Biochemical and Structural Studies of Dosage Compensation Members: MSL1, MSL3, and MOF from *Drosophila melanogaster*

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Masters of Science

In the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Kent Conrad Klemmer

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or part, for scholarly purposes may be granted by the professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of Saskatchewan in any scholarly use which may be made of any materials in my thesis.

Request for permission to copy or make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5

Abstract

Dosage compensation is the key regulatory process employed in *Drosophila melanogaster* to equalize the level of gene transcripts between the single X chromosome in males (XY) and the two X chromosomes in females (XX). Dimorphic sex chromosomes evolved by the severe degeneration of the Y chromosome, giving rise to an imbalance between the heterogametic sex and the homogametic sex. Vital to the viability of male *Drosophila* is the dosage compensation complex (DCC), a ribonucleoprotein complex that mediates the precise two-fold transcription of the single male X chromosome. The DCC is comprised of five proteins: male-specific-lethal proteins (MSL) 1, 2, and 3, male absent-on-the-first (MOF), maleless (MLE), and two non-coding RNAs. The complex specifically co-localizes along the male X chromosome in a reproducible manner, resulting in acetylation of lysine 16 of the N-terminal tail of histone H4. The exact mechanism of recruitment and spreading of the DCC along the male X chromosome remains unclear; recent studies propose a multi-step mechanism involving DNA sequence elements, epigenetic marks, and transcription. Understanding how dosage compensation functions provides insight into the interplay between gene regulation and chromatin remodelling. The goal of this project was to better understand how *Drosophila* MSL1, MSL3, and MOF interact and how their interaction modulates MOF's acetyltransferase activity. Recombinant protein constructs were cloned and over-expressed in a bacterial expression system permitting future structure determination by X-ray crystallography. The $dMSL1_{820-1039}$ construct consisted of the C-terminal domain, reported to be able to interact with both dMSL3 and dMOF. $dMSL3_{186-512}$ contained the domain required for the interaction with dMSL1 and dMOF. $dMOF_{371-827}$ was comprised of the catalytic domain, the CCHC zinc finger, and the chromodomain, as the N-terminal region does not encode any known domains. All three recombinant proteins were successfully cloned, over-expressed, and purified to homogeneity. Recombinant $dMOF_{371-827}$ was determined to acetylate histones. Interaction studies using GST pull-down assays and size exclusion chromatography determined that $dMSL1_{820-1039}$ and $dMOF_{371-827}$ did not interact above background levels. Moreover, size exclusion chromatography revealed $dMSL3_{186-512}$ and $dMOF_{371-827}$ did not interact nor did the three recombinant proteins form a stable complex.

Acknowledgements

I would like to thank Dr. Stanley Moore for being my supervisor for my master degree, and for providing both laboratory space and resources enabling a constructive experience and training for my future endeavours. I would like to acknowledge my co-workers during the course of my degree; Andrew Welham, Ewa Kerc, Michael Lane, Yunhua Jia, and Yurdagul Ferhatoglu for their help and advice.

I would like to dedicate my thesis to my friends and family. I am grateful to them for being their when needed. To my mom Cheryl thank you for all your help and dedication in making it possible for me to succeed at university. I want to extend my gratitude to Stephanie Worobec for her love and kindness. I want to thank Dean and Kevin Ziegler for all of their assistance and friendship.

Table of Contents

List of Abbreviations

Ac-CoA – acetyl-coA

BLAST – Basic Local Alignment Search Tool

BSA – bovine serum albumin

CES – chromatin entry site

ChIP – chromatin immunoprecipitation

ChIP-chip – chromatin immunoprecipitation hybrized to a DNA microarray

Chromo – chromatin organization modifier

CPM – counts per minute

CV – column volume

DAPI – 4’,6-diamidino-2-phenylindole

DCC- dosage compensation complex

DTT – dithiothreitol

Eaf3p – Esa1p-associated factor-3 protein

EDTA – ethylenediaminetetraacetic acid

EMS – ethyl methanesulfonate

EMSA – electrophoretic mobility shift assay

Esa1 – essential SAS2-related acetyltransferase

FISH – RNA fluorescence *in situ* hybridization

GS – Glutathione Sepharose

GST – Glutathione-S-transferase

HAS – high affinity site

HAT – histone acetyltransferase

HEPES – 4-(2-hydroxyethyl)-1-piperazine ethanesulphonic acid

HMTase – histone N-methyltransferase

HP1 – heterochromatin protien 1

H4K16 – lysine 16 of N-terminal tail of histone 4

H4K16Ac – acetylated lysine 16 of N-terminal tail of histone 4

IPTG – isopropyl-β-D-thiogalactopyranoside

K_d – equilibrium dissociation constant

LB – Luria-Bertani broth

MCS – multiple cloning site

MSL – male specific lethal

MLE – maleless

MLL1 – mixed lineage in leukemia 1

MOF – male absent on the first

MOZ – monocyte leukemia zinc finger protein

MPD – 2-methyl-2, 4-pentanediol

MRE – MSL recognition element

MRG – MORF-related gene

MRG15 – MORF-related gene on chromosome 15

MYST – MOZ, Ybf3/SAS3, SAS2, and Tip60

NCP – nucleosome core particle

NSL – non-specific lethal

OD_{600nm} – optical density measured at 600 nanometres

PBS – phosphate buffered saline

PCR – polymerase chain reaction

PC – polycomb

PDB – Protein Data Bank

PEV – position-effect variegation

PHI-BLAST – Pattern Hit Initiated BLAST

pI – isoelectric point

PSI-BLAST – Position Specific Iterative BLAST

PMSF – phenylmethylsulfonyl fluoride

RB – RNA binding

RING – really interesting new gene

roX – RNA on the X

SAS – something about silencing

SEC – size exclusion chromatography

SDS-PAGE – sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sxl – sex lethal protein

Tip60 – (HIV) Tat interactive protein 60 kDa

Tris-HCl – tris(hydroxymethyl)aminomethane hydrochloride

Xist – X-inactive specific transcript

List of Figures

List of Tables

1. Introduction

The introduction summarizes the literature of *Drosophila melanogaster* dosage compensation and the interplay between dosage compensation, chromatin and nucleosome structures, and epigenetics. Moreover, the review provides insight on how the *Drosophila melanogaster* dosage compensation complex (DCC) functions by reviewing the literature of individual members of the DCC, their interactions with one another, and their proposed mechanisms of action. Studying the DCC permits an understanding of the general principles of transcription and their relationship to chromosome gene wide regulation. This literature review will provide insight into epigenetic regulation, and how the cross-talk between post-translational modifications of histone tails leads to an equalization of X-linked gene expression between male and female *Drosophila melanogaster*.

1.1 Dosage compensation

1.1.2 Dosage compensation in metazoans

Worms, flies, and mammals each have evolved a unique mechanism of dosage compensation to resolve their unequal expression of sex-linked genes (Table 1.1). Disproportionate expression levels of genes encoded on respective sex chromosomes is generated in organisms with heterogametic males and homogametic females. The term dosage compensation was first used to describe how heterogametic males and homogametic females display identical phenotypes from X chromosomal sex-linked genes (Muller *et al.* 1931). *Homo sapiens* males, like male *Drosophila* are heterogametes, posses an X chromosome and a Y chromosome. Female *Homo sapiens*, like female *Drosophila*, are the homogametic sex, inheriting an X chromosome from both male and female parents (Table 1.1). In *Homo sapiens*, dosage compensation occurs in female cells, employing a mechanism termed X chromosomal inactivation. Inactivation is a phenomenon where one of the two X chromosomes undergoes severe condensation and forms a Barr body. This Barr body results in a silent X chromosome that equalizes the gene expression between the sexes through prevention of gene expression of one of the two X chromosome in females (Lucchesi *et al.*, 2005; Nguyen and Disteche, 2006; Wutz and Gribnau, 2007). The Barr body contains histone proteins that have undergone high levels of posttranslational modifications such as methylation and deacetylation, resulting in a tightly packed chromosome inaccessible to transcription. Similar to mammals, dosage

compensation in *Caenorhabditis elegans* takes place in the homogametic sex (Meyer, 2000). The homogametic sex is hermaphroditic, possessing two X chromosomes, and the heterogametic sex is male, possessing one X chromosome (Table 1.1). Dosage compensation acts on both of the X chromosomes of the hermaphrodite to effectively down regulate the level of transcription. The two-fold reduction in gene expression of both X chromosomes then equates to the single male X chromosome. Dosage compensation acts to balance the level of transcription between the sex chromosomes, but the down regulation of the X chromosomes in both mammals and worms creates an imbalance at the level of transcription between the X chromosomes and the autosomes (Gupta *et al.*, 2006; Meyer, 2000; Nguyen and Disteche, 2006).

1.1.2 Dosage compensation in *Drosophila melanogaster*

In *Drosophila* the male is the heterogamete possessing a single X sex chromosome and a single Y sex chromosome, from the male parent, whereas the homogametic female *Drosophila* possess two X chromosomes, one inherited from each parent (Table 1.1). In *Drosophila*, the ratio between X chromosomes and autosomes (X:A) dictates the sex-specific development (Lucchesi and Manning, 1987). The observation that RNA accumulated around the male X chromosome in *Drosophila* suggested that dosage compensation operated at the level of DNA transcription (Mukherjee and Beermann, 1965). The male Y chromosome appears to play a minimal role in *Drosophila*, because it requires no compensation mechanism; however, this creates additional pressure on an organism because the X chromosome becomes the dominant site of regulation. The DCC enables male *Drosophila* to express genes encoded on their single X chromosome to equivalent levels of the gene products of the two X chromosomes in female *Drosophila* (Lucchesi and Manning, 1987). Identification of the DCC in *Drosophila* illustrates a novel method for solving the unequal gene expression of sex chromosomes compared to other eukaryotes (Lucchesi and Manning, 1987).

1.1.3 Genetic and biochemical basis of the DCC

Genetic screens enabled researchers to identify key genes involved in dosage compensation.

Table 1.1: Overview of the different dosage compensation strategies, including the distribution of chromosomes from flies to mammals.

Organism	Female	Male
Drosophila	XX AA	$\mathbf{X}_{Y}$ AA
Mammalian	$\mathbf{X}_{X}$ AA	$\mathbf{X}_{Y}$ AA
C. elegans	XX AA	$\mathbf{X}_{O}$ AA

The sex chromosomes are indicated by an X or a Y and the autosomes are indicated by an A. Capitalization and bolding demarks the sex chromosome affected by dosage compensation. The 0 represents the absence of a sex chromosome.

As dosage compensation occurs in male *Drosophila* genes specifically involved in the dosage compensation process should cause lethality in males but have no effect on the survival of female offspring. Sex specific lethality, the process by which male *Drosophila* develop to third larval instar or the prepupal stage but fail to metamorphose or hatch, is an effective method to understand the biochemical or physiological differences between two sexes and their developmental biology (Belote and Lucchesi, 1980; Fukunaga *et al.*, 1975). Mutations generated by ethyl methanesulfonate (EMS) treatment, found while researchers were scanning the second and third chromosome, facilitated the discovery of male-specific lethal (MSL) 1, MSL2, and maleless (MLE) genes. The identification of these genes led to the characterization of the respective proteins encoded at the gene loci. Homozygous *Drosophila* males possessing a terminal deletion of the third chromosome resulted in male specific lethality, which led to the discovery of MSL3 (Gorman *et al.*, 1995). Scanning the X chromosome using EMS mutations led to the identification of male absent on the first (MOF) (Hilfiker *et al.*, 1997). Moreover, two non-coding RNAs were identified; both gene loci were determined to be in the X-chromosome and termed RNA on the X (roX1 and roX2) (Akhtar *et al*., 2000; Meller and Rattner, 2002; Meller *et al*., 1997).

Involvement of non-coding RNA in *Drosophila* dosage compensation suggests that non-coding RNAs play an important role in the two otherwise dissimilar dosage compensation processes (Stuckenholz *et al.*, 2003; Wutz and Jaenisch, 2000). The non-coding RNA involved in dosage compensation in *Homo sapiens*, termed X-inactive specific transcript (Xist) is also

encoded in the X chromosomes. Xist is central to regulating the X chromosomal inactivation as the inactive X chromosome is coated with Xist transcript.

Co-immunoprecipitation experiments combined with immunofluorescence assays illustrated that the components affecting dosage compensation formed a ribonucleoprotein DCC consisting of MSL1, MSL2, MSL3, MLE, MOF, roX1, and roX2 (Copps *et al*., 1998; Hilfiker *et al*., 1997; Scott *et al*., 2000). Dosage compensation is specific for male *Drosophila* and therefore, assembly of the DCC is only observed in males. The reason for male specificity is because of a master regulatory gene, sex lethal (*sxl*); the Sxl protein is constitutively expressed in female *Drosophila* and not males (Cline, 1979; Lucchesi and Skripsky, 1981). Moreover, a lethal phenotype is observed in female possessing a loss of function Sxl mutant and in males actively expressing Sxl. The Sxl protein is a female specific RNA-binding protein and a key member of the hierarchy of genes regulating sexual dimorphism. The primary way in which Sxl prevents dosage compensation in females is by inhibiting expression of MSL2 protein through preventing splicing of the MSL2 messenger RNA. Specific inhibition of MSL2 in females by Sxl was verified by the removal of the putative Sxl binding sites in the 3' untranslated region of MSL2 mRNA resulting in MSL2 to be actively expressed (Kelley *et al.*, 1995; Lyman *et al.*, 1997). All of the other components of the DCC are actively expressed in both female and male *Drosophila* suggesting additional roles of the dosage compensation components in female *Drosophila* (Kelley *et al*., 1995).

The primary function of the DCC is targeting the acetyltransferase activity of MOF towards K16 of the N-terminal tail of histone H4 (H4K16) specifically along the male X chromosome (Gu *et al.*, 1998; Hilfiker *et al.*, 1997). This posttranslational modification results in the in hyperacetylation of H4K16 precisely along the male X chromosome. Covalent attachment of acetyl groups onto lysine residues on the N-terminal tails of histones results in a neutralization or reduction of the positively charged histone tails leading to a relaxation or loosening of the DNA-histone or nucleosome-nucleosome contacts. Hyperacetylation of H4K16 therefore, results in downstream effects enabling expression of X-linked genes to be equalized between males and females by a process termed hypertranscription. For supplementary up to date reviews on the DCC, refer to (Gelbart and Kuroda, 2009; Hallacli and Akhtar, 2009; Straub and Becker, 2008).

1.2 Components of the *Drosophila* dosage compensation complex

The following subsections provide a review of each of the dosage compensation components: MSL1, MSL2, MSL3, MOF, MLE, and roX RNA (refer 1.2.1 – 1.2.6).

1.2.1 MSL1

Drosophila melanogaster MSL1 protein is 1039 amino acid residues (Palmer *et al.*, 1993) (Fig. 1.1A). The MSL1 protein was found to contain a coiled-coil domain comprised of amino acids 120 – 172, a highly acidic region defined by amino acid residues 368 – 391, and a PEST like sequence within amino acid residues 708 – 801 (Scott *et al.*, 2000). An additional protein domain, termed the PEHE domain was identified as a region extending from amino acids 865 – 1004, containing four characteristic amino acid residues proline, glutamate, histidine, and glutamate (Marin, 2003). The high sequence similarity between the PEHE domain of *Homo sapiens* and *Drosophila* indicates an evolutionarily conserved protein domain.

