains and Vectors

<table>
<thead>
<tr>
<th>Bacterial strains</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia Coli</em> Top10</td>
<td>Host for sub-cloning</td>
</tr>
<tr>
<td><em>Escherichia Coli</em> DH10 Bac</td>
<td>Host for Baculovirus recombination</td>
</tr>
<tr>
<td><em>Escherichia Coli</em> BL21 (DE3) Star</td>
<td>Host for over-expression of proteins</td>
</tr>
</tbody>
</table>

Table: 2.1.1 (b) Expression vectors

<table>
<thead>
<tr>
<th>Vectors</th>
<th>size</th>
</tr>
</thead>
<tbody>
<tr>
<td>pPROEX HTb</td>
<td>4779 bp</td>
</tr>
<tr>
<td>pETM-30</td>
<td>6346 bp</td>
</tr>
<tr>
<td>pFASTBAC HTb</td>
<td>4856 bp</td>
</tr>
<tr>
<td>pUC18</td>
<td>2686 bp</td>
</tr>
</tbody>
</table>
Figure: 2.1.1 (a) pPROEX HTb vector map. The vector contains ampicillin resistance gene, Trc promoter and multiple cloning sites (MCS). The multiple cloning sites are present in between Hind III and Ehel restriction sites. The desired gene product was cloned using the Ncol and Xbal restriction sites.
Figure: 2.1.1(b) pETM-30 vector map. The vector contains kanamycin resistance gene, T7 promoter and multiple cloning sites (MCS). The multiple cloning sites are present in between KpnI and XhoI restriction sites. The desired gene product was cloned using the NcoI and XhoI restriction sites.
Figure: 2.1.1 (c) pFASTBAC HTb vector map. The vector contains ampicillin resistance gene and multiple cloning sites (MCS). The multiple cloning sites are present in between EcoR I and Hind III restriction sites. The desired gene product was cloned using the NcoI and XbaI restriction sites. The sequence for the 6x histidine affinity tag, spacer region and rTEV protease cleavage sites are underlined.
Figure: 2.1.1 (d) pUC18 vector map. The vector contains ampicillin resistance gene, multiple cloning site (MCS). The multiple cloning sites are present in between EcoRI and HindIII restriction sites. The cloning/expression region of the coding strand is shown blue-marked.
2.1.2 DNA and Protein modifying enzymes

The enzymes and instruments used for the cloning, expression and purification purpose are shown in Table 2.1.2 (a) and the analytical instruments used are shown in Table 2.1.2 (b).

Table: 2.1.2 (a) Enzymes used for DNA manipulation

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pyrobest DNA Polymerase</td>
<td>Takara Bio Clontech, California, US</td>
</tr>
<tr>
<td>Pfu Turbo DNA Polymerase</td>
<td>Stratagene, California, US</td>
</tr>
<tr>
<td>Restriction enzymes (NcoI, XhoI, XbaI, DpnI, EcoRI)</td>
<td>New England Bio labs, Ipswich, Massachusetts, US</td>
</tr>
<tr>
<td>Ligation Mix</td>
<td>Takara Bio Clontech, California, US</td>
</tr>
<tr>
<td>Protease Inhibitor cocktail</td>
<td>Roche Diagnostics, Gmbh, Germany</td>
</tr>
<tr>
<td>Trypsin and Chymotrypsin</td>
<td>Bio chemicals, Mannheim Gmbh, Germany</td>
</tr>
</tbody>
</table>

Table: 2.1.2 (b) Analytical instruments used

<table>
<thead>
<tr>
<th>Analytical use</th>
<th>Device/Instrument</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein purification</td>
<td>Akta purifier (RT) Heparine HiTrap™ 5ml Superdex™ 75 10/300 GL Superdex™ 200 10/300 GL Superdex™ 75 16/60 GL</td>
<td>GE Healthcare Life Sciences, Uppsala, Sweden</td>
</tr>
<tr>
<td>Centrifugation</td>
<td>Beckman Coulter Avanti J-26xP AvantiJ-20xP Eppendorf Centrifuge 5804 R</td>
<td>Beckman Coulter, California, US</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Eppendorf Centrifuge, Hamburg, Germany</td>
</tr>
<tr>
<td>Process</td>
<td>Equipment/Software</td>
<td>Supplier/Location</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>---------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Concentrating Protein</td>
<td>Amicon ® Ultra</td>
<td>Millipore , Ireland</td>
</tr>
<tr>
<td>Polymerase chain reaction</td>
<td>Biometra Thermo cycler</td>
<td>Biometra GmbH, Gottingen, Germany</td>
</tr>
<tr>
<td>Ultra sonication</td>
<td>Misonix Ultrasonic liquid processors</td>
<td>Misonix Processors, Newtown, CT, US</td>
</tr>
<tr>
<td>Incubator Shaker</td>
<td>Infors HT Multitron</td>
<td>Infors HT , Allschwill, Switzerland</td>
</tr>
<tr>
<td>Spectrophotometer</td>
<td>Nanodrop ND-1000</td>
<td>Thermo scientific, Wilmington, US</td>
</tr>
<tr>
<td>Analysis of agarose gels</td>
<td>UviTec</td>
<td>Gel documentation, Cambridge, UK</td>
</tr>
<tr>
<td>Isothermal calorimeter</td>
<td>Microcal 200</td>
<td>GE healthcare life sciences, Uppsala, Sweden</td>
</tr>
<tr>
<td>Thermo-flour stability assay</td>
<td>MX 3005 P</td>
<td>Stratagene, California, US</td>
</tr>
<tr>
<td>Setting crystallization drops</td>
<td>Cartesian pixsys 4200</td>
<td>Genomic solutions, UK</td>
</tr>
<tr>
<td>Crystallization screens and optimizer</td>
<td>Tecan Evo 200/8</td>
<td>Tecan , Hombrechtikon, Switzerland</td>
</tr>
<tr>
<td>Crystals observation</td>
<td>Formulatix Rockimager</td>
<td>Formulatix, US</td>
</tr>
<tr>
<td>Cristallisation plates</td>
<td>Hampton plates</td>
<td>Hampton Research, California, US</td>
</tr>
<tr>
<td></td>
<td>Qiagen plates</td>
<td>Qiagen, California, US</td>
</tr>
</tbody>
</table>
2.2 Molecular biology methods

2.2.1 Cloning of individual domains as well as full length protein

The constructs for MLE protein expression are shown in Table 2.2.1 (a) and the constructs for CXC domain of MSL-2 protein are shown in Table: 2.1.1 (b).

Table: 2.2.1 (a) Different constructs of MLE protein

<table>
<thead>
<tr>
<th>Construct</th>
<th>Length(aa)</th>
<th>Tag used</th>
<th>Domain</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-86</td>
<td>N term His tag</td>
<td>1st dsRBD</td>
<td>Double stranded RNA binding domain (dsRBD)</td>
</tr>
<tr>
<td>2</td>
<td>1-92</td>
<td>N term His tag</td>
<td>1st dsRBD</td>
<td>Extended version of construct 1</td>
</tr>
<tr>
<td>3</td>
<td>151-254</td>
<td>N term His tag</td>
<td>2nd dsRBD</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1-254</td>
<td>N term His tag</td>
<td>Two dsRBDs</td>
<td>Two domains together</td>
</tr>
<tr>
<td>5</td>
<td>1-1293</td>
<td>N term His tag</td>
<td>Full length</td>
<td>Contains all domains of MLE Protein</td>
</tr>
<tr>
<td>6</td>
<td>321-1293</td>
<td>N term His tag</td>
<td>Truncated full length</td>
<td>Construct without dsRBDs</td>
</tr>
<tr>
<td>7</td>
<td>1-1125</td>
<td>N term His tag</td>
<td>Truncated full length</td>
<td>Construct without G-rich repeats</td>
</tr>
<tr>
<td>8</td>
<td>1-1153</td>
<td>N term His tag</td>
<td>Truncated full length</td>
<td>Extended version of construct without G-rich repeats</td>
</tr>
<tr>
<td>9</td>
<td>321-1125</td>
<td>N term His tag</td>
<td>ATPase and RNA/DNA helicase domain</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>321-1153</td>
<td>N term His tag</td>
<td>ATPase and RNA/DNA helicase domain</td>
<td>Extended version of construct 9</td>
</tr>
<tr>
<td>11</td>
<td>987-1125</td>
<td>N term His tag</td>
<td>OB fold(oligosaccharide binding) domain</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>987-1153</td>
<td>N term His tag</td>
<td>OB fold(oligosaccharide binding) domain</td>
<td>Extended version of construct 11</td>
</tr>
<tr>
<td>13</td>
<td>1011-1125</td>
<td>N term His tag</td>
<td>OB fold(oligosaccharide binding) domain</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1011-1153</td>
<td>N term His tag</td>
<td>OB fold(oligosaccharide binding) domain</td>
<td>Extended version of construct 13</td>
</tr>
</tbody>
</table>