1.2.2 MSL2

MSL2, a 773 amino acid protein, contains a really interesting new gene (RING) finger motif, residues 25 – 81, a cysteine rich (CXC) motif, residues, 525 – 561, and a C-terminal proline rich region (Fig. 1.1C). The RING finger motif is a type of zinc finger containing the consensus sequence (C_3HC_4 zinc finger) (Lyman *et al.*, 1997). Interestingly, the RING finger motif of *Drosophila* MSL2 has been identified in E3 ubiquitin ligases from yeast to humans (Hibbert *et al.*, 2009). However, a connection between ubiquitination and dosage compensation in *Drosophila* has not been proposed. The CXC motif is also found in a number of other chromatin-associated proteins (Bashaw and Baker, 1995; Fauth *et al*., 2010; Kelley *et al*., 1995; Zhou *et al*., 1995).

1.2.3 MSL3

The MSL3 protein of *Drosophila* is 512 amino acids in length (Fig. 1.1D). The major domains in the MSL3 protein includes a chromatin organization modifier (Chromo) domain, residues 1 – 90, and a MORF4-related gene (MRG) domain comprised of amino acid residues 186 – 490.

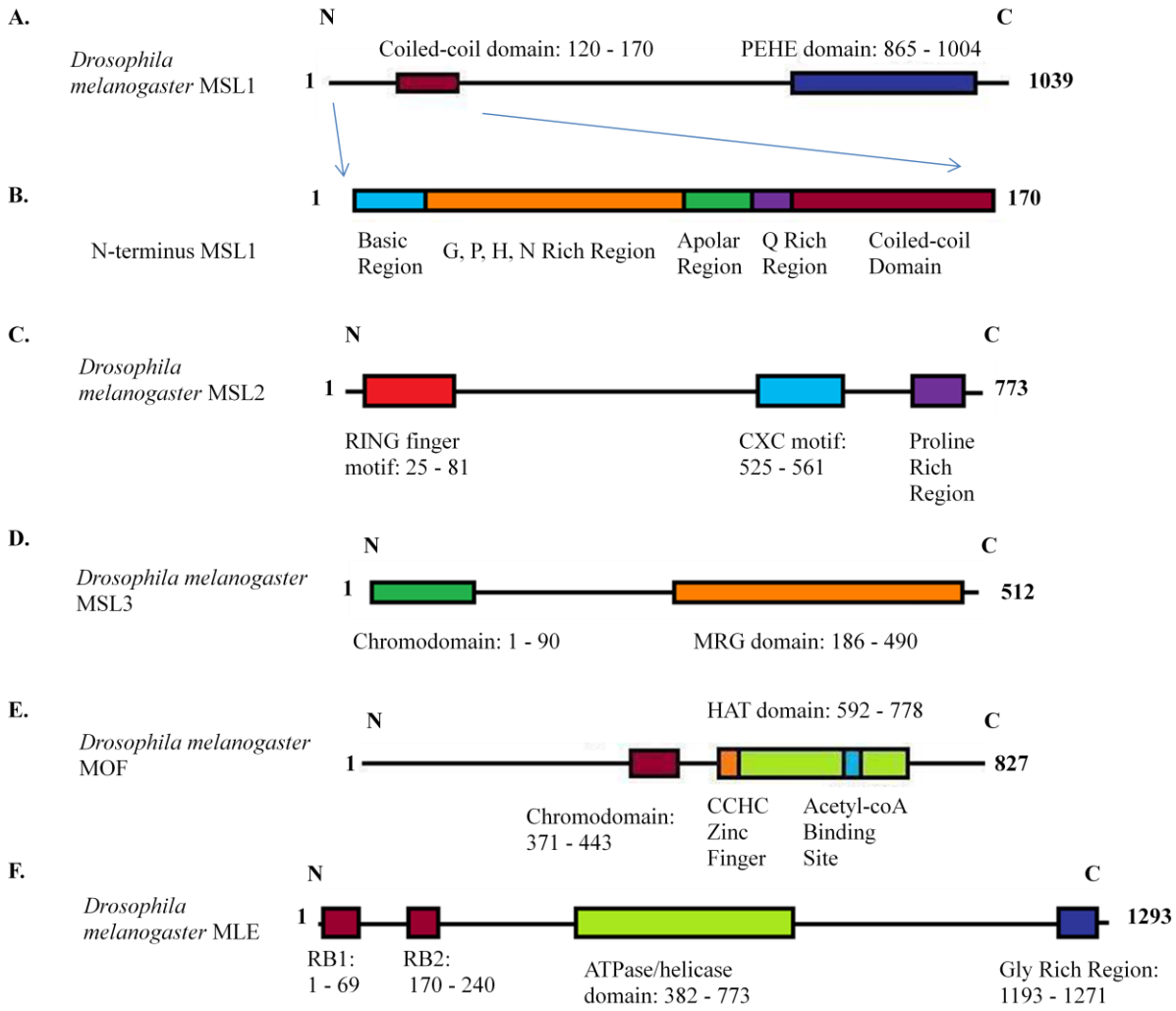


Figure 1.1: Schematic drawing of the protein components of the *Drosophila melanogaster* DCC. (A). Major protein domains of the full-length MSL1 protein. Coiled-coil domain comprised of amino acid residues 120 – 170 and the PEHE domain is amino acid residues 865 – 1004. (B). Regions described within the N-terminus of MSL1 described by Scott *et al*. 2005. The blue arrows indicate expanded region. (C). Major protein domains of the full-length MSL2 protein. Major protein domains of the full-length MSL2 protein including the RING finger motif comprised of amino acid residues 25 – 81, the CXC motif consisting of residues 525 – 561, and a C-terminal proline rich region. (D). Major protein domains of the full-length MSL3 protein. The chromodomain is comprised of amino acid residues 1 – 90 and the MRG domain consists of amino acid residues 186 – 490. (E). Major protein domains of the full-length MOF protein. Chromodomain amino acid residues 371 – 443, HAT domain 592 – 778, CCHC zinc finger residues 539 - 599, Acetyl-coA binding site. (F). Major protein domains of the full-length MLE protein. The two RNA binding domains are amino acid residues 1 – 69 and 170 – 240 respectively, the ATPase/helicase domain comprised of 382 – 773 and glycine rich region extends from 1193 – 1271.

1.2.4 MOF

MOF is a 827 amino acid protein possessing a chromodomain comprised of residues 371 – 443, a CCHC zinc finger comprised of residues 539 - 599, and a MOZ, Ybf3/SAS3, SAS2, and Tip60 (MYST) histone acetyltransferase (HAT) domain consisting of residues 592 – 778 (Fig. 1.1E). Within the MYST HAT domain is the acetyl-coA (Ac-CoA) substrate binding site.

1.2.5 MLE

MLE is a 1293 amino acid protein and a member of the ATP dependent RNA helicase (DExH family). The identified domains of MLE include two RNA binding (RB) motifs RB1 and (RB2), an ATPase/helicase domain, and a glycine rich region (Fig. 1.1F).

1.2.6 roX RNA

The roX1 RNA transcript was determined to be approximately 3,700 nucleotides in length, whereas the roX2 RNA was only approximately 600 nucleotides (Park *et al.*, 2002; Smith *et al.*, 2000). Poor sequence similarity between roX1 and roX2 and their respective homologues from other *Drosophila* suggests the functional properties of roX RNA work at the level of secondary and tertiary structure. Computational prediction combined with multiple sequence alignments against *roX* genes from other species of *Drosophila* revealed *roX1* and *roX2* genes contained a male-specific DNase I hypersensitive site (DHS), a stem-loop structure, and GUUNUAGG repeats (roX box) (Park *et al*., 2007; Park *et al*., 2008; Stuckenholz *et al*., 2003). Three roX box repeats were found at the 3' end of *roX1* and two roX box repeats at the 3' end of *roX2*.

1.3 Nucleosome and chromatin structure in *Drosophila*

The regulatory mechanisms of gene expression work at the level of individual genes, groups of genes or on entire chromosomes. Fundamental to the regulation of eukaryotic genomic DNA is the ability of the cell's DNA to fit within the nucleus and yet remain transcriptionally competent (Bassett *et al.*, 2009). In the nucleus, individual chromosomes are condensed into a 30 nm chromatin fibre (Kumaran *et al.*, 2008). Chromatin is comprised of nucleosome core particles (NCP), consisting of approximately 146 base pairs of duplex

genomic DNA wrapped around an octameric complex of histone proteins $(H2A:H2B:H3:H4)_2$ like beads on a string (Luger *et al.*, 1997). The organization of chromatin results in a barrier for the transcription of genes and plays an integral role in the organism's maintenance of cellular functions and development (Narlikar *et al.*, 2002). Eukaryotic chromatin exists in several compaction states that can be broadly divided into two main types, euchromatin and heterochromatin (Kumaran *et al.*, 2008). Euchromatin is mainly comprised of gene-rich regions packaged loosely to facilitate gene transcription, and is transposon poor (Kumaran *et al.*, 2008; Vermaak and Malik, 2009). Heterochromatin is densely packed, transposon rich, and gene sparse. Compaction of heterochromatin provides a means of regulating gene expression and segregating chromosomes into functional units. The 30 nm chromatin fibre was further proposed to be in two unique nucleosome packing densities, stating that the euchromatic fibre is 6 nucleosomes per 11 nm, whereas heterochromatic fibre is 12 – 15 nucleosomes per 11 nm (Fig. 1.2) (Bassett *et al.*, 2009). The relationship between spatial requirements of DNA interacting proteins and exposure of unique consensus sequences illustrates how nucleosomal packing regulates transcription. In *Drosophila,* two of the three autosomes, 2L/R and 3L/R are primarily in a euchromatic state with some distinct regions of heterochromatin such as the centromeres. The male Y sex chromosome and chromosome 4 are largely in a heterochromatin state. The X chromosome also has distinct regions of euchromatin and heterochromatin (Fig. 1.3).

In *Drosophila*, certain cells can permit an increase in their cell volume and elevate their gene expression levels by altering their chromosomes' structure; these highly specialized chromosomes are generated by replication of the genomic DNA thousands of times without cell division, termed polyploid cells. Several types of secretory cells of *Drosophila* adults, including salivary glands and several cells in *Drosophila* larvae, are polyploid cells. Under such conditions the homologous chromosomes align to form polytene chromosomes (Fig 1.4) (Zhimulev *et al.*, 2004). The polytene chromosomes make *Drosophila melanogaster* an ideal organism to study chromosome compaction, gene regulation, epigenetics, and dosage compensation. Furthermore, polytene chromosomes also maintain their organization into euchromatin and heterochromatin. Polytene chromosomes possess an additional arrangement, termed bands and interbands.

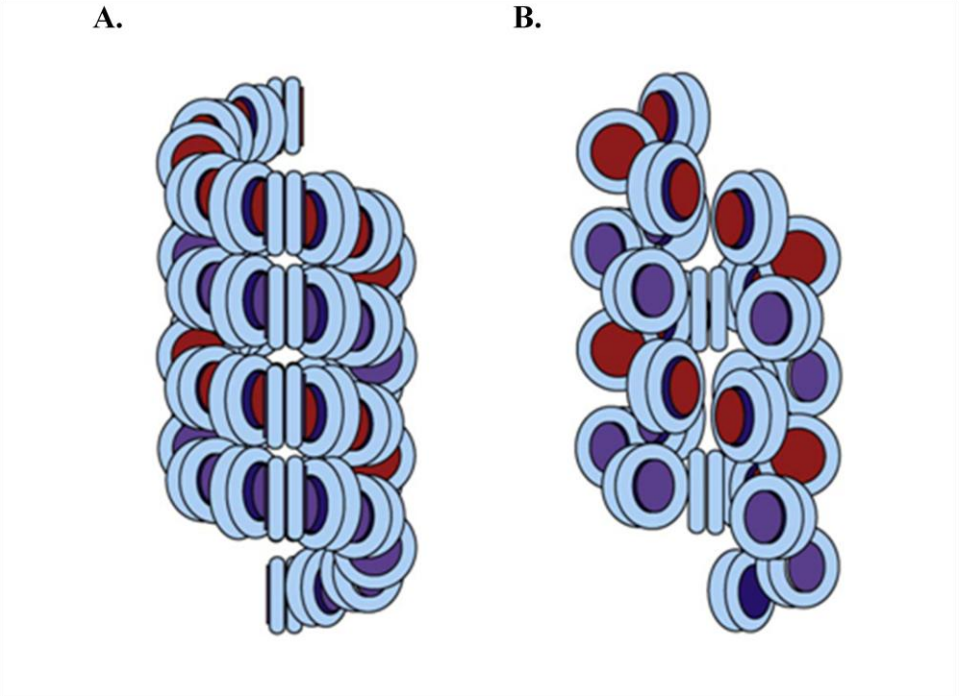


Figure 1.2: Alternative nucleosome packing in proposed 30 nm chromatin fibre. (A). Model representation of heterochromatin, a mass to unit ratio of 12 – 15/11 nm. (B). Representation of the 30 nm euchromatin fibre with a mass to unit ratio of 6/11 nm. Figure taken from (Bassett *et al.*, 2009). Individual NCP are coloured light blue with either a dark blue center or a burgundy center. Note the compaction of genomic DNA maintains a double helical conformation.

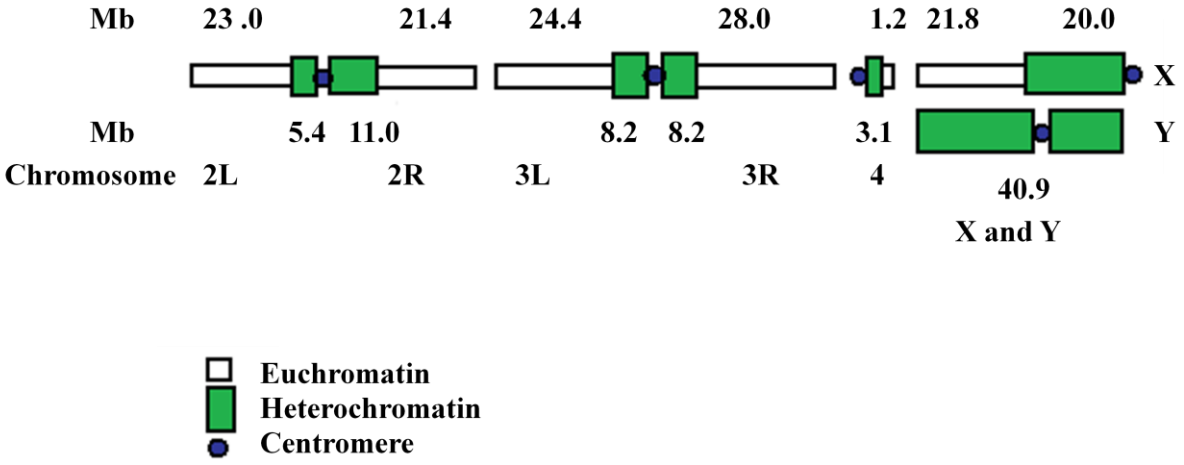


Figure 1.3: Overview of the chromosomes in *Drosophila melanogaster*. The autosomes include chromosome 2, 3 and 4. Both chromosome 3 and 4 are comprised of a left and right region joined by a centromere. The sex chromosomes are the X and Y chromosome. The three states of chromatin are illustrated along the individual chromosomes.