Table: 2.2.1 (b) Different constructs for CXC domain of male specific lethal-2 protein

Based on the limited proteolysis result, two different constructs were created for CXC domain of MSL-2. Both domains vary by their amino acid length.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Length(aa)</th>
<th>Tag used</th>
<th>Domain</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>440-566</td>
<td>N term His tag</td>
<td>CXC domain</td>
<td>DNA binding domain in MSL2 protein</td>
</tr>
<tr>
<td>2</td>
<td>440-590</td>
<td>N term His tag</td>
<td>CXC domain</td>
<td>Extended version of construct 1</td>
</tr>
</tbody>
</table>
2.2.2 Amplification of desired product by polymerase chain reaction

Polymerase chain reaction is an important technique in molecular biology field which is useful for the amplification of desired piece of DNA template. This technique is also known as *in-vitro* replication machinery. DNA template is amplified by the desired primers (forward and reverse primers). The primers were designed to introduce two restriction sites (for *Nco*I and *Xba*I restriction endonucleases) to the 5’ and 3’ end of the gene, respectively. Constructs 2 and 3 were cloned into the pETM30 vector using restriction enzymes *Nco*I and *Xho*I. Insoluble full length constructs (5 and 6) were cloned into pFASTBAC HTb vector using *Nco*I and *Xba*I restriction sites. The PCR reaction mixtures used for amplification are shown in Table: 2.2.2 (a) and the program used for the amplification is shown in Table: 2.2.2 (b).

Table: 2.2.2 (a) PCR reaction mixtures used to amplify the constructs

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X Pyrobest buffer II (KCl, 1M Tris-HCl, pH 9.0, MgCl₂, Triton-X 100)</td>
<td>5 μl</td>
</tr>
<tr>
<td>dNTPs 2.5 mM</td>
<td>5 μl</td>
</tr>
<tr>
<td>DNA template</td>
<td>0.5 μl (should be about 10 ng)</td>
</tr>
<tr>
<td>Forward primer (10 μM)</td>
<td>2 μl</td>
</tr>
<tr>
<td>Reverse primer (10 μM)</td>
<td>2 μl</td>
</tr>
<tr>
<td>Pyrobest DNA polymerase (125U)</td>
<td>0.2-0.5 μl</td>
</tr>
<tr>
<td>Water</td>
<td>35 μl</td>
</tr>
<tr>
<td><strong>Total reaction volume</strong></td>
<td><strong>50 μl</strong></td>
</tr>
</tbody>
</table>

Table: 2.2.2 (b) PCR program used to amplify the gene product

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial pre-heating</td>
<td>95°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>Annealing</td>
<td>55-60°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>1 minute per kb (repeat steps 2-4, 25-30 times)</td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>10 minutes</td>
</tr>
</tbody>
</table>

2.2.3 Purification of PCR products by Qiagen purification kit

Amplified PCR products were purified using Qiagen PCR purification kit. Based on the instructions in the manual, amplified PCR products were mixed with 5 volumes of PBI buffer (binding buffer: 5 M Gu-HCl, 30% isopropanol) and loaded into the spin column. The QIAquick spin column was centrifuged for 30-60 seconds and the flow through discarded. Then, the column was washed with 750 μl of PE buffer (wash buffer: 10 mM Tris-HCl pH 7.5, 80% ethanol), centrifuged for 60 seconds and the flow-through was discarded. The column was centrifuged for an additional minute to remove the residual ethanol from the membrane. The spin column was placed in a clean sterile eppendorf tube and DNA was eluted with 30 μl of sterile distilled water by centrifuging for 1 min. All centrifugation steps were carried out at 25000 x g in an Eppendorf 5417R tabletop centrifuge.
2.2.4 Restriction enzyme digestion of plasmid as well as desired PCR product

Amplified DNA fragments, which has the desired restriction sites at 5’ end as well as 3’ end were digested by appropriate restriction endonuclease enzyme at 37°C for 2-3 hours to create compatible cohesive ends with the vector, while the expression vectors also digested with the same restriction endonuclease enzyme pair for 2 hours at 37°C. (The amount of restriction enzymes can be adjusted according to the concentration of the enzymes; the enzymes are stored in 50% glycerol). Glycerol concentration in reaction mixture should not be over 5%; otherwise the enzymes might have a star activity. E.g. not more than 3 μl of enzymes in a 30 μl reaction).

The Protocols used for digestion are shown in Table: 2.2.4 (a) and (b).

Example of restriction digestion for pProEx HTb vector

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Volume (μl)</th>
<th>Ingredient</th>
<th>Volume (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEB Buffer 4</td>
<td>3.6 μl</td>
<td>NEB Buffer 4</td>
<td>4 μl</td>
</tr>
<tr>
<td>DNA</td>
<td>30 μl</td>
<td>DNA</td>
<td>2-4 μg (e.g. 20 μl)</td>
</tr>
<tr>
<td>XbaI</td>
<td>1 μl</td>
<td>XbaI</td>
<td>2 μl</td>
</tr>
<tr>
<td>NcoI</td>
<td>1 μl</td>
<td>NcoI</td>
<td>2 μl</td>
</tr>
<tr>
<td>BSA</td>
<td>0.36 μl</td>
<td>BSA</td>
<td>0.4 μl</td>
</tr>
<tr>
<td>Total</td>
<td>36 μl</td>
<td>Water up to</td>
<td>40 μl</td>
</tr>
</tbody>
</table>

2.2.5 Amplified products were analyzed by agarose gel electrophoresis

Agarose gel electrophoresis is commonly used in biochemistry as well as in molecular biology experiments. This technique is useful for the separation of both DNA as well as RNA based on their length. Using markers, we can easily identify the size of the fragments which are separated by this technique. The amplified DNA fragments migrate towards positive side due to their negative charge, when electric field is applied. Agarose gel acts as a sieve, which is composed of galactose and 3, 6-anhydrogalactose. It traps the DNA fragments based on their size. The smaller fragments move faster than the larger fragments. Agarose gel’s resolving power depends on the pore size of the matrix, which depends on the concentration of agarose added in the buffer 1X TAE (40 mM Tris acetate, 1 mM EDTA). In the process of agarose gel electrophoresis, 1% agarose gel was firstly prepared in 1X TAE buffer and then added appropriate amount of SYBR safe stain (Invitrogen). After solidification of gel, the casting tray was kept into the electrophoretic tank containing appropriate amount of 1X TAE buffer. The amplified PCR products with appropriate 6X gel loading buffer (0.09% bromophenol blue, 0.09% xylene cyanol, 60% glycerol, 60 mM EDTA) were loaded in the well. The electrophoresis was carried out at 80-120 volts. After gel electrophoresis, the gel was visualized using UviTec trans illuminator.
2.2.6 Purification of restriction enzyme digested PCR products and plasmid DNA from agarose gels using Qiagen gel extraction kit

Restriction enzyme digested PCR products and plasmid DNA separated by agarose gel electrophoresis were purified using the gel extraction kit (Qiagen, Hilden, Germany) according to the manual. Sliced fragments from agarose gel, which contain desired size of the fragments were cut and weighed. After that, QG buffer (6 M guanidine thiocyanate, 50 mM Tris-HCl pH 7.5, 20 mM EDTA pH 8.0) was added to the gel (e.g. 300 μl per 100 mg of gel). The tube was incubated at 60 °C for 5-10 min. After the gel slice had dissolved completely, the sample was loaded onto Qiagen quick spin column and centrifuged for 60 s. Spin columns were washed once again with 500 μl QG buffer and with 700 μl wash buffer and centrifuged for 60 seconds. The flow through was discarded and the spin columns were centrifuged once again for one minute to remove traces of the buffer. Then, columns were placed into a clean eppendorf tubes and DNA was eluted with 30 μl of distilled water. All these steps were carried out in an Eppendorf 5417R table top centrifuge at 25000 x g.

2.2.7 Ligation of DNA fragments

Formation of phosphodiester bonds between 5’ phosphor group and 3’ hydroxyl group is achieved by ligation reaction. This process needs ATP and is catalyzed by DNA ligase enzyme. We used the Takara ligase mighty mix, which already contains buffer as well as ATP. The tubes were incubated at 16°C for 30 min. The ligation reaction mixture is shown in Table: 2.2.7.

Table: 2.2.7 Ligation reaction mixture

<table>
<thead>
<tr>
<th>Ingredients</th>
<th>Volume (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digested Insert</td>
<td>4 μl</td>
</tr>
<tr>
<td>Digested Vector</td>
<td>1 μl</td>
</tr>
<tr>
<td>Takara mighty mix</td>
<td>5 μl</td>
</tr>
</tbody>
</table>

2.2.8 DNA sequencing

DNA sequencing was carried out in Cogenics (Beckman Coulter Genomics, Takeley, UK). The sequence was verified after sequencing using L-Align software. Plasmids containing desired sequences were used further for transformation process.

2.2.9 roX 2 stem loop RNA in vitro production for binding studies

2.2.9.1 Production of RNA
The strategy used for the production of large amounts of RNA includes the following steps

- Design of DNA template from RNA sequences
- Cloning of DNA template to plasmids
- Production of plasmid DNA in *E. coli* cells
- Linearization of the plasmid with corresponding restriction enzyme
- T7 RNA transcription from the template
- 15% analytical polyacrylamide gel (PAGE) containing 8 M urea by stained with methylene blue.

2.2.9.2 Test-transcription of the plasmid DNA inserts

The plasmid DNA, containing genes of roX2 stem loop RNA, was digested with 1 unit/μg DNA of EcoRI restriction enzyme (NEB: New England Biolabs). Digestion was performed at 37 °C for 4 h in buffers recommended by respective enzyme manufacturers. Purification of the linearized DNA was performed by phenol-chloroform extraction. The DNA solution was distributed in eppendorf tubes up to 200 μl. Two hundred μl of phenol/chlorophorm (25:24) were added in each tube. The tubes were then vortexed and centrifuged for 1 min at 20000 x g. The supernatant was transferred into a new tubes and the procedure repeated. In the last round of extraction, 200 μl of chloroform were added to the supernatant. After centrifugation, the supernatant was newly distributed into tubes; 1/10 volume of Na acetate (pH 5.0) and 3 volumes of ice-cold ethanol (95% EtOH) were added into each tube; the tubes mixed gently and stored 20 min at -20 °C. Thereafter, precipitated DNA was spun at 4 °C and 48000 x g; the supernatant was carefully removed with vacuum pump and a glass capillary tip. The DNA was air-dried and solubilized in milliQ water. The transcription reaction was carried out for 3-4 h at 37 °C in following reaction conditions:

<table>
<thead>
<tr>
<th>Volume (μl)</th>
<th>Ingredients</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 μl</td>
<td>10x buffer</td>
</tr>
<tr>
<td></td>
<td>(400 mM Tris/HCl pH=8.0 at 25 °C; 50 mM DTT; 10 mM spermidine; 0.1% Triton X100; 300 mM MgCl₂)</td>
</tr>
<tr>
<td>20 μl</td>
<td>PEG 8000(40% stock)</td>
</tr>
<tr>
<td>10 μl</td>
<td>4 mM of each nucleotide triphosphate (NTP)</td>
</tr>
<tr>
<td>5 μl</td>
<td>10 mg/ml T7 RNA polymerase</td>
</tr>
<tr>
<td>10 μl</td>
<td>5 μg of linear DNA</td>
</tr>
</tbody>
</table>

Beginning from the phenol-chloroform extraction, all procedures were performed in RNase-free conditions, which means using of special area for RNA work only, treating of surfaces of benches and glassware with RNase inactivating agents and 100% ethanol, wearing gloves and using of sterile, RNase-free plastic ware. The transcribed RNA was visualized on 15 % analytical polyacrylamide gel (PAGE) containing 8 M urea by staining with methylene blue.
2.2.9.3 RNA purification

Large scale production of RNA for further binding and crystallization experiments was performed using the procedure below: Total volume: 5 ml

<table>
<thead>
<tr>
<th>Volume</th>
<th>Ingredients</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 ml</td>
<td>10x buffer (400 mM Tris/HCl pH=8.0 at 25 °C; 50 mM DTT; 10 mM spermidine; 0.1% Triton X100; 300 mM MgCl₂)</td>
</tr>
<tr>
<td>1 ml</td>
<td>PEG 8000(40% stock)</td>
</tr>
<tr>
<td>250 μl</td>
<td>each 4 mM of each nucleotide triphosphate (NTP)</td>
</tr>
<tr>
<td>50 μl</td>
<td>10 mg/ml T7 RNA polymerase</td>
</tr>
<tr>
<td>200 μl</td>
<td>linear DNA (4.5 mg)</td>
</tr>
</tbody>
</table>

Reaction was performed under 37 °C for 3 hours, after that 4 M Urea and 50 mM EDTA were added for gel filtration purpose. The products of 5 ml transcription reaction were loaded on large gel filtration column S75 at 1.5 ml/min speed maintained at 82 psi pressure. Working buffer contained 4 M urea, 20 mM Tris pH 7.5 and 5 mM EDTA. Pool fractions were dialyzed against fresh milliQ water. Precipitation was carried out by 1:1000 (Glycogen, 20 mg/ml); 1/10 volume of 3 M Na acetate (pH 5.0) and 3 volumes of ice-cold ethanol (95% EtOH) were added and spin down for 30 min, after that washed with 70% EtOH.

2.3 Microbiology methods

2.3.1 Transformation of E. coli cells with ligation mixtures

Ligation mixtures of pPROEX HTb, pETM30, pFAST BAC recombinant plasmids were transformed into the E. coli competent cells (TOP10). 50 μl of competent cells were thawed on ice for 20-30 min and 1-2μl of ligation mixture was added and the cells were incubated on ice for 30 min. After the cells were treated with heat shock at 42 °C for 30 s and then transferred to ice bucket for 2 min, and then 250 μl LB medium (without antibiotics) was added and the tubes were incubated at 37 °C for 45-60 min. 200 μl of the cell suspensions was spread onto the LB plates containing 100 μg/ml ampicillin or 50 μg/ml Kanamycin based on the vector map and these plates were incubated overnight at 37 °C incubator.

2.3.2 Colony screening

From the LB_{Amp} or LB_{Kan} plates, colonies of E. coli TOP-10 cells containing pPROEX HTb, pETM30, pFAST BAC recombinant plasmids were selected; 2 ml overnight culture of each screened colony (usually 4 colonies per transformation reaction) was grown by incubating on a rotary shaker at 37 °C. Cells from the overnight culture were lysed and the recombinant plasmids were purified using the QIAPrep Miniprep Kit (Qiagen) according to the supplier recommendations, and purified recombinant plasmids were screened for having the correct insert using appropriate restriction endonucleases. After digestion with these enzymes digestion mixtures were analyzed by agarose gel electrophoresis. Plasmids containing an insert with correct size were submitted for DNA sequencing and plasmids with the wild type sequence were selected and used for transformation into the BL-21 cells for protein expression.
2.3.3 Plasmid DNA purification by Qiagen Miniprep Kit

Plasmid DNA was purified using the Qiagen miniprep kit. The purification was performed based on instructions in the manual. A single colony from the host \textit{E.coli} cell containing the plasmid of interest was inoculated into the 2 ml of LB medium which contains the appropriate antibiotic in adequate amount and cultured overnight at 37°C. In a clean and sterile 1.5 ml eppendorf tube the bacterial cells were pelleted using an Eppendorf 5417R table top centrifuge. The bacterial cell pellets were resuspended in 250 μl of buffer P1(resuspension buffer: 50 mM Tris-Cl pH 8.0, 10 mM EDTA, 100 ug/ml RNase A) followed by 250 μl of buffer P2 (lysis buffer: 200 mM NaOH, 1% SDS) and mixed by gently inverting the tubes 4-6 times. An additional 350 μl of buffer N3 (neutralization buffer: 3 M potassium acetate pH 5.5) was added and the sample in the tube gently mixed resulting in the formation of cloudy white precipitate. The tubes were then placed in the table top centrifuge, and centrifuged at 25000 x g for 10 minutes.

The cloudy white precipitate pellet was discarded and the supernatant was applied to Qiagen spin column and centrifuged for 1 minute. The column was washed with 750 μl of buffer PE and centrifuged for 1 minute. The flow through liquid collected in the tube below was discarded and further centrifuged for additional 1 minute to remove all traces of buffer PE. The Qiagen spin column was then placed in a clean and sterile 1.5 ml eppendorf tube and 30 μl of distilled water was added. The column was allowed to stand for 1 minute. The plasmid was eluted and collected in the 1.5 ml eppendorf tubes by centrifugation for 1 minute.

2.3.4 Large scale expression of recombinant proteins

\textit{BL21} plates containing recombinant plasmids were inoculated into the desired volume of LB medium containing appropriate antibiotics. The cultures were inoculated at 37°C until it reached the OD value of 0.6. After reaching 0.6, the cell cultures were then induced with 0.1M IPTG and the temperature was reduced to 23 °C and incubated overnight. The next morning the cell cultures were harvested using the JLA-8.1 rotor at 14000 x g at 4 °C for 20 minutes. The supernatant was discarded and the pellets were resuspended in appropriate column buffer and stored at -20 °C for further use.