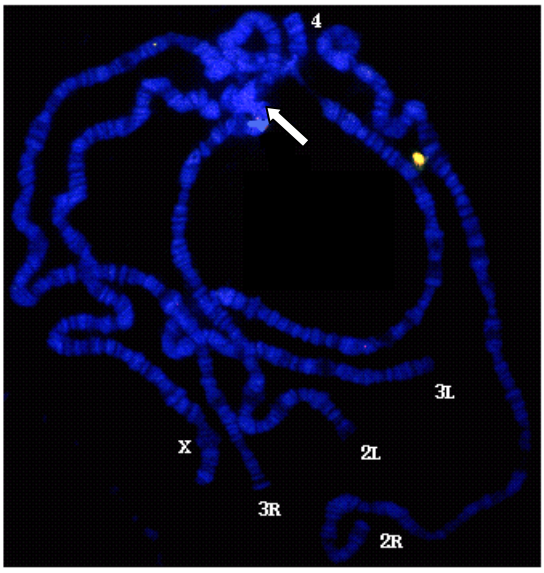


Figure 1.4: Illustration of the bands and interbands observed by fluorescent DAPI stained *Drosophila melanogaster* polytene chromosome (msg.ucsf.edu/sedat//polytene_chrom.html). Bands are stained blue and interbands appear black. The arrow demarks the chromocenter and the individual chromosomes are indicated as X, 2, 3, and 4. The right and left arm of chromosome 2 and 3 are shown by an L or a R.

Bands and interbands were discovered by the unique staining patterns of 4', 6-diamidino-2-phenylindole (DAPI), a fluorescent dye which strongly interacts with the minor groove of DNA (Fig. 1.4) (Bayani *et. al.* 2004). *Drosophila* polytene chromosomes also contain a chromocenter, defined as a heavily stained region where heterochromatic regions surrounding the centromeres align.

The consistent banding pattern of the polytene chromosomes is evidence that the *Drosophila* interphase chromosomes are in a highly organized arrangement (Zhimulev *et al.*, 2004). The decondensation of bands, termed chromatin puffs because of active transcription was originally observed in polytene chromatin, from *Drosophila* salivary glands (Semeshin *et al.*, 2008). Both bands and interbands are packaged into nucleosomes, but there are clearly significant differences in the higher order packing of the chromatin fibre. Bands are typically compacted regions that are transcriptionally silent. Interbands appear as more dispersed fibrils encoding non-coding regions and housekeeping genes. The decompaction of interbands has been proposed to result in activation of proteins possessing insulator or barrier functions. Furthermore, interbands may also interact with nuclear matrix factors, playing a role in nuclear

architecture, as regulatory regions within promoter regions for RNA polymerase II, and function as a way to divide the chromosome into subdomains. Lastly, it was determined that a band plus an interband does not exist as a distinct cytogenetic unit, rather that both bands and interbands act as autonomous units (Semeshin *et al.*, 2008). From cytological studies of the dosage compensation proteins, it has become evident that the DCC localizes to the interband regions of the male X chromosome, but not to the bands or the heterochromatic centromeres. Hence ascertaining how the DCC functions requires elucidation of the structural details governing bands and interbands, and their correlation with nucleosome packing, epigenetic markers, and gene transcription.

1.4 Epigenetic markers and states of chromatin

Regulation of chromosomes into heterochromatin and euchromatin does not stop there; there exists epigenetic mechanisms for controlling the information encoded in the DNA of chromosomes (Jenuwein and Allis, 2001; Marmorstein and Trievel, 2009; Suganuma and Workman, 2008). The mechanisms of epigenetic actions are regulated through histone modifications, histone variants, noncoding RNAs, and nucleosome remodelling. Each of the four histone proteins contains unstructured regions at their respective N- and C-termini termed histone tails that are rich in lysine and arginine amino acid residues. The tails of histone proteins are known sites of posttranslational modifications. Various permutations of histone tail modifications are thought to determine a histone code whereby the transcriptional competence of a gene is determined by a "bar code" of modifications maintained on the histone tails (Jenuwein and Allis, 2001). Histone modifications, more specifically, are posttranslational modification of the N-terminal tails of the histone proteins H3 and H4, and, in some cases, H2A and H2B; types of modification include lysine acetylation, lysine or arginine N-methylation, ubiquitination, phosphorylation, sumolyation, arginine deimination, ADP-ribosylation, and lysine deacetylation. Histone modifications are mainly targeted towards the ε-amine group of lysine residues and the θ-amine of arginine residues (Smith and Denu, 2009).

Solving the crystal structure of the *Xenopus laevis* NCP (core histone proteins bound to a 147 base-pair DNA fragment, modified from human α-satellite DNA) enabled visualization of the potential mechanisms of the histone code (Luger *et al.*, 1997; Luger and Richmond, 1998) (Fig. 1.5A - C).

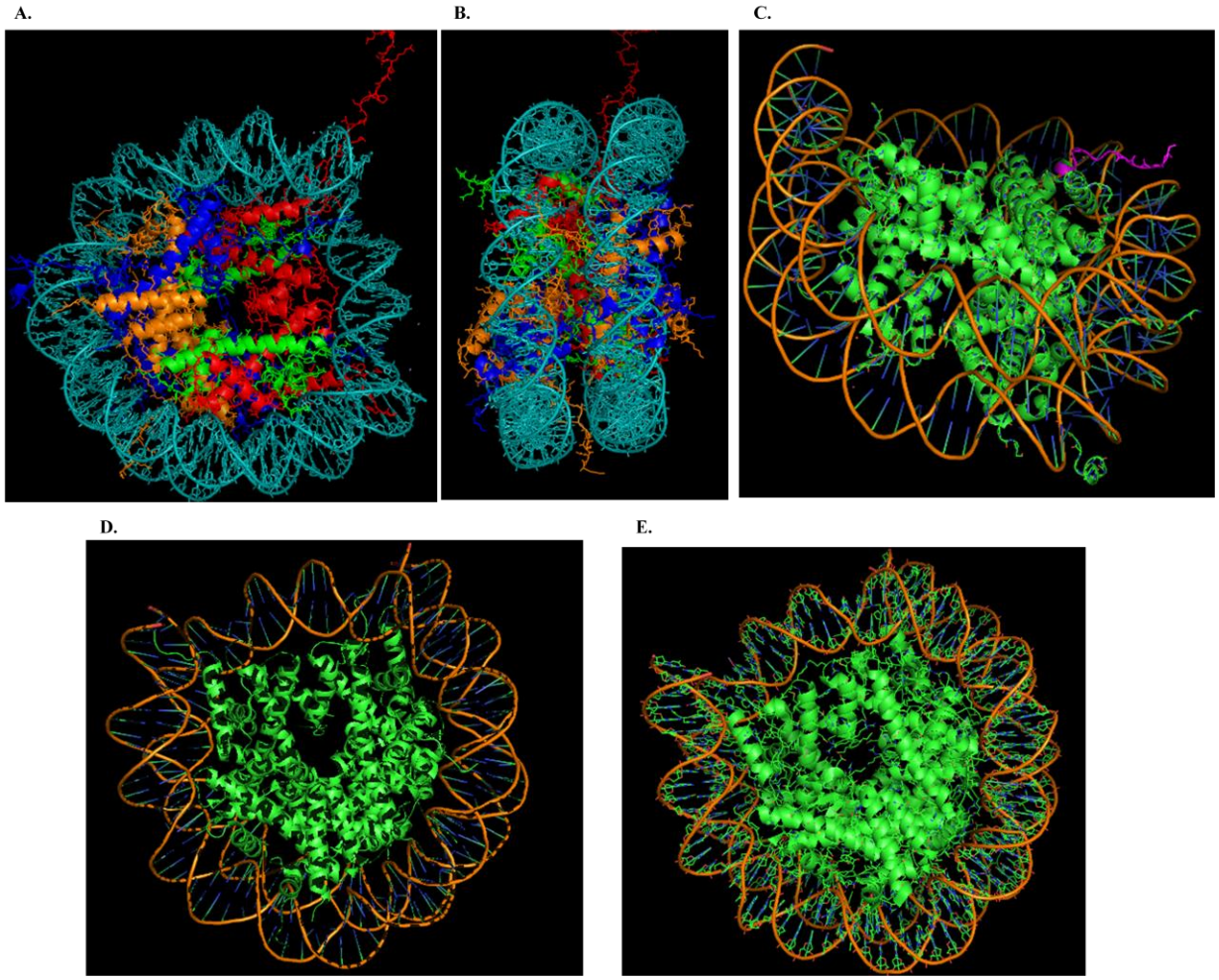


Figure 1.5: Solved nucleosome core particles from *Xenopus laevis*, *Drosophila melanogaster*, and *Homo sapiens* by x-ray crystallography. (A). NCP from *Xenopus laevis* (PDB accession number: 1KX3, (Davey *et al.*, 2002) including 147 bp double stranded DNA fragment (cyan), H2A (orange), H2B (dark blue), H3 (red), and H4 (green). (B). Side view of NCP from *Xenopus laevis* emphasising the N-terminal tails from each of the histone proteins. (C). NCP from *Xenopus laevis* emphasising the H4 N-terminal tail including the conserved amino acid residues from 16 – 25 (magenta). (D). NCP from *Drosophila melanogaster* (PDB accession number: 2PYO,(Clapier *et al.*, 2008)). (E). NCP from *Homo sapiens* (PDB accession number: 2CV5, (Tsunaka *et al.*, 2005).

The histone tails account for approximately 30% of the core histone protein sequence and are largely comprised of lysine and arginine residues; consequently, they have a prominent positive charge at neutral pH (Luger and Richmond, 1998). The interaction between histones and DNA is mediated in part through hydrogen bonds from positively charged amino acids on the histone proteins to the phosphodiester backbone of the DNA double helix. The N-terminal

tails prefer interaction with the DNA minor groove. The N-terminal tail of histone H4 is a highly conserved region encoding 4 invariant lysine residues at positions 5, 8, 12, and 16 (Megee *et al.*, 1995). More specifically, the amino acid residues 16 – 24 of the histone H4 amino-terminal tail are critical for crystallization of the NCP, where the basic polypeptide of the histone H4 N-tail interacted with an acidic region formed by H2A and H2B on the surface of the NCP. Intriguingly, the acidic region of H2A is highly conserved among histone variants and is the only highly acidic patch on the surface of the histone octamer (Luger and Richmond, 1998) (Fig. 1.5). The importance of the histone H4 tail was stressed in mutagenesis studies in *Saccharomyces cerevisiae*; the absence of the N-terminal tail resulted in a reduced viability and genome integrity (Megee *et al.*, 1995). Solving the NCP structure of *Drosophila* and *Homo sapiens* revealed a high degree of conservation of the overall tertiary and quaternary structure (Fig. 1.5D, E) (Clapier *et al.*, 2008; Tsunaka *et al.*, 2005). The strict conservation of the NCP across evolution was illustrated in the 99% sequence identity of histone H4 amino acid sequences (Clapier *et al.*, 2008). How the DNA is contorted about the NCP results in different lengths of genomic DNA per NCP. This translates into a functional relationship between NCP compaction and transcription, because the average lengths of exons are approximately equal to the average length of DNA wrapped around a NCP (Schwartz *et al.*, 2009). Acetylation is thought to relax nucleosomes, reduce the number of internucleosomal interactions, by an alteration of the electrostatic interactions between the positively-charged histone N-terminal tails and the negatively charged phosphodiester backbone of nucleosomal DNA. Classically, acetylation is thought to promote accessibility of the genomic DNA of promoter regions. Moreover, acetylation may also function as a sight for the recruitment of specific macromolecular complexes involved in chromatin remodelling, cell cycle, splicing, and nuclear transport (Choudhary *et al.*, 2009). With regard to dosage compensation, it has been suggested that the acetylation of the H4K16 by MOF likely opens up the chromatin structure on the male X chromosome by altering the interaction between the histone H4 N-terminal tail and the acidic region of H2A and H2B of the neighbouring nucleosome, thereby facilitating an increase in transcription of X chromosomal genes.

1.4.1 Overview of histone modifications implicated in dosage compensation

Regulating the indexing of chromosomes into functional euchromatin or heterochromatin subdomains is critical to maintaining distinct types of chromatin. Within the *Drosophila* genome there are greater than 50 suppressor of position-effect variegation (PEV) [Su(var)] loci that encode structural components of chromatin. All of the Su(var) genes are thought to all be involved in defining, regulating, and maintaining heterochromatin (Ebert *et al.*, 2006; Ebert *et al.*, 2004). PEV was originally described in white mottled-4 translocation (w[m4]), where an inversion on the X chromosome results in the w[m4] gene to relocate adjacent to pericentric heterochromatin, resulting in a red eye colour compared to wild-type white eye colour. Inactivation or altered expression of specific genes by PEV is mediated through the rearrangement of gene loci: normally expressed euchromatic gene loci relocate either into silent heterochromatin or adjacent centromeric regions (Girton and Johansen, 2008). Altered expression is thought to be through histone modifications (H3K9 tri-methylation) and the recruitment of heterochromatin related proteins, results in spreading of heterochromatin, which overrides the basal expression level. Moreover, PEV exemplifies how subdomains of chromosomes, nucleosome packing, and histone modifications regulate gene expression. Polytene squashes of *Drosophila* carrying PEV mutations revealed differentiation from a structured banded morphology into a diffuse banding morphology. Three of the most established Su(var) gene products that affect dosage compensation are the zinc finger protein Su(var)3-7, the histone methyltransferase (HMTase) Su(var)3-9, and the chromodomain protein Su(var)2-5 encoding heterochromatin protein 1 (HP1). Importantly, disruption of the modifiers of variegation also alters the morphology of the male polytene chromosome and dosage compensation of the male X chromosome (Spierer *et al.*, 2005).

In *Drosophila*, there are five HP1 protein isoforms (Vermaak and Malik, 2009). The defined protein domains of HP1 consist of a chromodomain, required for binding H3K9Me_2, a chromo shadow domain, involved in homodimerization and protein-protein interactions, and a hinge domain, involved in mediating localization and interaction with genomic DNA. It has been documented that HP1 interacts with numerous proteins in connection with transcription, DNA replication, DNA repair, nuclear organization, and telomere maintenance (Vermaak and Malik, 2009). For instance, loss of Su(var)3-7 or HP1 results in specific male lethality. The connection between HP1 and the DCC likely has to do with the maintenance of

heterochromatin boundaries on the X chromosome as defined by the presence of repressive methylation markers on H3K9 and the phosphorylation of H3S10 by kinase Jil-1 (see below) (Spierer *et al.*, 2005).

Jil-1 encoded in the Su(var)3-1 locus is a chromosomal histone H3 protein tandem kinase. From immunofluorescence colocalization experiments, it has become evident that both Jil-1 and the DCC are enriched at the interbands along the male X polytene chromosome (Ebert *et al.*, 2004; Jin *et al.*, 2000). Notable pull-down assays proved that the kinase domains of Jil-1 interacts with MSL1 (Jin *et al.*, 2000). In *Drosophila* homozygous for Jil-1 mutated alleles, the euchromatin bands are absent and all of their chromosome arms are drastically condensed. Ectopic expression of Jil-1 can rescue the null Jil-1 phenotype, resulting in complete restoration of chromatin to the appearance of wild-type chromatin (Wang *et al.*, 2001). In the case of Jil-1 null hypomorphic alleles, the reduced levels of H3S10 phosphorylation permitted dimethylation of H3K9 by Su(var)3-9 and allowed a subsequent spread of HP1 to ectopic locations along all chromosomes, including the highest levels along the male X chromosome (Zhang *et al.*, 2006b). Therefore, the proposed model of Jil-1 kinase activity is that it functions in the maintenance of the euchromatin, and establishment of chromatin structure. Furthermore, phosphorylation of H3S10 was found to be critical in transcription elongation by release of RNA Polymerase II from promoter-proximal pausing complicating the issue of the DCC and Jil-1 crosstalk (Ivaldi *et al.*, 2007). This reflects the ability of both the DCC and Jil-1 to help to regulate the compaction state and transcriptional competence of the X-chromosome.