2.4 Recombinant protein purification and analysis

2.4.1 Preparation of cell lysate for protein purification

Cells were harvested at a speed of 14000 x g at 4 °C for 20 min and the cell pellet was resuspended in 40 ml of lysis buffer containing 200 mM Tris, 20 mM NaCl, 5 mM 2-mercaptoethanol to lyse the cells. One protease inhibitor cocktail tablet (Roche Diagnostics, Mannheim, Germany) was added to the lysis buffer to prevent protein degradation. After that the solution was ultrasonicated with 80 W for 2 min on ice to enhance the lysis efficiency. The cell lysate was ultracentrifuged at a 60000 x g for 30 min at 4 °C using a rotor 25.5 ultracentrifuge to remove cellular debris.
2.4.2 Immobilized metal affinity chromatography column for purification of recombinant protein contains 6X His tag

IMAC column is an efficient purification technique based on their binding affinity with nickel ions. This is the best method for recombinant proteins containing polyhistidines His-tags. The ÄKTA purifier FPLC™ system (GE Healthcare) was used for the purification of proteins. The supernatant from the crude lysate was loaded onto a prepacked 20 ml Ni- HisTrap™ column (GE Healthcare).

The column was washed with washing buffer, containing 200 mM Tris, 20 mM NaCl, 5 mM 2-mercaptoethanol, after that column was washed with 1M salt buffer (1M NaCl) to remove nucleic acids from the proteins. Elution was carried out with the increased concentration of imidazole (from 20 mM to 500 mM). The elution fractions were analyzed by Bio-Rad protein assay as well as 15% SDS-PAGE gel. The purified proteins were cleaved further by TEV cleavage enzyme (ration of 1:100) for overnight, to remove the His tag from recombinant protein.

2.4.3 Dialysis of Protein from Ni-chelating Sepharose column

The cleaved protein sample from first Ni-chelating Sepharose column was dialyzed against the washing buffer (200 mM Tris, 20 mM NaCl, 5 mM 2-mercaptoethanol) to eliminate the high imidazole content in the protein and to exchange the buffer components suitable for the next purification step. Collected fractions were placed in the dialysis membrane (Spectra/por® Dialysis membrane 6-8000 MW) which was dialyzed against 2 liters of washing buffer (200 mM Tris, 20 mM Nacl, 5 mM 2-mercaptoethanol) for 5-6 hours. The buffer was exchanged two times during the dialysis step. After that the cleaved protein was purified by second run on the nickel chelating Sepharose column.

2.4.4 Heparin Column for Nucleic acid removal

Heparin Sepharose column chromatography is an important technique to remove nucleic acid binding proteins, growth factors and steroid receptors. The collected fractions from the nickel column were loaded onto a HiTrap™ heparin 5 ml column and equilibrated with gel filtration buffer (200 mM Tris, 10 mM NaCl, 5 mM 2-mercaptoethanol). The column was run at a flow rate of 5 ml/min. The proteins were eluted by the increased concentration (gradient elution) of salt in buffer (10 mM NaCl -1 M NaCl). The purified fractions were analyzed by 15% SDS-PAGE gel.

2.4.5 Gel filtration/Size exclusion chromatography

Gel filtration chromatography columns separate molecules based on their molecular weight. Larger molecules are eluted earlier and smaller molecules are retained for longer period. Pure fractions of the 2nd nickel column were collected and concentrated. The concentrated protein solutions were loaded onto a Superdex™ 75 or 200 columns, these columns were equilibrated with gel filtration buffer (200 mM Tris, 20 mM NaCl, 5 mM 2-mercaptoethanol). The column was run at a flow rate of 0.5 ml/min. The purity of the eluted fractions was analyzed by 15% SDS-PAGE gel and the purest fractions were collected together for crystallization experiments as well as for a RNA binding study.
2.5 Baculo virus expression for full length MLE Protein

2.5.1 Blue/white screening

Full length as well as truncated versions (construct 5 and 6) of MLE protein were expressed using Bac-to-Bac® Baculovirus Expression System (Invitrogen). Recombinant baculovirus generation was carried out in two steps. The genes of the constructs (5 and 6) were first cloned into a small transfer vector pFastBac™ and propagated in E. coli. The recombinant plasmid was then transformed into E. coli DH10Bac™ competent cells which contained a baculovirus shuttle vector bMON14272. This bacmid carried a mini-attTn7 attachment site from the transfer vector as well as a kanamycin resistance marker and lacZα gene.

The mini-Tn7 element on the pFastBac™ can transpose to the mini-attTn7 target site on the bacmid in the presence of transposition protein provided by the helper plasmid (Figure: 2.5.1). If so, the reading frame of the lacZα gene, where the attachment site for the bacterial transposon is inserted, gets disrupted. Such lacZα deletion causes formation of white colonies (Lac-) in the presence of a chromogenic substrate Blue-gal and the inducer IPTG. Expression of lacZα peptide by cells with unaltered bacmid leads to formation of blue colonies.

DNA prepared from selected E.coli clones containing the recombinant bacmid was then used to transfect insect cells. Luria Agar plates were prepared containing 50 μg/ml kanamycin, 7 μg/ml gentamicin, 10 μg/ml tetracycline, 100 μg/ml Blue-gal and 40 μg/ml IPTG. The DH10Bac™ competent cells were thawed on ice. One hundred μl of the cells were dispensed into 15-ml round-bottom polypropylene tubes. Approximately 1 ng of recombinant donor plasmid was added. The mixture was incubated on ice for 30 min, further shock heated by transferring to 42°C water bath for 45 sec and cooled down on ice for 2 min. 900 μl of LB medium were added to the mixture. The vials were placed in a shaking incubator at 37 °C with medium agitation (225 rpm) for overnight. The cells were then serially diluted 1:0, 1:10 and 1:100, using LB medium, and spread evenly over the surface of the plates and incubated for 24 to 48 h at 37°C.

Figure: 2.5.1 Generation of recombinant baculoviruses and gene expression with the Bac-to-Bac® Expression System
2.6 Limited proteolysis

Limited proteolysis is the technique to find out the stable subdomains from the protein of interest using the desired protease enzyme. The digested fragments were analyzed by SDS-PAGE gels. Normally we can use proteases like chymotrypsin, trypsin, and proteinase K for digestion purpose. First of all the protein of interest was diluted using the appropriate buffer system to obtain the desired volume. A protease enzyme (1:1000) was added to this solution to start the reaction, and then aliquots of the reaction samples were removed at a specific time intervals (5, 10, 30, and 60 minutes). The removed aliquots were boiled to inactivate the protease enzyme activity and then the samples were analyzed by SDS-PAGE gels.

2.7 Western blotting

After the samples were resolved by SDS-PAGE, the plates were removed and the stacking gel was cut from the gel. The resolving gel was transferred to a container and soaked into the 1x Transfer buffer (25 mM Tris, 192 mM glycine, 10% methanol). During this time, four pieces of Whatman 3 mm chromatography paper having same dimensions as the resolving gel were cut and a piece of reinforced nitrocellulose membrane of same size and cut on one corner to mark orientation of the gel was prepared. Two 3 mm paper sections were soaked in transfer buffer and placed on the semi-dry transfer blotting apparatus. A 10 ml plastic pipette was used to roll out the air bubbles from underneath. The gel was stacked on the nitrocellulose membrane, followed by two more sheets of soaked 3 mm paper.

The transfer was run for 3 hours at 150 mA. The stack was slowly disassembled and the membrane removed. The transfer of the pre-stained marker will serve to indicate if the transfer worked properly or not. The membrane was blocked for 1 hour in 5% milk solution in PBS (1.4 mM KH₂PO₄, 8 mM Na₂HPO₄, 140 mM NaCl, 2.7 mM KCl, pH 7.3), washed 3 times for 5 min in PBS containing 0.05% Tween-20. The membrane was then incubated for 1 hour and slight shaking with 0.2 μg/ml of primary anti-His-tag mouse monoclonal antibodies diluted in a 5% milk/PBS solution. The membrane was further washed and incubated with 0.1 μg/ml of the secondary HRP-conjugated anti-mouse antibodies diluted in a 5% milk/PBS solution. The membrane was washed one more time and signals were detected.

2.8 Electrophoretic Mobility Shift Assay

The EMSA is based on the ability of a DNA or RNA binding protein to retard the mobility of its specific DNA substrate during electrophoresis on a non-denaturating polyacrylamide gel or agarose gel. EMSAs provide a simple method for assessing the binding affinity and specificity of a given DNA or RNA interaction. Using an EMSA, several different substrates can be assessed for suitability for use in a co-crystallization trial with a DNA or RNA binding protein. The mobility of protein-nucleic acid complex through a non-denaturating gel is only slightly affected by change in the size of the DNA fragment. Mobility is primarily determined by the size and charge of the protein and by conformation of the protein-nucleic acid complex. We were used 1% agarose gel for double stranded RNA-Protein binding study.
2.9 Thermo Fluor Stability Assay

Thermo Fluor stability assay useful for the optimization of experimental buffers by screening. A protein is screened against a matrix of buffer and ligands/additives using a thermal denaturation assay to identify rapidly conditions, additives and ligands offering improved solubility, reduced aggregation and, in some cases, improved crystallization. The assay is performed in a 96 well format. The sample plate contains the protein, various buffers, additives and fluorescent dye. SYPRO orange is a commercially available dye, which is interacts with the hydrophobic molecules and give the fluorescence. Due to denaturation process, hydrophobic core residues become exposed upon unfolding. A real time PCR machine is helpful for the gradual heating of samples. We used this method to identify the melting point value for mutant and wild type proteins. Based on the result, we decided the best mutants for further study.

2.10 Isothermal Titration Calorimeter

Isothermal titration calorimeter (MicroCal iTC200) unit directly measures the heat evolved or absorbed in liquid samples as a result of mixing precise amounts of reactants. We can measure the thermodynamic parameters from this instrument (Figure: 2.10). A spinning syringe is used for injecting and mixing of reactants. Spin rates are user selectable; the usual range is 500 to 1000 rpm. The normal temperature operating range is 20°C to 80°C.

![Figure 2.10 Isothermal titration calorimetry setup](image)

Sample and reference cells are accessible for filling and cleaning through the top of the unit. During the experiment, ligand is spread (titrated) into the sample cell based on the reaction (heat may be taken up or released) and we can measure the temperature differences. Based on the availability of the ligand, we can get a peak which corresponds to the binding; from this graph we can measure the binding constants ($K_B$), reaction stoichiometry ($n$), enthalpy ($\Delta H$) and entropy ($\Delta S$) values in a single experiment.
2.11 Crystallization trials

2.11.1 Concentration of proteins

Purified protein samples from gel filtration were concentrated by centrifugation using centrifuge 5804R with the cutoff value of 10 kDa at 5000 x g to reach a concentration of approximately 10-15 mg/ml.

2.11.2 Setting up crystallization drops

Crystallization trials were carried out for MLE subunits (two dsRBDDs, dsRBDI, dsRBDDII) as well as MSL2 subunits (CXC 566, CXC 590). Crystallization drops were set up by Cartesian pixsys 4200 robot using Hampton and Qiagen screens. The sitting drops vapor diffusion technique was used for initial crystallization drops.

The reservoir in the plate was filled with mother liquor containing precipitant, buffer or salt in higher concentration. The crystallization plates were automatically created by the Cartesian pixsys 4200 robot.

Fundamental concept in this procedure is the evaporation of water from the protein sample and its transfer to the reservoir solvent by vapor diffusion. This will increase the concentration of precipitant in the small drop. At concentration above the saturation point, protein molecule starts to accumulate and form the crystals. Based on the sitting drop result, we can optimize a few crystallizations conditions using hanging drop method. Apart from the orientation of droplet there is no much more difference in the fundamental concept of vapor diffusion. The reservoir contained 500 μl of mother liquor; from this liquor 1 μl was transferred to a cover slip which was mixed with 1 μl of 15-20 mg/ml protein solution. The cover slip was inverted over the reservoir to maintain a closed system. The water vaporized from the protein droplet was absorbed by the reservoir. The protein crystallized when the precipitant reached an optimum level of saturation.

Microcrystals were observed for MLE subunit (dsRBDD, dsRBDDII) under some conditions with solutions containing lithium sulfate as well as ammonium sulfate.
3. RESULTS

3.1 MLE expression and purification:

The main aim of this project was to produce soluble MLE protein or its individual domains in sufficient amounts for RNA binding and crystallography study. Constructs for this project work were designed based on the sequence alignments with human and yeast MLE orthologs whose structures are known. Fourteen constructs (Figure: 3.1(a)) were designed coding for essentially all the individual domains as well as the full length MLE protein.

![Diagram of constructs](image)

**Constructs:**

1(1-86) →
2(1-92) →
3(151-254) →
4(1-254) →
5(1-1293) →
6(321-1293) →
7(1-1125) →
8(1-1153) →
9(321-1125) →
10(321-1153) →
11(987-1125) →
12(987-1153) →
13(1011-1125) →
14(1011-1153) →

Figure: 3.1(a) The constructs used for MLE project work
All the fourteen fragments were amplified using PCR with the respective primers and cloned into the pProEx-Htb vector to generate fusion proteins containing an N-terminal His-tag and a TEV cleavage site. The cloned constructs were transformed into BL21 (DE3) Star *E.coli* cells for over-expression using the protocols which are available in microbiology methods (Section 2.3). The cell extracts were collected and the presence of the MLE fragments was monitored by immuno-blotting with an antibody that recognizes the His-tag of the recombinant proteins. Western blot results show a band for construct 4 (30 kDa). Unfortunately, for constructs 5, 6 and 9 we observed multiple bands of smaller size, indicating possible degradation during expression. No expression was visible for all the other constructs (Figure: 3.1(c)).
Since crystallization trials can now be performed with low quantity of protein, we wanted to verify that these constructs are not expressed at least in low amounts. Thus, we performed large scale expression test for all the 14 constructs. For this purpose proteins from 1L E. coli culture were purified using Ni-chelating Sepharose column. SDS-PAGE analysis of the 500 mM elution fractions of all constructs (Figure: 3.1(d)) clearly showed that the desired product can only be obtained for construct 4 (two double stranded RNA binding domains, 30 kDa), confirming the result of western blot. The lack of soluble expression and degradation of the other constructs might be caused by not correct construct boundaries and flexible regions present between individual domains. E. coli might also not be an appropriate host for expression of these MLE constructs.

3.2 Purification of the construct containing two double stranded RNA binding domains

The MLE construct 4 (254 aa) containing two N-terminal double stranded RNA binding domains was expressed in E. coli BL21 at 23°C after induction with IPTG and purified by His-tag affinity chromatography using a Ni-chelating Sepharose column. After His-tag removal with TEV protease the protein was further purified by another passage through the Ni column.
In order to remove nucleic acid contaminations, the protein was loaded onto a heparin column. The last part of the purification was gel filtration using a Superdex 75 column (Figure: 3.2 (b)). All fractions of the 11 ml peak were collected and concentrated to 5 mg/ml. The pictorial representation of two double stranded RNA binding domains are below.

Figure: 3.2 (a) Left: SDS-PAGE of the fractions from the Ni column and Heparin column (E- elution from first Ni column, AC- after TEV cleavage, AD- after dialysis, FT- flow-through, 20 mM imidazole wash, 500 mM imidazole wash, F1- peak1 from heparin column, F2- peak 2 from heparin column). Right: SDS-PAGE of fractions from the peaks of the Superdex 75 gel filtration (F1- F6).

Figure: 3.2 (b) Chromatogram showing the elution profile from the Superdex 75 column by monitoring the absorbance at 280 nm and 260 nm for the MLE (construct 4).
3.3 Expression and purification of constructs 2 (dsRBD1) and 3 (dsRBD2) with His-GST in pETM30 vector

The constructs coding for the first and second dsRBD (construct 2 and 3) cloned into pPROEX Htb vector did not yield soluble protein. In order to increase the solubility, these fragments were re-cloned into pETM30 vector to generate their His-GST-tag fusions. The MLE constructs 2 and 3 containing individual double stranded RNA binding domains were expressed in *E.coli* BL21 at 23°C after induction with IPTG and purified by His-tag affinity chromatography using a Ni-chelating Sepharose column. After His-tag removal with TEV protease the protein was further passage through the 2nd Ni column. The proteins were further purified by gel filtration using Superdex 75 column (Figure: 3.3 (b) & (c)).