ATPase dependent nucleosome remodelling complexes are thought to affect dosage compensation by affecting the structure and spacing of nucleosomes. The hyperacetylation of H4K16 along the male X chromosome has been found to antagonize ISWI function (Corona *et al.*, 2002). ISWI is a member of the SWI2/SNF2 family of nucleosome remodelling ATPase complexes, which are thought to promote chromatin folding. The relationship between dosage compensation and other nucleosome modifying complexes is in the maintenance of higher-order chromatin structure, in crosstalk between histone modifications, and in the subdivision of chromosomes. Particular to the function of the DCC is the intricate balance of the histone tail modifications along the male X chromosome.

Encoding epigenetic information in higher eukaryotes is in part carried out by DNA methylation (Mandrioli and Borsatti, 2006). Predominantly DNA methylation occurs on the

CpG dinucleotide, which is thought to mediate chromatin silencing by recruitment of methyl-CpG binding proteins. However, the occurrence of DNA methylation in *Drosophila* is only beginning to be acknowledged, subsequently how it affects dosage compensation is not yet known.

1.4.2 Histone acetyltransferase: mechanism and recognition

One of the most abundant histone posttranslational modifications (epigenetic marker) is the reversible acetylation of lysine residues on histone N-terminal tails. The specific acetylation of H4K16 along the *Drosophila* male X chromosome by MOF requires an appreciation of histone acetylation. The exposure of genomic DNA for replication or transcription depends on local nucleosome organization. It is understood that acetylation of histone tails by HATs induces micro- and macro-specific gene regulations. The targeting of specific lysine residues by HAT enzymes within the N-terminal tails of the histone proteins is poorly understood. Collectively it has been observed that some HATs precisely target promoters to mediate activation or repression of specific genes, whereas other HATs act globally on large domains of chromatin, including both coding regions and non-promoter regions (Dekker and Haisma, 2009; Marmorstein and Trievel, 2009). The involvement of HATs in developmental defects and cancer has inspired recent attempts to understand the targeting mechanism of HATs and the cascade that results from the acetylation of the lysine residues of the N-terminal tails.

The majority of HAT enzymes participate in large multiprotein complexes (Marmorstein and Trievel, 2009). HAT enzymes do not always target histone proteins; for instance p300/CBP also acetylates transcription factors and non-transcription factors (Bannister and Kouzarides, 1996; Kalkhoven, 2004; Marmorstein and Trievel, 2009). Moreover, the human homologue of MOF, hMOF, is known to acetylate p53 (Li *et al*., 2009). The poor sequence conservation between different HAT families and the different catalytic mechanisms attests to the diversity HATs (Marmorstein and Trievel, 2009).

The DCC component MOF belongs to the MYST family of histone N-acetyltransferase enzymes. The MYST family of HATs include human HIV-1 Tat interactive protein (Tip60), human monocyte leukemia zinc finger protein (MOZ), yeast something about silencing proteins (SAS2 and SAS3), yeast essential SAS2-related acetyltransferase (Esa1), human

HBO1, mouse Querkopf, and fly MOF (Table 1.2) (Thomas and Voss, 2007). The fact that MOF, a key member of the DCC and dosage compensation in male *Drosophila*, remains expressed in female *Drosophila* implies additional roles, such as acetylation of other proteins and/or histone proteins. In *Drosophila* females, H4K16 acetylation occurs in the 5' promoter regions of genes, which is atypical in relation to what occurs in males where acetylation occurs in the coding 3' region of genes (Kind *et al*., 2008; Prestel *et al*., 2010). Superposition of the solved protein structures of hMOZ, hMOF, Tip60, and Esa1 illustrates the high degree of conservation of the tertiary structure and the active site in the MYST family of HATs (Fig. 1.6A, B). A MOF construct, comprised only of amino acid residues 518 to 827, maintained its ability to acetylate histones and specificity for histone H4 (Smith *et al.*, 2000). Recombinant MOF with a G691E mutation in the acetyl-Co binding site, possessed a 10 fold decrease in HAT activity towards histone substrates compared to recombinant full-length wild type MOF (Fig. 1.6C, D) (Akhtar and Becker, 2000; Hilfiker *et al.*, 1997). Also, the recombinant full-length MOF was shown to preferentially monoacetylate H4K16 on both free and nucleosomal histones; these results were corroborated by the inability of the G691E mutant to acetylate histones. Expression of G691E MOF, containing an inactive HAT domain, greatly reduced the level of DCC assembly on the X chromosome in females (Gu *et al.*, 2000). Therefore, MOF has been determined to be the sole acetyltransferase within the DCC.

Table 1.2: Summary of the MYST HAT family in metazoans.

HAT	Organism
MYST family	
Sas2	Yeast
Sas3	Yeast
Esa1	Yeast
MOF	Fruit fly
HBO1	Human
MOZ	Human
Tip60	Human

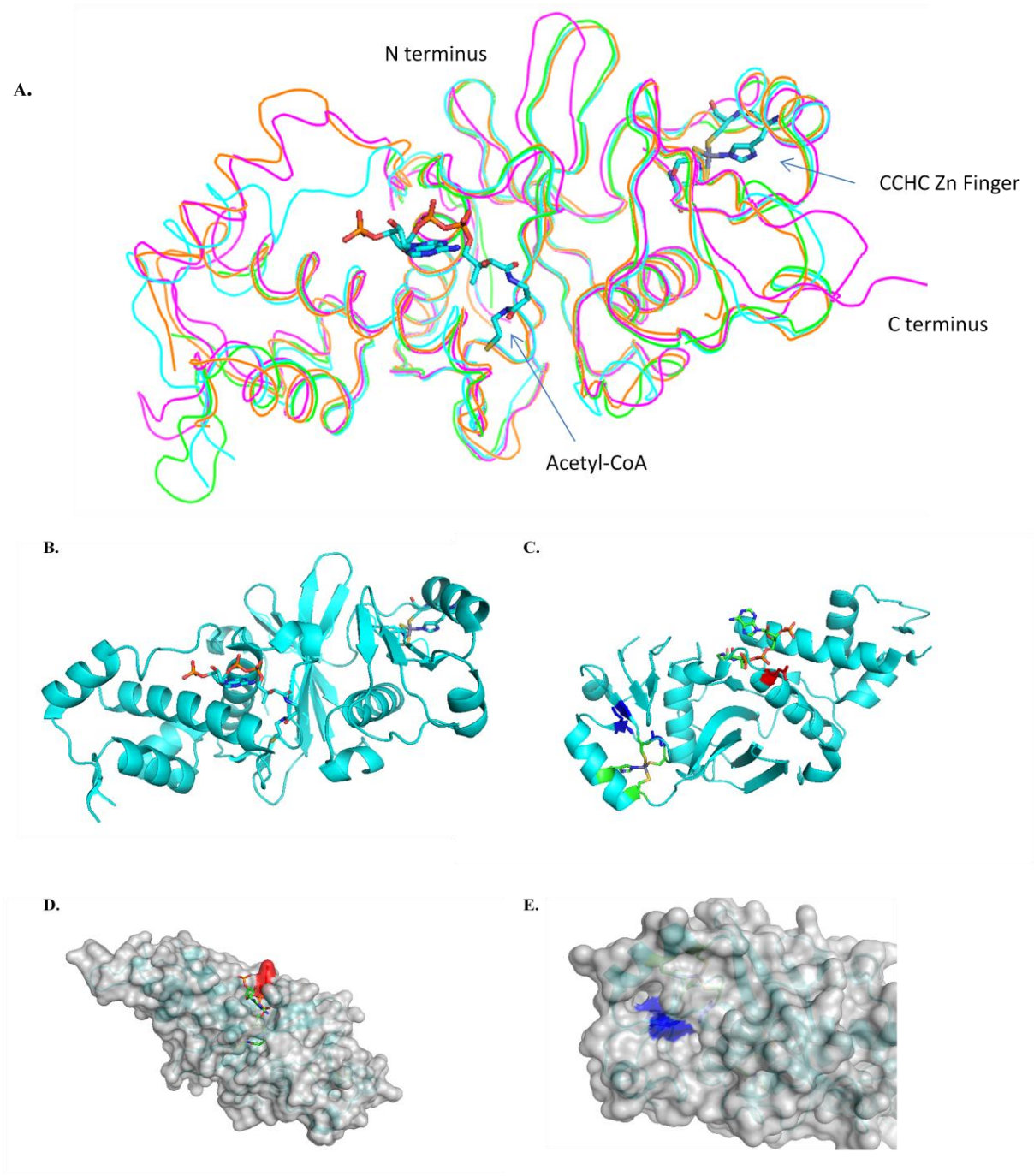


Figure 1.6: Solved protein structures of MYST family HATs. (A). Superposition of Esa1 (magenta) (PBD accession id: 1MJA), Tip60 (green) (PBD accession id: 2OU2), *Homo sapiens* MOZ (orange) (PBD accession id: 2RC4), and *Homo sapiens* MOF (cyan) (PBD accession id: 2PQ8) interacting with acetyl-coA, and indicating the conserved zinc finger. (B). Ribbon representation of *Homo sapiens* MOF illustrating the secondary structure, the CCHC zinc finger, and Acetyl-CoA. (C). Ribbon representation of *Homo sapiens* MOF showing the G691E mutation (red), and the Y572G, L578G and Y580G mutations within the zinc finger (blue). (D). Surface accessibility of the G691E mutation. E. Surface accessibility of the Y572G, L578G, and Y580G mutations (Akhtar and Becker, 2000; Akhtar and Becker, 2001).

Insight into the how MOF catalyzes the acetylation of H4K16 comes largely from the proposed mechanism for Esa1 (Marmorstein and Roth, 2001; Marmorstein and Trievel, 2009; Yan *et al*., 2002). In the reaction mechanism, the first step is deprotonation of the sulphydryl group of C304 by E338 (Yan *et al.*, 2002). This enables a nucleophilic attack by C304 on the thioester of Ac-CoA generating an acetyl-cysteine enzyme intermediate. The acetyl moiety is then transferred to the lysine residue of the histone N-terminal tail by a nucleophilic displacement reaction involving the deprotonated ε-amino group of the lysine substrate. Moreover, both the cysteine and the glutamate residues are conserved all of the members of the MYST family suggesting conservation of the catalytic mechanism (Marmorstein and Trievel, 2009).

One study has hypothesized that the specific covalent modification of the histone tails generates high affinity docking surfaces for proteins to interact with chromatin (Nielsen *et al.*, 2002). MOF and MSL3 each contained unique chromodomains (Fig. 1.1D, E). Chromodomains are known to be a common protein domain found in chromosome modifying/interacting proteins; the chromodomain was originally identified in *Drosophila* modifiers of variegation, HP1 and polycomb (PC) protein (Paro and Hogness, 1991). The chromodomains of MOF and MSL3 have been described as a chromo-barrel domain as they are both thought to be unique compared to the chromodomains of HP1 and PC (Bertram and Pereira-Smith, 2001; Nielsen *et al*., 2005). The chromodomain of HP1 consists of a three-stranded β-sheet whereas the chromo-barrel domain of MSL3 and MOF are comprised of a four-stranded β-sheet (Nielsen *et al*., 2005). However, the functions of chromodomains are typically grouped in with the canonical chromodomain of HP1 and PC proteins; the chromodomains from both proteins are known to bind methylated N-terminal tails of histone proteins through a conserved aromatic pocket. The canonical chromodomain interacting with a histone tail via an induced-fit mechanism consists of a three-stranded anti-parallel β-sheet and an α-helix packed against the β-strand of the histone H3 tail (Jacobs and Khorasanizadeh, 2002). Chromodomains differentiate the different methylated lysine residues by unique recognition grooves; for example, for K9 the recognition motif is $QTARK_9$, whereas for K27 the motif is $KAARK_{27}$ (Fischle *et al.*, 2003; Nielsen *et al.*, 2002). In the *Drosophila* HP1 structure a three walled cage generates the methyl-amino binding site, which consists of Y24, W45, and Y48; mutation of the three aromatic residues drastically reduced the binding affinity

of the H3K9Me$_3$ peptide (Jacobs and Khorasanizadeh, 2002). The *Homo sapiens* chromodomain of MORF-related gene on chromosome 15 (MRG15) was solved revealing that the chromodomain is also composed of a β-barrel (four-stranded β-sheet) and a α-helix (Zhang *et al.*, 2006a). MRG15 a paralogue of MSL3, was found also to possess a methyl-lysine binding cage. Preliminary data also suggests the MRG15 binds to H3K36Me$_3$ peptides. *In vitro* binding assays with GST-MRG15 determined MRG15 specifically interacts with H3K36Me$_2$ in the presence of other di-methyled lysine residues. Moreover, the chromodomain of MRG15 was found to align with the chromodomain of HP1 (Fig. 1.7A, B). Alignment of the chromodomain of HP1 and MRG15 confirmed that both proteins shared the same three stranded β-sheet (β-barrel) and a short α-helix, and the ability to bind tri-methylated lysine. Moreover, within the β-barrel of MSL3 there is a hydrophobic pocket generated by Y26, Y46, and Y49, which is thought to enable interaction with methylated histone N-terminal tails (Nielsen *et al.*, 2005). The presence of an aromatic pocket analogous to the methyl binding pocket of MRG15, HP1 and PC suggested the chromodomain of MSL3 may also bind to methylated histone tails. Furthermore, the solved chromodomain of *Drosophila* MOF revealed that the chromodomain was again comprised of a β-barrel and a short α-helix, but lacked the aromatic pocket (Nielsen *et al.*, 2005). Interestingly, it is possible to super-impose the structures of the chromodomain of MOF and the chromodomain of HP1 suggesting structural similarities (Fig. 1.7C, D).

1.5 Function of the DCC

1.5.1 Protein interactions and assembly of the DCC

Biochemical analysis coupled with immunofluorescence studies has enabled great insight into the function and mechanism of the DCC. The step by step assembly of the DCC is not completely understood; however, the initial step requires interaction between MSL1 and MSL2, which enables their binding to the X chromosome (Scott *et al.*, 2000). Yeast two-hybrid assays determined that MSL1 and MSL2 interact via amphipathic coiled coil α-helical regions, amino acid residues 85 – 186 of MSL1 and amino acids 39 – 82 of MSL2 (Fig. 1.8) (Copps *et al.*, 1998). The amphipathic coiled coil α-helical region of MSL1 includes the apolar region and the coiled-coil domain, and of MSL2 the RING finger (Fig. 1.1A, B, C) (Li *et al.*, 2005). In terms of chromatin association MSL1 and MSL2 possesses overlapping binding

patterns, suggesting both proteins target and tether the core complex to the male X chromosome (Lyman *et al.*, 1997).