![Figure 3.3 (a) Left: SDS-PAGE of the fractions from the first Ni column. Elution was achieved in 500 mM imidazole wash [BI- before induction, Lys- lysate, S- supernatant, FT- flow-through, saltW-1 M NaCl wash, 500 mM imidazole elution]. Right: SDS-PAGE of the fractions from the second Ni column, after dialysis and TEV cleavage to remove His-GST from the protein. Lower band indicates the desired protein of interest [AC- after TEV cleavage, AD- after dialysis, FT- flow-through, 20 mM imidazole wash, 100 mM imidazole wash, 500 mM imidazole wash].](image)

![Figure 3.3 (b) Chromatogram showing the Superdex 75 column elution profile of purified dsRBD I (construct 2).](image)
3.4 roX2 stem loop RNA preparation by *in-vitro* transcription

The 72 nucleotide long roX2 stem loop was cloned into the pUC18 vector and the plasmid DNA was amplified by Mega prep (Qiagen). After purification of the plasmid DNA, the vector was linearized with EcoRI restriction endonuclease enzyme. An *in vitro* transcription reaction was performed as described in Table: 2.2.9 with a reaction volume of 5 ml. The quality of the produced RNA was analyzed with a 15% analytical polyacrylamide gel (PAGE) containing 8 M urea and stained with methylene blue (Figure: 3.4 (a)). The roX2 RNA was purified with a large gel filtration column (S75). Working buffer contained 4 M urea, 20 mM Tris pH 7.5 and 5 mM EDTA (Figure: 3.4(b)). Finally I have obtained the concentration of 4.5 mg/ml RNA.
Figure: 3.4 (a) Left: Secondary structure prediction of roX2 stem loop by MFOLD; dG value for this prediction is -23.60. Right: 72 nucleotide roX2 RNA verified by analytical gel, the band of the roX2 RNA verified with the marker. 8 M urea gel was used for this purpose.

Figure: 3.4 (b) The gel filtration chromatogram showing the elution profile of the roX2 RNA, middle peak represents the roX2 RNA (75ml of RNA elution).
3.4.1 MLE binding of roX2 RNA

The purified RNA was dialyzed twice in a cassette against fresh milliQ water. To form secondary stem loop structure the RNA was heated up to 95°C and cooled down to 60°C and then 20 mM MgCl$_2$ was added. RNA binding study was performed by modified EMSA (Electrophoretic Mobility Shift Assay). The EMSA is based on the ability of a DNA or RNA binding protein to retard the mobility of its specific nucleic acid substrate during electrophoresis on a non-denaturating polyacrylamide gel or agarose gel.

The reaction mixture was prepared mixing 0.2 mM-1.6 mM protein with 0.2 mM RNA in 20 mM Tris, 200 mM NaCl containing buffer and incubated for 15 min at 20°C. First, the construct 4 (1-254) was mixed with roX2 stem loop RNA and resolved using 1% agarose gel. Gel analysis with UV illuminator (for RNA) and coomassie staining (for protein) revealed clear band shift, indicating the interaction between MLE and roX2 RNA (Figure: 3.4.1).

The dsRNBs are considered to be sequence non-specific RNA binders (Tian, et al., 2004). However, ADAR2 dsRBD was shown to bind dsRNA in a sequence specific manner (Stefl, et al., 2010). To test the sequence specificity of MLE dsRBD we tested whether MLE interacts also with another unrelated dsRNA, the Alu RNA of signal recognition particle (Brooks, et al., 2009). Figure: 3.4.1 shows that MLE (1-254) binds equally efficiently also the Alu RNA, indicating that the RNA binding by MLE is sequence independent.

![Figure: 3.4.1 MLE binding study with different RNA by EMSA, Left: RNA and protein staining of construct4. The lane 1 represents RNA alone; 2 and 3 shows the increased concentration of protein and RNA. (a) RNA staining: RNA binding of roX2 and ALU stem loop RNAs under UV illumination. Right: Protein staining: protein migration was verified after protein staining for both RNAs (rox2 and ALU).]

3.4.2 rox2 RNA binding by individual dsRBDs of MLE

Next, we tested whether the individual dsRBD domains can bind roX RNA. The EMSA results (Figure: 3.4.2) show that both dsRBD1 and 2 (construct 2 and 3) do bind RNA in a way comparable to the one of MLE (1-254). This was rather unexpected as Izzo et al. (2008) reported that the first dsRBD of MLE did not bind RNA.
3.4.3 Isothermal calorimetric (ITC) results

We measured the affinities of two dsRBD construct (MLE 1-254), dsRBD I and dsRBD II for the 72 nt roX2 RNA using isothermal calorimetry (ITC). ITC experiments were performed by titration of the protein solution into the roX2 RNA solution and a clear binding was observed for all three constructs (Figure: 3.4.3). As the binding seems to be sequence independent and we do not know how many binding sites are present on the 72 nt RNA construct or whether these bind RNA in the same way, it was not possible to choose a proper binding model for the data analysis and obtain an exact dissociation constant ($K_d$) value.

We used this experiment to obtain a general idea about the MLE RNA binding affinity. The approximate $K_d$ values of double dsRBDs (construct4), dsRBD I (construct 2) and dsRBD II (construct 3) were 7.1 µM, 22 µM and 40 µM, respectively. From this result we concluded that both individual domains can bind RNA. The affinity towards the roX2 RNA is further increased when both domains are present in one construct.
Figure: 3.4.3 (a) ITC results of construct 4 (two dsRBDs), dissociation constant (Kd) value: 7.1 µM, stoichiometry factor value: -3.4 (b) ITC results of construct 3 (dsRBDII), dissociation constant (Kd) value: 22 µM, stoichiometry factor value: -2.2 (c) ITC results of construct 2 (dsRBDI), dissociation constant (Kd) value: 40 µM, stoichiometry factor value: -2.6
3.5 Mutation study of MLE protein

The role of the MLE dsRBDs in dosage compensation is not clear. Thus, we wanted to prepare MLE mutants that could not bind RNA with their dsRBD and test the impact of such mutants on the dosage compensation in transgenic flies.

3.5.1. Double stranded RNA binding domains

The double stranded RNA binding domains adopt a α1-β1-β2-β3-α2 topology. All dsRBMD contain three regions making contact with the target RNA. Region 1 is present in the helix α1, region 2 present in the loop (between β strands β1 and β2) and region 3 in the loop between β2 and the α2 helix. Regions 1 and 2 interact with the minor groove of the dsRNA via 2’ hydroxyl groups (Figure: 3.5.1 (a)), while region 3 interacts with the dsRNA via phosphate groups of the major groove (Tian, et al., 2004). All double stranded RNA binding motifs and their conserved residues are represented in the sequence logo (Figure: 3.5.1 (b)).

Figure: 3.5.1 (a) Schematic drawing of dsRBD of *Xenopus laevis* (1DI2) with dsRNA. Hydroxyl groups are represented as red balls and phosphate groups are shown in green.

Figure: 3.5.1 (b) Sequence logo of dsRBM from 1,428 species
3.5.2 Multiple sequence alignment of MLE dsRBDs

The multiple sequence alignment was performed by ESPript software (Figure: 3.5.2). We have found out the conserved residues for mutagenesis experiments. The promising residues were identified based on the superimposition of structures which is homolog to MLE protein.

Figure: 3.5.2 Multiple sequence alignment of MLE dsRBDs. Multiple sequence alignment of the first 275 amino acids of Drosophila, Tribolium, Nasonia, Apis, Bombyx, Anopheles, Canis, Xenopus, Bos, homo and Equus. dsRBD I and dsRBD II are shaded in the alignment. The amino acids are displayed in their single letter code. Conserved residues were pointed out in red and mutated residues were pointed out in green boxes. Secondary structures of the human protein is shown in blue.
### 3.5.3 MLE mutagenesis

#### 3.5.3.1 First round of mutagenesis

The structures of the individual dsRBDs of human RNA helicase are known. We compared them with the known structures of dsRBDs in complex with RNA and found that the *Xenopus laevis* RNA binding protein A (Xlrbpa) structure is the most similar. We superimposed the Xlrbpa onto the RNA helicase A structures, and identified several RHA residues that correspond to the RNA binding residues in Xlrbpa. We also verified whether these residues are well conserved among species and in MLE. Since we wanted to abolish the RNA binding simultaneously in both dsRBD domains, we decided to mutate some of the selected residues into similar residues, which would just disrupt the local hydrogen bond network in the interface without affecting solubility of the protein. We generated a first set of mutants - KNQ, KNQHR and KNQHRNKK. The residues mutated in the first round of mutagenesis are shown in Table: 3.5.3.1 and the interacting residues are shown in Figure: 3.5.3.1(a).

<table>
<thead>
<tr>
<th>Mutant name</th>
<th>Domains</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNQ</td>
<td>dsRBD I &amp; dsRBD II</td>
<td>K4E, N175E, Q179D</td>
</tr>
<tr>
<td>KNQHR</td>
<td>dsRBD I &amp; dsRBD II</td>
<td>K4E, N175E, Q179D, H196E, R198E</td>
</tr>
<tr>
<td>KNQHRNKK</td>
<td>dsRBD I &amp; dsRBD II</td>
<td>K4E, N175E, Q179D, H196E, R198E, N52D, K53E, K54E</td>
</tr>
</tbody>
</table>

The KNQHRNKK mutant was not stable and precipitated suggesting that the introduced mutation affected the overall structure of this construct. We tested the RNA binding of the KNQ and KNQHR mutants in the band shift assay (Figure: 3.5.3.1 (b)) which showed that the KNQ mutations were not sufficient to abolish the RNA binding while the KNQHR mutant did not bind RNA anymore. We also tested the RNA binding of the KNQHR using ITC (Figure: 3.5.3.1 (c)). As expected no roX2 binding was observed for this mutant.
We noticed that the gel filtration peak of Superdex 75 column of the KNQHR was broader and eluting earlier than the one of the wild type protein (9.5 ml compared to 12 ml), suggesting possible aggregation problems (Figure: 3.5.3.1 (d)).