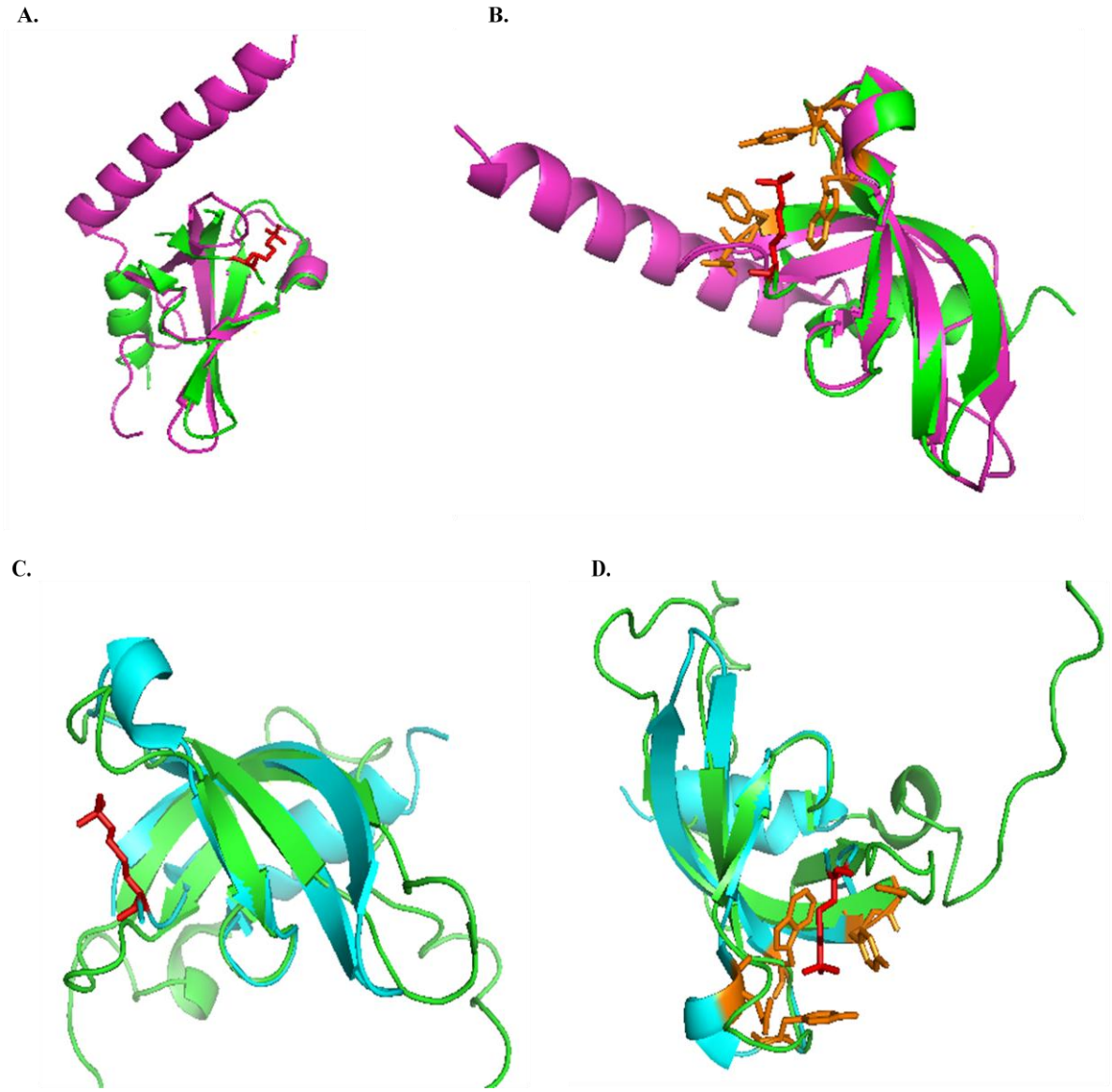


Figure 1.7: Solved structures of the chromodomains of MRG15, MOF, and HP1. (A). Structural superposition of chromodomains of MRG15 (magenta) (PBD accession id: 2F5K) and HP1 (green) (PBD accession id: 1KNE) and the interaction with trimethylated lysine peptide ($H3K9Me_3$) (red). (B). Structural superposition of the chromodomain of MRG15 against HP1 showing the methyl lysine binding site of HP1 comprised of three aromatic amino acids Y24, W45, and Y48 of HP1 (orange) interacting with $H3K9Me_3$ (red). (C). Structural superposition of the chromodomains of MOF (cyan) (PBD accession id: 2BUD) and HP1 (green), and the $H3K9Me_3$ (red). (D). Structural superposition of the chromodomain of MOF against HP1 showing the methyl lysine binding site of HP1 comprised of three aromatic amino acids Y24, W45, and Y48 of HP1 (orange) interacting with $H3K9Me_3$ (red).

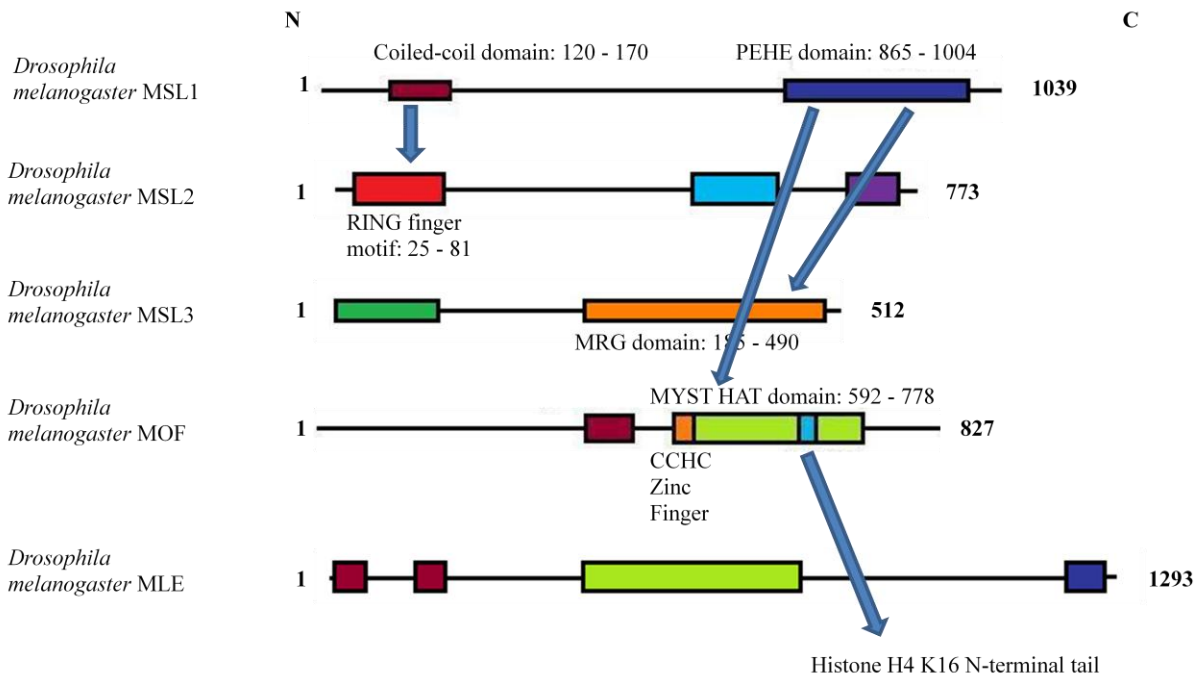


Figure 1.8: Illustration of the protein-protein interactions between the individual proteins of the DCC. See text for further detail, section 1.5.1.

The C-terminus of MSL1 consisting of the PEHE domain has been shown to interact via a pull-down assay with both MOF and MSL3 with *Drosophila* cell extracts (Fig. 1.8) (Scott *et al.*, 2000). Further studies found that the MRG domain of MSL3, residues 186 – 490, interacted with MSL1, residues 973 – 1039 (Morales *et al.*, 2005; Morales *et al.*, 2004). Also, amino acid residues 766 – 939 of MSL1 were found to be required for the interaction with MOF. Protein interaction between MSL1 and MOF was found to be dependent on the CCHC zinc finger domain of MOF, but the exact amino acid residues of MOF required for this interaction remains unknown (Morales *et al.*, 2004). The interaction between MSL3 and MOF in *Drosophila* is dependent on prior interaction with MSL1 as neither MSL3 nor MOF can directly interact with one another (Buscaino *et al.*, 2003; Morales *et al.*, 2004).

The incorporation of roX RNA into the DCC by MLE has remained poorly understood. Studies have indicated that roX RNA mediates the association of MLE with the DCC and, in reciprocal fashion MLE incorporates roX RNA into the DCC (Lee *et al.*, 1997). Another study proposed that MSL2 played a role in the transcription of roX RNA (Rattner and Meller, 2004). Studies with mutant MLE have shown the absence of MSL3 and MOF from the male X

chromosome indicating MLE also plays a role incorporating MOF and MSL3 into the DCC (Gorman *et al.*, 1995; Gu *et al.*, 1998). In a reciprocal fashion the integration of MLE with the DCC has also been linked to MSL3. As observed in male *Drosophila* lacking MSL3, MLE no longer associated with the X chromosome (Gorman *et al.*, 1993). Moreover, MSL3 has been shown to be involved in the correct localization of MOF, as the elimination of MSL3 by RNAi knockdown resulted in absence of MOF along the male X chromosome (Buscaino *et al.*, 2003).

1.5.2 The DCC requires multiple protein interactions for proper activity

Direct interaction between MSL1, MSL3, and MOF has been shown to promote activation of MOF's nucleosomal HAT activity and increase substrate specificity towards histone H4 (Morales *et al*., 2004). Moreover, the inclusion of MSL1 or MSL3 with MOF did not result in any positive changes in MOF's HAT activity. Although H4K16 is the key target for the MOF's HAT activity, both MSL1 and MSL3 can be acetylated by MOF *in vitro* (Morales *et al*., 2004). As MSL3 and MOF cannot directly interact in *Drosophila* it is thought that the MSL1 scaffold brings MSL3 and MOF into direct contact, which could be expected to result in a conformational change of MOF that induces a more concise binding of substrate and enhanced activity of the acetyltransferase. A relation, therefore, must exist between the histone acetyltransferase activity of MOF and the other proteins of the DCC. Studies on Tip60, a member of the MYST HAT family, established a relationship between HAT activity and histone recognition by the Tip60 chromodomain (Sun *et al.*, 2009). The chromodomain of Tip60 was found to interact directly with histone H3K9Me$_3$ and this increased HAT activity towards Tip60 substrate. Therefore, the chromodomain must be able to interact with a precise target, H3K9Me$_3$, which acts as an allosteric regulator of Tip60 catalytic activity instead of a binding motif for the recruitment of Tip60. The example of Tip60 supports a role of MOFs chromodomain in dosage compensation by affecting the activity of HAT towards H4K16, and the possibility that the chromodomains of MOF and MSL3 may interact with methylated histone tails.

The function of the MRG domain of MSL3 is thought to be involved not only in the integration of MSL3 into the DCC but also in the regulation of the HAT activity of MOF towards H4K16. Protein interaction studies between MSL3 and MSL1, both expressed using baculovirus, showed that the MRG domain was required (Morales *et al*., 2005). Studies on

MSL3 without the MRG domain, by immunostaining and RNA fluorescence *in situ* hybridization (FISH), revealed that the MRG domain enabled MSL3 to specifically target the male X chromosome (Buscaino *et al.*, 2006). In the absence of MSL3, MOF's HAT activity was determined to be significantly lowered and not specific to H4K16. Studies with MSL3 possessing only the MRG domain, or lacking the chromodomain revealed that acetylation of K16 on H4 was similar to wild-type levels, indicating that the MRG domain and not the chromodomain was required for proper regulation of MOF's HAT activity (Morales *et al.*, 2005). The crystal structure of the MRG domain of MRG15 from *Homo sapiens* has been solved, which revealed a domain comprised of eight alpha helices (α1, α2, α3, α4, α5, α6, α7, and α8) (Fig. 1.9) (Bowman *et al.*, 2006). Moreover, the core of the domain consists of two orthogonal helix hairpins formed by two sets of two antiparallel helices (α2 & α3 and α5 & α6 - 8). The α6 - 8 helix of MRG domain contains a conserved hydrophobic patch thought to participate in protein-protein interaction. The sequence identity between the MRG domain of *Homo sapiens* MRG15 and the MRG domain of MSL3 is 25 % it is likely that the MSL3 MRG domain has a similar arrangement of helices (secondary structure) in its tertiary structure (Fig. 1.9).

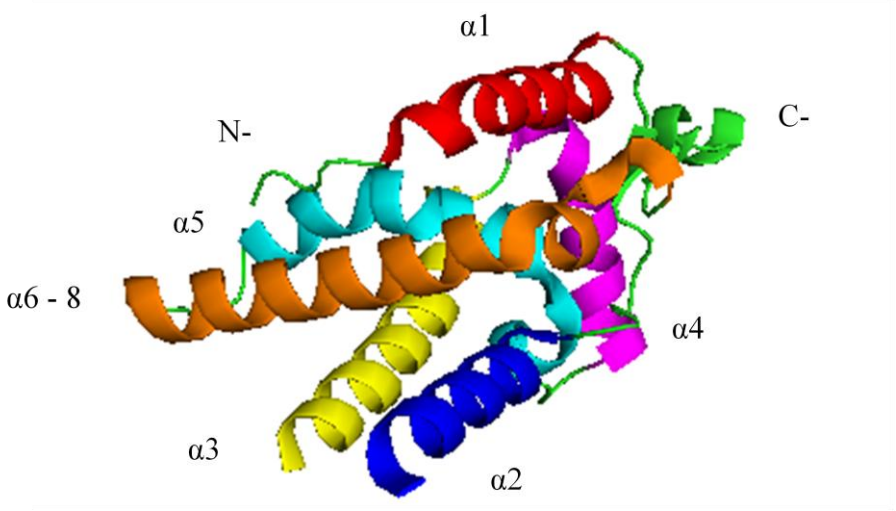


Figure 1.9: Structure of the MRG domain of MRG15 from *Homo sapiens* (PBD accession id: 2F5J). The structure contains eight unique α-helices (α1, α2, α3, α4, α5, α6, α7, and α8), and two orthogonal helix hairpins formed by two sets of two antiparallel helices (α2 and α3 and α5 and α6 -8). The helices are respectively coloured: α1 (red), α2 (blue), α3 (yellow), α4 (magenta), α5 (cyan), and α6 - 8 (orange).

Moreover, sequence alignments between the MRG domains of MRG15 to that of MSL3 revealed that helices α1, α5, α6 -7 are well conserved, α2 and α3 are moderately conserved, and α4 and α8 are not conserved (refer to section 3.4).

1.5.3 Localization of the DCC along the male X chromosome

The process of how the DCC specifically mediates its targeting of actively transcribed genes along the male X chromosome remains largely unknown. Three hypotheses have been proposed for how the DCC localizes to the X chromosome: chromatin entry site (CES), multiple-binding-sites, and transcription activation. The CES theory proposes that the 150 highly reproducible CESs enables recruitment of the DCC to the male X chromosome (Larschan *et al*., 2007; Legube *et al*., 2006). A CES retains MSL1 and MSL2 binding even in the absence of MSL3, MOF, and MLE. The CES may regulate recruitment, as the DCC was observed to associate with various CESs when inserted into autosomes (Kageyama *et al*., 2001; Kelley *et al*., 1999; Larschan *et al*., 2007; Park *et al*., 2003). The multiple-binding-sites model proposes that there are many binding sites with a wide range of binding affinities all along the male X chromosome. The observation that numerous sites along the male X chromosome retain a high affinity for MSL1, MSL2, or roX RNA when one of the other proteins of the DCC was mutated led to the proposition of an endogenous affinity spectrum along the X chromosome (Demakova *et al*., 2003). Also, in studies where the level of MSL2 was varied, a unique distribution pattern along the X chromosome was observed, indicating a hierarchy of targeting sites. The transcription activation model relies on active transcription of genes on the X chromosome. The model proposes that the act of transcription itself signals the recruitment of the DCC along the male X chromosome. The transcription activation model is supported by the observation that the MSL proteins bind to a transgene when they are inserted into a interband, a region of active transcription, along the X chromosome (Legube *et al.*, 2006). Moreover, it was shown that genes on the X chromosome actively being transcribed, in a number of cell lines, were enriched with H4K16Ac at the 5' end of the respective gene loci (Gelbart *et al.*, 2009). In addition, colocalization of the DCC and H4K16Ac was observed at the 3' and middle regions of active genes. In conclusion, the distribution of the DCC cannot be

explained simply by a single principle, and is explained more accurately by all three models in a multimodal mechanism (Straub *et al.*, 2008).