<table>
<thead>
<tr>
<th>wild type</th>
<th>KNQ</th>
<th>KNQHR</th>
</tr>
</thead>
</table>

Figure: 3.5.3.1 (b) RNA binding study of wildtype and mutants by EMSA. Lanes in both gels: 1- RNA alone; 5, 9 and 13- Protein alone; remaining lanes- increased concentration of wild type, KNQ and KNQHR proteins (0.2 mM – 2.4 mM) Left: RNA staining: RNA binding of wild type and mutant (KNQ, KNQHR) proteins verified with roX 2 RNA under UV illumination. Right: Protein staining: protein migration was verified after coomassie staining.

Figure: 3.5.3.1(c) left: Isothermal calorimetric result of wild type protein (1-254), dissociation constant ($K_d$) value: 7.1 µM, stoichiometry factor: 3.43. Right: Isothermal calorimetric result of the KNQHR mutant.
To compare stability of this mutant and the wild type protein we performed limited proteolysis with trypsin (Figure: 3.5.3.1 (e)). The observed pattern of tryptic fragments seems comparable, indicating that the mutant is not less stable than wild type. Finally, we analyzed the KNQHR mutant stability using the thermal shift assay (Figure: 3.5.3.1 (f)), which revealed it to be significantly less resistant to increasing temperature than the wild type protein.

**Wild type** | **KNQHR**
---|---
0’ | 0’
5’ | 5’
20’ | 20’
60’ | 60’
M | M

Figure: 3.5.3.1 (e) Limited proteolysis result of wild type and mutant (KNQHR) protein, SDS-PAGE gels shows the samples from samples from 0’, 5’, 20’, 60’ (minutes) for this study.
Figure: 3.5.3.1 (f) The values of fluorescence intensity vs temperature were plotted, the middle point of the experimental values represent the $T_m$ value. Thermo Fluor stability (TSA) result comparison of wild type and mutant (KNQHR) proteins, Wild type ($T_m$) value is 48°C but KNQHR ($T_m$) value is 41°C.

### 3.5.3.2 Second round of mutagenesis

To ensure that the introduced mutations do not affect the overall stability of the protein we prepared another set of mutants with fewer mutations – KNQH, KHR and KH. The residues mutated in the second round of mutagenesis are shown in Table: 3.5.3.2

Table: 3.5.3.2 Second round of mutants in human dsRBP

<table>
<thead>
<tr>
<th>Mutant name</th>
<th>Domains</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNQH</td>
<td>dsRBD I &amp; dsRBD II</td>
<td>K4E, N175E, Q179D, H196E</td>
</tr>
<tr>
<td>KHR</td>
<td>dsRBD I &amp; dsRBD II</td>
<td>K4E, H196E, R198E</td>
</tr>
<tr>
<td>KH</td>
<td>dsRBD I &amp; dsRBD II</td>
<td>K4E, H196E</td>
</tr>
</tbody>
</table>

Gel filtration elution profiles (Figure: 3.5.3.2 (b)) of all the new mutants were comparable with the wild type protein. Also the thermal shift assay (Figure: 3.5.3.2 (c)) confirmed that they are not less stable than the wild type. We tested the ability of these mutants to bind roX2 RNA in the band shift assay (Figure: 3.5.3.2 (a)). The RNA binding of KH mutant was reduced and the KNQH and KHR mutants did not bind RNA anymore.
Figure: 3.5.3.2 (a) RNA binding study of wild type and mutants by EMSA. The lanes: 1- RNA alone; lanes 4, 7,10 and 13- Protein alone; remaining lanes- increased concentration of wild type, KNQH, KHR and KH proteins (0.3 mM – 1.8 mM) (Left) RNA staining: RNA binding of wild type and mutant (KNQH, KHR and KH) proteins verified with roX 2 RNA under UV illumination (Right) Protein staining: protein migration was verified after coomassie staining.

Figure: 3.5.3.2 (b) Chromatogram showing the elution profile by monitoring the absorbance at 280 nm for the purified wild type and the mutants (KNQH, KHR, and KH) proteins through Superdex 75 column; elution volume: 11 ml.
Thus the most efficient mutant appears to be the KHR mutant (K4E, H196E, and R198E). The introduced mutations specifically interfere with the RNA binding, while the overall stability of the protein is unaffected. To be able to test the impact of these mutations on dosage compensation in vivo we had to generate corresponding mutations in the full length MLE protein. Using the quick change mutagenesis kit we introduced these mutations into the MLE gene that had been previously cloned into the pCasper4-pw8 vector. This vector is routinely used by our collaborators at MPI in Freiburg to generate transgenic flies that are then subject to chromatin precipitation experiments aimed to analyze the binding of the MSL complex and MLE to the X chromosome.

3.6 MSL2 – a key player in Dosage Compensation process

_Drosophila_ MSL2 fragment covering the CXC domain spanning residues 381-590 was expressed in bacteria as an N-terminal His-tag fusion using the pPROEX-Htb. To identify the exact limits of the CXC domain we performed limited proteolysis with trypsin (Figure: 3.6 (a)). The trypsin resistant fragments were analyzed by mass spectrometry (ESI TOF). The observed molecular mass of 13,672 Da that corresponds to MSL2 fragment 440-566. Thus, we prepared two new constructs - 440-566 and 440-590. The plasmids coding for these two fragments were transformed into _E.coli_ BL21 and purified by His-tag affinity chromatography using a Ni-chelating Sepharose column (Figure: 3.6 (b & c)). After His-tag removal with TEV protease protein was further purified another passage through a Ni column. In order to remove nucleic acid contaminations, the protein was loaded onto a heparin column. The last part of the purification was gel filtration using a Superdex 75 column (Figure: 3.6 (d & e)). All peak fractions were collected and concentrated to 16 mg/ml for crystallization trials.
Figure 3.6 (a) Limited proteolysis result of *Drosophila* MSL2 (473-590 & 381-590) protein, SDS-PAGE gels shows the samples from 0’, 5’, 20’, 60’ (minutes) for this study.

The pictorial representation of full length MSL 2 protein and two constructs of C-terminal cysteine rich CXC domains are given below.
Figure: 3.6 (b) Left: SDS-PAGE of the fractions from the first Ni column, elution was achieved in 500 mM imidazole wash. Right: SDS-PAGE of the fractions from the second Ni column, after dialysis and TEV cleavage to remove His tag from the protein. Lower band indicates the desired protein of interest.

Fig 3.6 (c) Chromatogram showing the elution profile by monitoring the absorbance at 280 nm and 260 nm for the purified CXC 566 domain through Superdex 75 column.
DNA Binding Study of CXC domain

Previous studies showed that the MSL2 domain can recognize GA (guanine-adenine) repeats in the high affinity sites (HAS), and that this recognition is important for MSL1-MSL2 core complex formation (Li, et al., 2008). However, it has also been suggested that DNA binding of MSL2 CXC domain might be sequence independent (Fauth, et al., 2010). The minimum dsDNA length required for MSL2 binding was 20 nt.

We have designed several synthetic DNA sequences which contain GA repeats with various lengths (based on the DBF-12 and CES11D1 high affinity sites (Becker, et al., 2007 and Alekseyenko, et al., 2008). We tested MSL2 (440-590) for dsDNA binding using the EMSA method. First, we tested the interaction with 21-32 nt long dsDNA and could show that all the DNA constructs were bound by MSL2. We could also show that MSL2 can bind even dsDNA lacking the GA repeat sequence (Figure: 3.6 (e)). To define the minimal length of DNA bound by MSL2 we then used DNA fragments of 11-18nt. Figure: 3.6 (e) shows that all the DNAs tested were bound by the MSL2 CXC domain (20-32 nt), even though the binding seems to be getting slightly weaker with decreasing the DNA length (11-18nt) Figure: 3.6 (f).
**Figure: 3.6 (e) DNA binding study of CXC domain. Lanes: 1- protein alone (CXC -590) ; 2, 5, 8, 11, 14, 17 and 20-synthetic DNA alone; remaining lanes with different concentration of protein with DNA (0.2 mM- 2 mM) (DBF12a – 32 nt, DBF12b - 28nt, DBF12c - 24nt, DBF12d- 21nt, CES-1- 21nt, CES-2- 28nt, CES mut- 28nt (mutation in the GA repeat)).**