1.5.4 Localization of the DCC requires particular protein domains

A supplementary model of DCC assembly is described as "cotranscriptional assembly" at the *roX1* and *roX2* loci; the individual protein subunits of the DCC recognize the non-coding RNA transcript as its actively being transcribed by RNA polymerase II, functioning as a beacon tethering the protein subunits and driving assembly of the complex (Kelley *et al.*, 2008). An important note is that both roX loci have been categorized as a CES. The properties of the two RB domains of MLE were recently examined in a study researching the recombinant MLE proteins purified from insect cells; the study found that the RB2 domain is dominant for interaction with roX RNA and its subsequent integration into the DCC (Izzo *et al.*, 2008). Removal of the RB2 domain was shown to result in a MLE protein that could not strongly interact with RNA; even though the glycine rich region was determined to aid in RNA binding. The RB1 domain was found to be involved in targeting the MLE to the X chromosome and not for interacting with RNA. Interestingly, MLE lacking the RB1 domain and the glycine rich region was determined to possess a heightened helicase activity. A GKT to GET mutation of the ATPase site A within the ATPase/helicase domain, identified as MLE^{GET}, resulted in a recombinant protein that no longer retained NTP hydrolysis activity and effectively prevented the helicase activity on both RNA and DNA (Lee *et al.*, 1997). The MLE^{GET} mutant not only maintained co-localization with MSL1 on the X chromosome, but it also led to significant redistribution towards autosomes. Furthermore, immunofluorescent staining of cells expressing the MLE^{GET} mutant revealed that MSL1, MSL3, and MLE mostly colocalized to the CESs along the X chromosome; however, roX1 RNA transcripts were also found throughout the nucleus, indicating an improper incorporation of roX1 RNA into the DCC (Gu *et al.*, 2000). Mutation of the DEAD box within the helicase domain from DEIH to DQIH revealed a MLE protein, termed MLE^{DQIH}, which had a high variability in X chromosome binding, and a significant inconsistency of the distribution of MSL1 along the male X chromosome (Richter *et al.*, 1996). The MLE^{GET} mutant and the MLE^{DQIH} mutant collectively revealed the importance of the ATPase/helicase domain for proper dosage compensation.

Explanation of how the DCC is targeted to the male X chromosome is further complicated by the ambiguity of DNA recognition sequences. The absence of any obvious DNA sequences encoding a DCC binding site may be the result of degenerate DNA elements, which were therefore, lost over time (Gilfillan *et al.*, 2006). Insight into how MSL1 regulates DCC assembly along the male X chromosome was observed by the deletion of amino acid residues 1 – 84, which resulted in an MSL1 protein that no longer bound to the male X chromosome (Fig. 1.1B) (Scott *et al.*, 2000). Further deletion studies in transgenic flies showed that the removal of amino acid residues 1 – 26 resulted in an MSL1 protein which only interacted with the CESs (Li *et al.*, 2005). Moreover, point mutation of three conserved basic amino acids, K3, R4, and K6, all mutated to alanine resulted in a MSL1 protein which was only found to bind to five of the CESs. Mutation of two conserved aromatic acids within the basic region, F5A and W7A, produced a protein that lost specificity for the X chromosome, and bound to significantly more autosomal sites. Therefore, it was determined that, amino acids 1 – 84 were required to interact with the X chromosome, the three conserved basic residues modify CES binding, and the two conserved aromatic residues implicitly distinguish the X chromosome from autosomes. In summary, the N-terminal domain of MSL1 is the region required for interaction with the X chromosome and specific regions within N-terminus function in a cooperative mechanism to identify and properly bind to specific sites along the X chromosome.

MSL2 cooperates with MSL1 to stably associate with the male X chromosome as deletion of the amino acid residues 19 – 140, including the RING finger, inhibited MSL2 and MSL1 from binding to the X chromosome (Kelley *et al.*, 1997; Lyman *et al.*, 1997; Zhou *et al.*, 1995). Therefore, the RING finger must be important in the interaction between MSL1 and MSL2 and interaction with the X chromosome. The CXC motif is implicated in affecting the association with genomic DNA as recombinant MSL2 lacking the CXC motif had a drastic reduction in its affinity for dsDNA (Fauth *et al.*, 2010).

MSL3 was also shown to interact with both free and nucleosomal DNA, which were immobilized on paramagnetic beads, suggesting MSL3 may also be important in targeting the DCC to the male X chromosome (Morales *et al.*, 2004). *In vitro* nucleic acid binding experiments addressing RNA or DNA interaction, where a competitor was included, illustrated MSL3 preferentially interacts with RNA over DNA (Akhtar *et al.*, 2000; Morales *et al.*, 2005).

Experiments with recombinant MSL3, comprised of residues 1 – 259 including the chromodomain and part of the MRG domain, was found to interact with RNA, interestingly with a binding affinity greater than full-length MSL3 (Morales *et al.*, 2005). Moreover, it was determined that residues 1 – 140 of MSL3 containing the chromodomain could not interact with RNA (Morales *et al.*, 2005). However two subsequent studies, using EMSAs, demonstrated the ability of MSL3 to interact with DNA and unmodified recombinant nucleosomes and required the presence of the chromodomain (Buscaino *et al*., 2006; Sural *et al*., 2008). Findings from our lab demonstrated recombinant MSL3 comprised only of amino acid residues 1 – 90 strongly interacted with nucleic acid (personal communication with Dr. Moore).

1.5.5 Distribution of the DCC along the male X chromosome

The distribution patterns of the DCC along the X chromosome have also been studied with chromatin immunoprecipitations hybridized to DNA microarrays of *Drosophila* polytene chromosomes (ChIP-chip) (Alekseyenko *et al.*, 2006; Gilfillan *et al.*, 2006; Legube *et al.*, 2006). The distribution of MSL3 along the male X chromosome from embryonic SL2 cells, larval wing imaginal disc cells, and late embryos were examined (Alekseyenko *et al.*, 2006). The distribution of MSL1 along the X chromosome of embryos was analyzed (Gilfillan *et al.*, 2006). Lastly, the distribution pattern of both MSL1 and MSL3 on the both the male and female X chromosomes from 16 hrs embryos, 4 – 6 hrs early embryos, and salivary glands from third instar larvae was investigated (Legube *et al.*, 2006).

The DCC has been identified to interact selectively in a discontinuous pattern with greater than half of the genes, between 600 – 900 of the 1389 known genes, encoded in the X-chromosome (Alekseyenko *et al.*, 2006; Gilfillan *et al.*, 2006). Discrepancy in determining the precise number of genes along the X chromosome that the DCC interact with is rooted in the experimental procedure including cell type, developmental stage, and the protein of the DCC that is tested MSL1 versus MSL3. The types of gene product, whether lethal or non-lethal, bound by the DCC (MSL1 or MSL3) was not indicative of DCC association (Legube *et al.*, 2006). Contrarily, it was shown that MSL1 interacts with more genes encoding lethal proteins than non-lethal genes (Gilfillan *et al.*, 2006).

MSL1 ChIP hybridization to custom microarrays revealed an almost complete bias of MSL1 binding to exons; moreover, within exons MSL1 bound to coding regions rather than of the 5' or 3' untranslated regions (Gilfillan *et al.*, 2006). Within the coding regions of exons MSL1 displayed a clear preference for the 3' end. ChIP-chip assays with a modified MSL3 tagged protein showed that 90% of the DCC binding clusters were within expressed genes with a preference for the middle and the 3`end of the genes (Alekseyenko *et al.*, 2006). Analysis of ChIP-chip studies investigating colocalization of MSL1 and MSL3 found that 94 % of the binding sites were the same (Legube *et al.*, 2006). Localization of MSL1 and MSL3 to the 3' end of genes led to the proposition that the DCC's mechanism of action takes part in transcription elongation or termination; in a similar way as transcriptional elongation factors. Recent publications on exon-intron structure argues that the average length of exons in eukaryotes coincides with the average 147 nucleotides wrapped around the histone octamers (Schwartz *et al.*, 2009). Moreover, in *Drosophila* it was observed that exons contain a higher GC content compared to flanking introns. The significance of the GC content is due to the correlation between nucleosome occupancy and exons. Because exons are associated with the NCP it was reasoned they may possess specific modifications of histones to mark the exons; it was determined that histone H3K36 tri-methylation ($H3K36Me_3$) was the most prominent modification, but also mono-methylated H4K20, H3K79, and H2BK5 was observed. In *Saccharomyces cerevisiae* and in *Drosophila* $H3K36Me_3$ was found to be enriched at the 3' end of exons (Bell *et al.*, 2008). Moreover, the $H3K36Me_3$ modification in *Homo sapiens*, *C. elegans*, and again in *Drosophila melanogaster* was consistently observed to at the 3' end of exons (Kolasinska-Zwierz *et al.*, 2009; Schwartz *et al.*, 2009). $H3K36Me_3$ is a reported mark of transcription elongation. The association of MSL1 and MSL3 to the 3' end of exons led to the postulation that elevated gene expression of the male X chromosome by the DCC may not be involved in the initiation of transcription, but instead the promotion elongation and termination of transcription. Acetylation of H4K16 along the male X chromosome is also found at the 3' end of exons (Alekseyenko *et al.*, 2006; Gilfillan *et al.*, 2006; Kind *et al.*, 2008).

A possible link between active transcription and the DCC was investigated through examining if MSL1 and RNA polymerase II colocalize along the male X chromosome. RNA polymerase II was observed to interact with approximately 65 % of the genes of the X-

chromosome, and with 87 % of the genes bound by MSL1. However, the binding profiles were rarely similar, indicating that transcription alone is inadequate for the recruitment of the DCC (Alekseyenko *et al.*, 2006). Experiments with DNA microarrays found that 87 % of genes bound by the DCC were actively transcribed; furthermore, 50 % of actively transcribed genes were also bound with the MSL3-tagged protein, whereas only 8 % were unbound. The effect of the DCC bound to genes actively being transcribed were studied through MSL2 RNAi knockdown. Analysis of the microarrays found that genes associated with the DCC displayed a significant decrease in gene expression after treatment, whereas the expression levels of genes not associated with the DCC were less affected. A weak correlation was determined between active transcription and MSL1 as only 15 % of the actively expressed genes encoded in the X-chromosome were bound by MSL1 in larvae salivary glands of 4 – 6 hour embryos when comparing ChIP-chip data and the expression data from the Yale *Drosophila* lifetime course (Legube *et al.*, 2006). These findings indicated that expression does not necessarily lead to MSL1 binding. The decreased requirement of the DCC may instead be due to the maternal stockpiles of mRNA, which are critically involved in the early stages of embryogenesis (Kugler and Lasko, 2009). Polytene chromosomes from salivary glands were immunostained for MSL1, RNA polymerase II, and two elongation factors and no apparent association between active transcription and MSL1 binding was observed (Legube *et al.*, 2006). Therefore, in DCC binding and recruitment transcription is important but not essential.

Expression patterns among different cell types of *Drosophila* and the respective DCC bindings were studied; it was determined that approximately 600 genes were commonly bound by the DCC in embryonic cells, larval wing derived cells, and transgenic embryos, ranging between 12 and 17 hours (Alekseyenko *et al.*, 2006). Therefore, common DCC binding sites exist in different cell types. In addition to DCC binding patterns between different cell types the DCC binding patterns between different developmental stages were studied. Interestingly, the number of genes targeted by MSL1 between early embryogenesis and later developmental stages increased, indicating that the DCC played a greater and greater role as *Drosophila* progress through the developmental stages (Legube *et al.*, 2006). Moreover, association of a specific gene and the DCC appears to be an inherent property; in other words, the specific MSL1 binding patterns seen in previous developmental stages remain in subsequent stages. Research showed 40 % of the MSL1 binding sites in the late embryonic stage, are also bound

sites in the early larvae stage. Interestingly, 40 % of the DCC targeted sites in late larval stage X chromosome comprised 83 % sites targeted along the X chromosome of the early larval stage. In this sense, sequences may have evolved along commonly expressed genes; however, identification of 14 genes which were clearly bound in certain cell types were not bound by the DCC in other cell types. These results indicate that sequence alone is not an absolute in DCC binding and recruitment (Legube *et al.*, 2006).

In order to investigate the reported CESs the binding pattern of MSL2, from embryos possessing mutated MSL3, was studied with ChIP-chip (Alekseyenko *et al.*, 2008). Moreover, the MSL1 and MSL2 distribution along the X chromosome from SL2 cells possessing reduced MSL3, MLE and MOF by RNA interference was examined with ChIP-chip (Straub *et al.*, 2008). High resolution ChIP-chip studies significantly increased the number of reported CESs (35 – 45 sites based on cytological studies of polytene chromosome squashes) along the male X chromosome; studies using SL2 cells identified 131 CESs, and studies using male embryos identified 150 CESs (Alekseyenko *et al.*, 2008; Straub *et al.*, 2008). Current research argues a CES is defined as a site targeted by MSL1-MSL2 in the absence of MSL3, MOF, or MLE (Alekseyenko *et al.*, 2008; Demakova *et al.*, 2003; Gelbart *et al.*, 2009; Straub *et al.*, 2008). The average size of the 150 CES were determined to be 1.5 kb in length. Of the 150 CES, 135 overlapped with the position of known genes (Alekseyenko *et al.*, 2008). These studies lead to the identification of CES to be enriched with GA and CA dinucleotides, termed a MSL recognition element (MRE). When the GA or CA dinucleotide motif or a MRE was expanded to 10 nucleotides, only 40 % of the CES contained the DNA sequence motif. Therefore, a large percentage of the recruitment of the DCC may be sequence independent (Gelbart and Kuroda, 2009; Straub and Becker, 2008). Surprisingly, when the MRE was inserted into an autosome, a novel DCC binding site independent of MSL3 was produced. The clustering of weak targeting elements can function collectively to produce a strong binding site in *Drosophila* (Straub *et al.*, 2008). Interestingly, within all of the 150 CESs the MRE region was depleted of histone H3 proteins; suggesting that the absence of nucleosomes is fundamental to a CES; it is thought that this would enable quick and easy recruitment and assembly of the DCC (Alekseyenko *et al.*, 2008).

1.5.6 Interplay between H3K36Me$_3$ and the DCC

H3K36Me$_3$ is a chromatin modification that correlates with the recent transcriptional elongation of a gene. Control of transcription was originally thought to be essentially only regulated by the recruitment of RNA polymerase pre-initiation complexes; however, a novel method has emerged to affect transcription by means of protein complexes that modify chromatin conformation. Since H3K36Me$_3$ is enriched at the 3' end of genes actively being transcribed in yeast, it may be a possible chromatin modification, which has the ability to recruit the DCC in *Drosophila* (Larschan *et al.*, 2007). Huntingtin interacting protein B (Hypb) or Set2 was determined to be an essential HMTase because knockout resulted in lethality in both male and female *Drosophila*. Hypb has been identified to catalyze the H3K36Me$_3$ chromatin modification along the male X chromosome (Bell *et al*., 2008; Larschan *et al*., 2007). A significant reduction of the posttranslational modification H4K16Ac and H3K36Me$_3$ along the male X chromosome was observed when Hypb was severely reduced by RNAi interfence, this suggests the DCC is affected by Hypb activity. With ChIP-chip profiling experiments it was determined that MSL3 and H3K36Me$_3$ colocalized at 93 % of the DCC binding sites, and at 83 % of the binding sites, MSL3 and H3K4Me$_2$ colocalized (Larschan *et al.*, 2007). A homologue of *Drosophila* MSL3 the Esa1p-associated factor-3 protein (Eaf3p), part of the Rpd3(S) histone deacetylase complex, of *Saccharomyces cerevisiae* has been found to be functionally linked with H3K36Me$_3$. The chromodomain of Eaf3p was determined to bind H3K36Me$_3$, which then recruits the histone deacetylase Rpd3(S) complex. It is thought that the Rdp3(S) complex removes transcription-coupled hyperacetylation and in doing so it inhibits unwanted RNA polymerase II initiation. Therefore, the colocalization of the DCC and the histone modification, H3K36Me$_3$ both appear to be part of the process though which DCC recognizes actively transcribed genes.

The role of the MSL3 chromodomain and its ability to interact with methylated histone tails, specifically nucleosomes with H3K36Me$_3$ was investigated (Sural *et al.*, 2008). It was determined that full length wild-type MSL3 had a higher affinity for H3K36Me$_3$ modified nucleosomes compared to unmodified nucleosomes. The MRG domain of MSL3 was shown again not to interact with nucleosomes and MSL3 lacking the chromodomain was determined to not interact with both modified and un-modified nucleosomes. Interestingly, studies on the DCC binding pattern where MSL3 lacked its chromodomain, or possessed the SYD62A

mutation within the chromodomain it was demonstrated that the DCC only interacted with the CESs, supporting the importance of the chromodomain of MSL3 in interaction with other regions along the male X chromosome (Buscaino *et al.*, 2006; Sural *et al.*, 2008). However, our lab examined recombinant MSL3 expressed in *E. coli* and found it did not interact appreciably with $H3K36Me_3$ (personal communication with Dr. Moore).

1.5.7 Spreading of the DCC along the X chromosome

Within the mechanisms proposed for proper dosage compensation is first the recruitment of the DCC to male X chromosome and subsequent bi-directional but non-continuous spreading of the DCC to the regions along the male X chromosome. Bi-directional spreading of the DCC from a known entry site was first observed when the *roX* loci were inserted into an autosome (Kelley *et al.*, 1999).

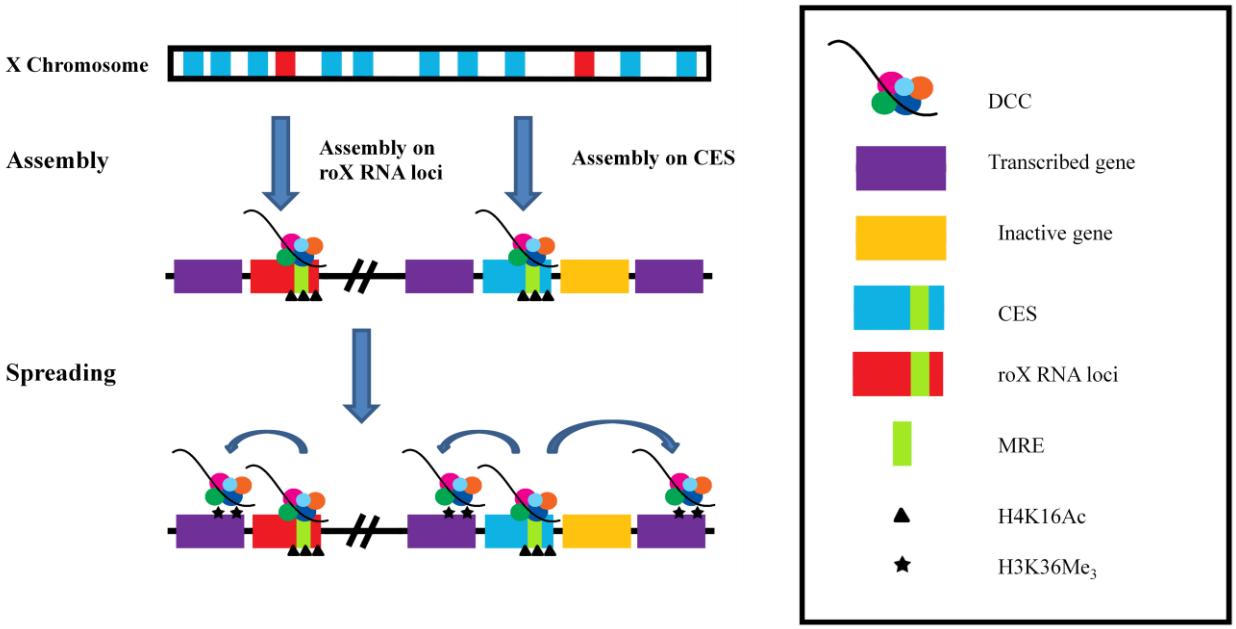


Figure 1.10: Proposed mechanism of how the DCC is recruited to and subsequently spreads along the male X chromosome. The DCC members specifically associate along the male X chromosome assembling at the CESs. The assembled DCC acetylates H4K16 at the CESs, and then spreads to neighbouring exons along the X chromosome targeted by $H3K36Me_3$. See text for further details, sections 1.5.5 – 1.5.7.

The *roX* loci are known to contain DNase I hypersensitive CESs for the DCC, and both *roX1* and *roX2* are found on the X chromosome, suggests their importance in both recruitment and spread of the DCC along the X chromosome. Moreover, bi-directional spreading aids in explaining how the observed wide spread distribution of DCC is obtained along the male X chromosome in accordance with the CES model and the multiple-binding-sites model (Kelley *et al.*, 1999; Oh *et al.*, 2003; Park *et al.*, 2002). Furthermore, the involvement of DNA sequence elements in spreading of the DCC was observed when a candidate CES possessing a MRE motif was inserted at an ectopic autosomal site, which resulted in recruitment and subsequent bi-directional spreading of the DCC from the autosomal site (Alekseyenko *et al.*, 2008). Interestingly, the rate of spreading appeared to be correlated to the rate of transcription of the roX RNA; as elevated levels of transcription inhibited local spreading and localization of the DCC at other sites on the X chromosome suggesting active transcription at other sites may prevent spreading along the male X chromosome (Kelley *et al.*, 2008). However, active transcription of other genes on the X chromosome, enriched with the chromatin modification H3K36Me$_3$ at the 3' end, may aid in targeting the DCC to those genes along the X chromosome. A recent study proposed a mechanism to explain how the DCC spreads along the X chromosome (Fig. 1.10) (Sural *et al.*, 2008). The initial step is the assembly of the DCC at the CES, which coincides with the CES model. The adjacent regions of the CESs contain H3K36Me$_3$ which then recruits the DCC away from the CESs. The vacated CES then enables assembly of another DCC. This continuous assembly and spreading results in the enrichment of H4K16Ac along the male X chromosome resulting in the two-fold transcriptional elevation of the genes encoded in the X chromosome.

1.6 Homologues of *Drosophila* dosage compensation proteins

Comparative genomics has identified a number of DCC orthologues in vertebrate species, suggesting that the DCC protein sequences have been conserved through evolution (Marin, 2003). All of the proteins involved in the DCC have homologues in *Homo sapiens* (Taipale and Akhtar, 2005). The known protein domains within the individual proteins of the *Drosophila* DCC except MLE were all found within their respective *Homo sapiens* homologue (Mendjan *et al.*, 2006). Experiments using co-immunoprecipitation identified a HAT complex consisting of hMSL1, hMSL2, hMSL3, and hMOF (Smith *et al.*, 2005). The complex was

confirmed by use of co-immunoprecipitation and tandem affinity purification, determining that targeting any of the four proteins resulted in co-purification of the remaining three members. Moreover, *in vitro* pull-downs determined hMOF and hMSL3 stably associated with one another, suggesting a direct interaction (Taipale *et al*., 2005). Immunoprecipitations consistently revealed two hMSL3 isoforms; surprisingly the shorter splice isoform was determined to not possess the chromodomain. Since the *Homo sapiens* complex acetylates the identical target of *Drosophila* MOF, H4K16, it is possible that the DCC is also highly conserved through evolution (Smith *et al.*, 2005).

Paralogues of *Drosophila* MSL3 have been identified with chromatin remodelling complexes, analogous to the *Drosophila* DCC all of these complexes also contain a MYST family HAT (Buscaino *et al.*, 2006). Eaf3p, the MRG15 orthologue in budding yeast, is part of the NuA4 complex containing the MYST HAT protein Esa1. In addition, MRG15 associates with the Tip60 complex in *Homo sapiens*, which includes the MYST HAT protein Tip60.

The acetylation activity of hMOF akin to dMOF has been shown to be specific H4K16 (Taipale *et al*., 2005). hMOF has been identified in additional complexes including the mixed lineage in leukemia 1 (MLL1) complex and the non-specific lethal (NSL) complex (Cai *et al.*, 2009). Other MYST HAT members found in humans include: HBO, INO, MOZ, MORF, and Tip60 (Voss and Thomas, 2009). RNAi knockdown of hMOF in HeLa cells caused a decrease in acetylation of H4K16 and the cells to arrest their cell cycle at the G(2)/M phase; therefore it may have a direct role in regulating cell cycle checkpoints (Gupta *et al.*, 2005; Smith *et al.*, 2005; Taipale *et al.*, 2005).

A stable interaction between hMOF and MLL1 has been mapped to the zinc finger of MOF and the C-terminal region of MLL1 (Dou *et al.*, 2005). The complex maintains MLL driven histone methyltransferase activity for H3K4 and MOF and mediates histone acetyltransferase activity for H4K16. Interestingly, the complex *in vitro* was shown to enable optimal activation of transcription on a chromatin template. hMOF and MLL1 complex were determined to associate with the ATM, a protein involved in DNA repair (Gupta *et al.*, 2005).

Affinity purification, coupled with mass spectrometric analysis, identified an enzymatically active NSL complex purified from human HeLa cells (Mendjan *et al.*, 2006). The interactions between hNSL1 (coiled coil, PEHE domains), hNSL2 (two C/H-rich domains), hNSL3 (seven WD40 repeats) and hMOF were confirmed. Furthermore, the NSL

complex was isolated from *Drosophila* embryos and Schneider cells, indicating an evolutionarily conserved complex from flies to humans. The NSL complex was found to be comprised of 9 subunits, including WD repeat domain 5 (WDR5), host cell factor 1 (HCF1), and MCRS1 (Cai *et al.*, 2009). The hMOF that associated with NSL complex was determined to acetylate not just H4K16 but also K5 and K8 on H4. The MCRS1 protein has been previously found in the humans INO chromatin remodelling complex.

hMLE orthologues have been linked to diverse roles in other eukaryotes. hMLE has been identified as a nuclear DNA helicase II or RNA helicase A (Lee and Hurwitz, 1993). Studies on MLE orthologues RNA helicase A of *Homo sapiens* and nuclear DNA helicase II of *Bos taurus* have led researchers to postulate on the function of MLE. Sequence alignments between MLE and RNA helicase A showed a 50% identity and alignment of MLE with RNA helicase II a 85% similarity (Lee *et al.*, 1997; Lee and Hurwitz, 1993; Richter *et al.*, 1996). Evidence has shown hMLE to be involved in multiple features of cellular and viral DNA and RNA metabolism (Aratani *et al.*, 2008). Moreover, it has also been shown to recruit the RNA polymerase II complex to promoters, activate transcription through interaction with cis-acting transcription factors and their promoters, and mediate transactivations through interaction with specific nuclear receptors.

1.7 Thesis Objectives and Hypothesis

The key function of the DCC is to specifically target the acetylation activity of MOF to the H4K16 along the male X chromosome (Gu *et al*., 1998; Hilfiker *et al*., 1997; Smith *et al*., 2000). Therefore, modulating activity and specificity of MOF's acetylation activity is pivotal for survival of male *Drosophila*. The C-terminus of MSL1, consisting of the PEHE domain, co-immunoprecipitates both MOF and MSL3 from *Drosophila* cell lysates (Scott *et al.*, 2000). The MRG domain of *Drosophila* MSL3 comprised of amino acid residues 185 – 490, or 141 – 490 was found to interact with residues 973 – 1039 of MSL1 (Morales *et al.*, 2005). Furthermore, amino acid residues 766 – 939 of MSL1 were determined to interact with MOF (Morales *et al*., 2004). The CCHC zinc finger and the MYST HAT domain are both required to interact with MSL1; however, the exact amino acid residues of MOF which interact with MSL1 remain unknown. Moreover, MSL3 and MOF do not directly interact, but are thought to interact once both are bound to MSL1 (Buscaino *et al*., 2003; Morales *et al*., 2004).

The purpose of this project was to gain a better understanding of the protein interactions between MSL1, MSL3, and MOF. Utilization of recombinant proteins and *in vitro* protein interaction studies enabled investigation of the protein complex. The project pursues four main objectives. To start we will need to construct a MSL1 C-terminal domain that can be over-expressed and purified to homogeneity. Also, to study the interactions of the human MSL1 components, it is necessary to generate a *Homo sapiens* MSL1 C-terminal construct that can be over-expressed and purified. Studying the interactions of the DCC components will also require the purification of two other proteins of the DCC: the MRG domain of MSL3 and the catalytic domain of MOF (chromodomain and MYST HAT domain). Characterizing protein interactions first requires ascertaining the stability and purity of the recombinant proteins by biochemical analyses including: characterization of the protein's structure using calibrated size exclusion chromatography, Far-UV spectrophotometry, and other techniques. It was hoped we could demonstrate interaction between the three DCC components using various combinations of the purified recombinant domains. Protein-protein interaction studies will be carried out with GST pull-downs and size exclusion chromatography. An important part of this work will be to obtain structures for the C-terminal domain of MSL1 and the MRG domain of MSL3 or any of the relevant complexes with MOF. Crystallization trials will be carried out on the recombinant proteins, and if possible co-crystallization trials of the complex for future structure determination by X-ray crystallography. Accomplishment of these goals will allow progress toward comprehending dosage compensation, and will also facilitate future structural studies.

2. Materials and Methods

2.1 Cloning of *Drosophila Melanogaster* MSL1, MSL3, and MOF

Table 2.1: List of the PCR primers for the relevant *Drosophila* MSL1, MSL3, and MOF gene fragments cloned in this work.

Recombinant Protein	Primer Sequence	Restriction Endonuclease
dMSL1	(AJW009) 5′-GACCTG*AGATCT***TCGGGAAAACTGCAGTCCAGAT**-3′	*Bgl*II
(820 – 1039)	(KK012) 5′-GACGAG*GAATTC***CTAACGATTCTTCTGGCGCTT**-3′	*Eco*RI
dMSL1	(KK011) 5′-GCCGAG*AGATCT***GGCTCAACGCCACAGCATGCG**-3′	*Bgl*II
(754 – 1039)	(KK012) 3′-CTCGTC*GAATTC***CTAACGATTCTTCTGGCGCTTGCG**-5′	*Eco*RI
dMSL1	(KK010) 5′-GCCGAG*AGATCT***TCACAGGTTACGCTAAGAAAATAAGAGAG**-3′	*Bgl*II
(712 – 1039)	(KK012) 3′-CTCGTC*GAATTC***CTAACGATTCTTCTGGCGCTTGCG**-5′	*Eco*RI
dMSL1	(KK009) 5′-GCCGAG*AGATCT***CAGAGCCCAGATCAAGAAATAGATGTG**-3′	*Bgl*II
(688 – 1039)	(KK012) 3′-CTCGTC*GAATTC***CTAACGATTCTTCTGGCGCTTGCG**-5′	*Eco*RI
dMOF	(SAM040) 5′-GACCTG*GGATCCC***AAAAGATCGATATAAGCGAAAATCCC**-3′	*Bam*HI
(370 – 827)	(SAM043) 5′-GACGAG*GAATTC***CTAGCCGGAATTTCCCGGAGCTCT-3’**	*Eco*RI
dMSL3	(BJM007) 5′-GACCTG*GGATCC***AATGATGTATCGGTCTATAATCATGTG** -3′	*Bam*HI
(185 – 512)	(BJM004) 5′-GACGAG*GAATTC***CTAAGCAGCAATACCATCCAGGGA** -3′	*Eco*RI

Note that in the synthetic oligonucleotides the clamps are underlined, the endonuclease restriction sites are italicized, and the complementary sequences are depicted in boldface.