<table>
<thead>
<tr>
<th>CXC 590</th>
<th>DBF 12a</th>
<th>DBF 12b</th>
<th>DBF 12c</th>
<th>DBF 12d</th>
<th>CES-1</th>
<th>CES-2</th>
<th>CES-mut</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 3 4</td>
<td>5 6 7</td>
<td>8 9 10</td>
<td>11 12 13</td>
<td>14 15 16</td>
<td>17 18 19</td>
<td>20 21 22</td>
</tr>
</tbody>
</table>

**Figure: 3.6 (f) DNA binding study of CXC domain. Lanes: 1- protein alone (CXC -590) ; 2, 5, 8, 11, 14 and 17-synthetic DNA alone; remaining lanes with different concentration of protein with DNA (0.2 mM- 2 mM) (DBF12a - 32 nt, P18- 18nt, P16- 16nt, P12- 12nt, P11-11nt).**

<table>
<thead>
<tr>
<th>CXC 590</th>
<th>DBF 12a</th>
<th>P18</th>
<th>P16</th>
<th>P14</th>
<th>P12</th>
<th>P11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2 3 4</td>
<td>5 6</td>
<td>7</td>
<td>8 9</td>
<td>10</td>
<td>11 12</td>
</tr>
</tbody>
</table>

### 3.7 Preliminary crystallization trials

The MLE protein constructs (construct 4 (1-254), construct 2 (1-92) and construct 3 (151-254) as well as two constructs of CXC domains (CXC 440-566 and CXC 440-590) were concentrated after gel filtration. The concentrated proteins were set up for crystallization trials using 480 different screening conditions from Hampton research and Qiagen screens. The crystallization experiments were regularly controlled for crystal formation. Microcrystal needles were observed for construct 3 (dsRBD II) in two conditions containing 0.05 M magnesium sulfate, 0.05 M Hepes pH 7.0, 1.6 M lithium sulfate and 0.025 M magnesium sulfate, 0.05 M Tris pH 7.0, 1.8 M ammonium sulfate. Unfortunately, we were unable reproduce these crystals using a manual crystal screen set up. For MLE (151-254) we obtained a long crystal in 0.2 magnesium formate (Figure: 3.7 (a)).
The crystal was frozen in mother liquor supplemented with 30% glycerol and tested for diffraction at the ID14EH4 beamline of the European Synchrotron Radiation Facility (ESRF), but only a high resolution diffraction spots typical for salt crystal were observed. No crystals were obtained for the MSL2 constructs.

Figure: 3.7 (a) Salt crystal of dsRBDII. Salt crystal was obtained for construct 3 which has double stranded RNA binding domain II.
4. DISCUSSION:

The role of the MLE protein in dosage compensation is not well established. The structural information would clearly help understand the way MLE functions. First, I tried to express 14 constructs which covered all the domains of MLE protein. Only the double stranded RNA binding domains (dsRBDs) were expressed soluble while other constructs were not expressed. I also tried to express the longer constructs in the insect cells, but I did not succeed to recombine these genes into the baculovirus genomic DNA.

Then, I focused more on the dsRBD domains. I prepared and purified three constructs; construct 2 (1-92), construct 3 (151-254) and construct 4 (1-254). I have performed the ITC and EMSA experiments with the roX2 RNA and I could show that these domains bind dsRNA. Since MLE binds also Alu RNA, the dsRNA recognition by the MLE dsRBDs seem to be sequence independent. This makes the structural studies more complicated because due to their sequence independency, the proteins can bind in various places on RNA. Before starting the crystallography study on the RNA complex, we have planned to identify the minimal RNA fragment sufficient for MLE binding. We have used a 21 nt long portion of stem loop RNA for the binding study but we could not get a conclusive result for this experiment.

We have tried to obtain crystals of the individual domains and found conditions where the second dsRBD can crystallize. These conditions still need to be optimized to obtain well ordered crystals. Nevertheless, these structures are likely to be similar to the ones of the homologous domains of human RNA helicase A (RHA), which are already known. Thus, we have used the known structures of RHA and Xlrbpa-2/RNA complex to generate models for MLE dsRBD binding to the RNA and prepared specific mutants which can abolish the binding.

The stability of these structure based mutants was verified by gel filtration and thermal stability assay (TSA). I have also prepared the mutation in the GST fusions of dsRBDI. The dsRBDII must still be mutated (HR). These mutants in individual domains will be tested for RNA binding. I have cloned these mutations into the full length MLE (in pCasper4- pws vector). I have prepared three mutants K (domain I), HR (domain II) and KHR (domain I & II). The impact of these mutations on dosage compensation will now be tested in Akthar’s group in MPI, Freiburg. The genome wide chromatin immune-precipitation study will be performed which will help us understand the significance of the RNA binding by dsRBDI and dsRBDII for the dosage compensation complex assembly and function.

Meanwhile, I was working on the CXC domain of MSL2. We identified trypsin resistant domain covering the predicted cysteine-rich domain, which was expressed, purified and used for crystallization trials. So far, no crystal was obtained.

Probably, better constructs need to be prepared. As DNA binding by this domain is sequence independent it will be a challenge to identify a suitable dsDNA fragment that can be successfully used in crystallization trials. The eventual structure of MSL2 with dsRNA will be very important for us to understand the interaction between the MSL complex and the nucleosome.
5. ACKNOWLEDGEMENTS

First of all, I would like to give my sincere thanks to Stephen Cusack, who provides me this wonderful opportunity to being part of this world class research environment in EMBL, Grenoble outstation.

My uncountable thanks to my supervisor Jan Kadlec, during this period I have learnt not only scientific techniques from him, learned motivation as well as determination to achieve the goal in my life. I was inspired by his scientific excellence. He helped me lot to understand myself in a better way. He was given lot of support during my tough times during this period. His guidance regarding scientific thinking, scientific writing and presenting science was very helpful for me. I can feel the transformation which he made on me. He was not only act as a supervisor, as a friend, brother as well as well-wisher.

I would like to thank Thomas Lunardi and other group members in my group. Thomas was very helpful and supportive during this period, especially I would like to thank for his music composition in the lab hours. It was very helpful for me during my stressful period.

I would like to thank Lars Liljas for his continuous guidance from my first year to end of my master thesis work. A big thanks to the EMBL administration staffs Mary-Jane Villot and Dominique Lancon, especially I would like to give my special thanks to Dominique to arrange accommodation for me in such a nice French family. This was very helpful for me to understand the French culture in a better way.

In addition, my sincere thanks to Ramesh Pillai and his group members. I was really enjoyed lot with his group members by their scientific as well as funny discussions during lunch hours. They arranged lot of parties and get-together during this period.

Last but not least, I would like to thank my parents, sister and brother-in-law for their continuous support throughout my career. I am very lucky to have such a nice family members. I would like to thank my friends also (Thiyagarajan Gnanasekaran, Sridharan ganesan, Prabhu Babu, Nazer Ghani, Francesca kosia and Anna Adamiok) for their valuable support and advice.
6. REFERENCES


46. Li, F., Parry, D.A.D., & Scott, M.J. The amino-terminal region of Drosophila MSL1 contains basic, glycine-rich, and leucine zipper-like motifs that promote X chromosome binding, self-association, and MSL2 binding, respectively. Molecular and Cellular Biology 25, 8913-8924 (2005).


### 7. ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>absorbance</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium persulfate</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>CBD</td>
<td>chromo barrel domain</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CES</td>
<td>chromatin entry site</td>
</tr>
<tr>
<td>DCC</td>
<td>dosage compensation complex</td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>double distilled water</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>dsRBD</td>
<td>double stranded RNA binding domain</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>ESRF</td>
<td>European Synchrotron Radiation Facility</td>
</tr>
<tr>
<td>HAC</td>
<td>heparin affinity chromatography</td>
</tr>
<tr>
<td>HAT</td>
<td>histone acetyltransferase</td>
</tr>
<tr>
<td>HIC</td>
<td>hydrophobic interaction chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal ion affinity chromatography</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodaltons</td>
</tr>
<tr>
<td>LA</td>
<td>luria agar</td>
</tr>
<tr>
<td>LB</td>
<td>luria bertani</td>
</tr>
<tr>
<td>MLE</td>
<td>maleless</td>
</tr>
<tr>
<td>MOF</td>
<td>males absent on first</td>
</tr>
<tr>
<td>MSL</td>
<td>male specific lethal</td>
</tr>
<tr>
<td>Short Form</td>
<td>Full Form</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>poly acrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>ROX</td>
<td>RNA on X-chromosome</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>TBE</td>
<td>tris borate EDTA</td>
</tr>
<tr>
<td>TE</td>
<td>tris EDTA</td>
</tr>
<tr>
<td>TEMED</td>
<td>tetramethylmethylenediamine</td>
</tr>
</tbody>
</table>