2.2 Cloning of *Homo sapiens* MSL1

Table 2.2: List of the PCR primers for the relevant *Homo sapiens* MSL1 gene fragments cloned in this work.

Recombinant Protein	Primer Sequence	Restriction Endonuclease
hMSL1 (404 – 614)	(KK005) 5'- GCCGAG*GGATCC***ATGGAAAGGCGGATGCAGCTGG**-3	*Bam*HI
	(KK008) 5'- GAGGCC*GTCGAC***CTATTTCCTACACGTCGGTGAGG**– 3'	*Sal*I
hMSL1 (364 – 614)	(KK006) 5'- GCCGAG*GGATCC***GAGGAACCCTGTGGTTCC** – 3'	*Bam*HI
	(KK008) 5'- GAGGCC*GTCGAC***CTATTTCCTACACGTCGGTGAGG**– 3'	*Sal*I
hMSL1 (262 – 614)	(KK007) 5'- GCCGAG*GGATCC***GGACCCAGCACCCATCCCAAG**– 3'	*Bam*HI
	(KK008) 5'- GAGGCC*GTCGAC***CTATTTCCTACACGTCGGTGAGG**– 3'	*Sal*I
hMSL1 (465 – 614)	(AJW017) 5'-GCCGAGGGATCC**AGTGTTGCAGGAGAAACTTCAGTC**-3'	*Bam*HI
	(KK008) 5'- GAGGCC*GTCGAC***CTATTTCCTACACGTCGGTGAGG**– 3'	*Sal*I

Note that in the synthetic oligonucleotides the clamps are underlined, the endonuclease restriction sites are italicized and the complementary sequences are depicted in boldface.

2.3 Cloning Procedure

Constructs were individually designed by first searching for homologues and then aligning the protein sequences to identify the conserved regions in the protein sequences. Protein homologues were found through the use of the Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Multiple alignments of the relevant protein sequences were generated using EMBL-EBI ClustalW2 (Larkin *et al.*, 2007) (http://www.ebi.ac.uk/Tools/clustalw/index.html), and adjustments were done manually when necessary. The appropriate gene nucleotide sequences were then used to design the corresponding primers. ExPASy Prot Param (John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005) http://www.expasy.ch/tools/protparam.html) was used to predict the recombinant protein molecular mass, isoelectric point, and extinction coefficient for each construct.

Four unique constructs of MSL1, encompassing the C-terminal domain, from both *Drosophila melanogaster* and *Homo sapiens* were designed (Table 2.1 & 2.2). A MSL3 construct and a MOF construct from *Drosophila melanogaster* were also designed (Table 2.1).

All constructs were designed to be inserted into the multiple cloning sites (MCS) of pGEX-6P-3 downstream of the glutathione S-transferase gene and the PreScission protease cleavage site (Fig. 2.1).

2.3.1 Overview of pGEX-6P-3 vector

The utilization of the pGEX-6P-3 expression vector (GE Life Science) enabled the cloning of a variety of protein constructs, transformation of *E. coli*, and selection of transformants by the appropriate antibiotic (ampicillin). Insertion of a gene fragment of interest into the MCS of the vector enables expression of an in-frame GST-fusion protein. The recombinant GST tag, composed of 228 amino acids, contains a PreScission protease cleavage site at the 3' end of GST DNA, upstream of the MCS. PreScission protease is a recombinant fusion protein comprised of GST and human rhinovirus (HRV) type 14 3C protease. The protease is specific for the core amino acid sequence Leu-Phe-Gln/Gly-Pro and cleaves between the Gln and the Gly residue, which has been engineered into the pGEX-6P-3 MCS. The GST tag permitted purification of the recombinant protein by means of Glutathione Sepharose (GS) affinity chromatography.

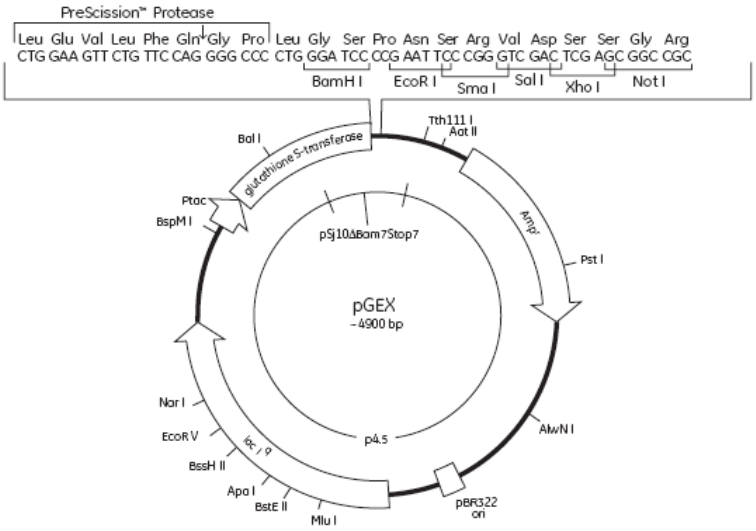


Figure 2.1: Illustration of the pGEX-6P-3 vector. The MCS, the GST gene, the Amp restistance gene, the lac I gene, the pBR322 ori, and the endonuclease restriction sites are all shown. The figure was adapted from the Recombinant Protein Handbook (Amersham Biosciences).

2.3.2 DNA amplification by Polymerase Chain Reaction

Each gene fragment of interest was generated by DNA polymerase chain reaction (PCR). All of the PCRs were done using 2 ng of cDNA template diluted into 50 uL of 1X PCR reaction mixture. The 1X PCR reaction mixture contained 1X PCR reaction buffer (200 mM Tris-HCl [pH 8.8 at 25 °C], 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1 % [v/v] Triton X-100, 1 mg/mL BSA, 1.5 mM $MgSO_4$), 0.5 μM 5' forward primer, 0.5 μM 3' reverse primer, 200 μM dNTPs, 1 μL (5 U) *Pyrococcus furiosus* (*Pfu*) DNA polymerase (Fermentas), and H_2O, for a total volume of 100 μL. Amplifications by PCR were carried out using a MyCycler™ Thermal Cycler (Bio-Rad) for 30 – 35 cycles. All primers were purchased from Sigma Genosys and are listed in Tables 2.1 and 2.2. A representative PCR amplification program is illustrated in Figure 2.2. Note that adjustments were made to facilitate the melting temperature of the primers and the nature and length of the template cDNA.

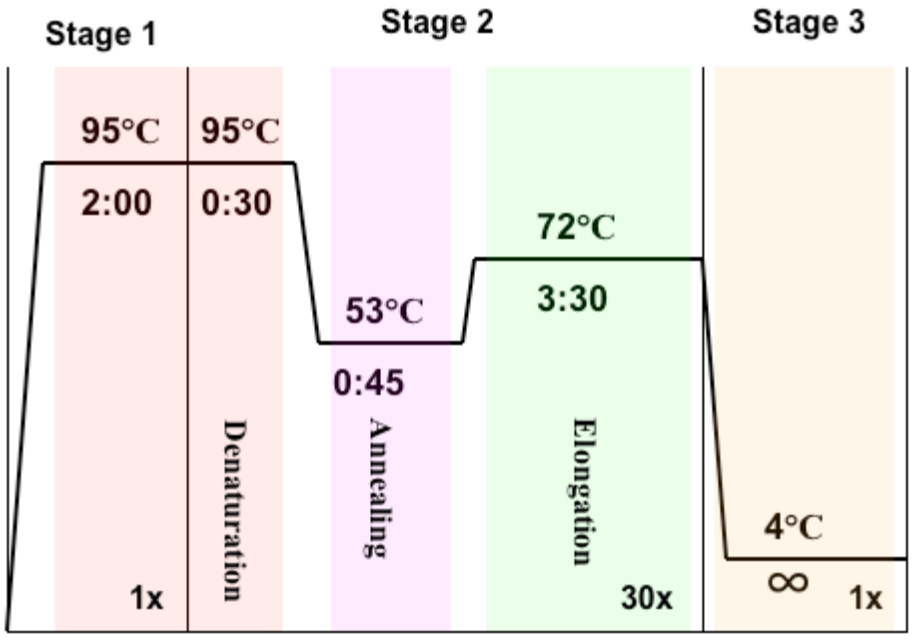


Figure 2.2: Typical PCR amplification program. The program consists of three stages and 30 cycles. Stage 1 is an extended 2 min denaturation step. Stage 2 is made up of the 30 cycles where each cycle includes a 30 s denaturation step, a 45 s annealing step, followed by a 3.5 min. elongation step. Stage 3 holds the reaction mixture at 4 °C indefinitely.

2.3.3 Ethanol Precipitation of PCR DNA

Following PCR, the reaction mixture was purified by ethanol precipitation which was performed by adding 10 % v/v 3 M sodium acetate (pH 5 at 25 °C) to the PCR mixture, followed by a wash of 3X v/v of the PCR mixture with 95% ethanol. The resultant solution was stored at -20 ° C for 10 min, and then pelleted via centrifugation at 17,418 x g (r_{max} 13,000 rpm) for 10 min at 4 °C. The subsequent supernatant was discarded and the pellet was then washed with 2X 70 % ethanol. The remaining liquid within the pellet was removed by rotary-evaporation and then the pellet was dissolved in 20 μL TE buffer (10 mM Tris-HCl, [pH 7.5 at 25 °C] and 1 mM EDTA).

2.3.4 Agarose gel electrophoresis

Agarose gel electrophoresis was the method employed for further purifying the PCR product DNA and estimating the size, quantity, and purity of the desired PCR product, or vector DNA. The agarose matrix was always 1 % agarose (w/v) in 1X TAE solution (40 mM Tris-acetate [pH 8.0 at 25 °C] and 1 mM EDTA). 1X TAE solution was diluted from a 50X stock solution. To dissolve the agarose the mixture was heated for 2 min in a microwave, after which the solution was cooled and 0.5 μg/mL ethidium bromide was added. The solution was poured into an Easycast™ Horizontal Electrophoresis (Owl Separation Systems) gel casting tray containing a 10 well comb. The gel was set for 30 min, the comb was removed, and the gel was placed into the electrophoresis apparatus, and submersed in 1X TAE buffer. DNA samples, ranging between 5 – 100 ng, were combined with DNA loading buffer (0.04% [w/v] bromophenol blue and 5% [v/v] glycerol) to a final concentration of 1X. DNA samples were loaded into the agarose gel wells and resolved at 100 V for 30 – 45 min. Visualization of double stranded DNA was by means of ethidium bromide, which intercalates the nucleic acid and fluoresces under ultra-violet light, 365 nm. Photographs of the resultant agarose gels were taken with a Gel Logic 200 Scientific Imaging Systems (Eastman Kodak Company) and analysed with Kodak Molecular Imaging Software.

2.3.5 DNA analysis: restriction digestion and DNA purification

The cloned insert of interest was analysed by an agarose gel ran with a FastRuler™ DNA ladder (Fermentas) of known size and concentrations. All of the *Drosophila* MSL1

constructs were inserted into the MCS of pGEX-6P-3 by means of the *Bam*HI/*Eco*RI restriction endonuclease sites. However, as the *Drosophila* MSL1 fragments intended to be amplified possessed an internal *Bam*HI site, the insert primers contained *Bgl*II and *Eco*RI restrictions sites to generate complementary ends. In this way, *Bam*HI digested vector can be ligated to *Bgl*II sticky ends on the PCR product. The *Homo sapiens* MSL1 constructs and the pGEX-6P-3 vectors were all digested with *Bam*HI/*Sal*I. The *Drosophila* MSL3 construct and the pGEX-6P-3 vector were both digested with *Bam*HI/*Eco*RI. All restriction endonucleases were purchased from Fermentas or New England Biolabs. Digestions were done at 37 °C for a minimum of 4 hours. Agarose gel electrophoresis permitted qualification of the restriction endonucleases digestions. The DNA fragments (vector and insert) were further purified prior to DNA ligation by employing the Qiagen Gel Extraction Kit, following the manual provided. The DNA was stored in 50 µL of TE buffer.

2.3.6 DNA ligation

Ligation reactions enable the generation of a recombinant pGEX-6P-3 vector containing an in-frame fusion of the gene fragment of interest. Ligations were carried out using minimal volumes, consisting of 1 µL insert (at a 4:1 insert: vector molar ratio), 2 µL vector, 1 µL T4 DNA ligase, 2 µL 5X reaction buffer, and H_2O to a final volume of 10 µL. The ligation mixture was incubated at 4 ^{o}C for 8 hrs using the MyCycler^{TM} Thermal Cycler (Bio-Rad). The correct insert to vector molar ratio that resulted in a suitable number of single colonies resulting from transformation of competent *E. coli* cells was determined empirically for each insert.

2.3.7 Competent Cell preparation

Competent XL-1, Rosetta 2, and BL-21 DE3 (BL-21) *E. coli* cells were prepared in the lab using the following procedure. Cells were first plated onto Luria-Bertani (LB) broth agar plates containing the appropriate antibiotics (if required) and incubated for 14 - 16 hours at 37 °C. Following incubation, a single colony was picked and then inoculated in a 5 mL LB broth, possessing the appropriate antibiotic, and incubated for 14 - 16 hours at 37 °C and 300 rpm. The culture was diluted 1 in a 100 with fresh LB broth, which possessed the appropriate antibiotic. The diluted culture was grown at 37 °C and 300 rpm to an optical density (OD_{600}

nm) of 0.4 to 0.6. Cells were harvested through the use of a Beckman-Coulter Centrifuge; centrifugation was done for 5 minutes at 2,688 x g (r_{max} 4,000 rpm) and 4 °C. The cell pellet was resuspended into ice cold 0.1 M $MgCl_2$ in a 1 to 5 v/v ratio of the original cell volume (100 mL), and incubated on ice for 1 hour. Cells were harvested by centrifugation for 5 minutes at 2,688 x g (r_{max} 4,000 rpm) and 4 °C. The cell pellet was resuspended in 0.1 M $CaCl_2$ plus 14 % glycerol, with a 1 to 50 v/v ratio of the original cell volume (100 mL), and incubated on ice for 1 hour. The cell solution was aliquoted and flash frozen with liquid N_2 and stored at - 80 °C. The transformation efficiency of the competent cells was determined by transformation of 1 μg of vector DNA resulting in either 1 x 10^6 cell/μg DNA for XL-1 cells, or 1 x 10^4 cell/μg DNA for BL-21, and Rosetta 2 cells, each resulting in 10 – 25 individual colonies.

2.3.8 Transformation of *Escherichia coli* cell lines

The entire ligation reaction mixture was mixed with 50 – 100 μL of competent *E. coli* XL-1 cells (stored at -80 °C). The mixture was thoroughly mixed, cold-shocked by incubating the mixture at 4 °C for 45 min, and then plated onto pre-warmed LB plates containing ampicillin (100 μg/mL), and incubated for 14 hrs at 37 °C.

For protein expression, competent *E. coli* Rosetta 2 or BL-21 cells were transformed with the purified recombinant plasmid. Approximately 20 – 50 ng of plasmid DNA was combined with 50 – 100 μL competent cells and cold-shocked for 45 min. The solution was plated onto pre-warmed LB plates containing the appropriate antibiotic(s) and incubated at 37 °C for 14 hrs.

2.3.9 Recombinant screening

Colony PCR was utilized to screen single colonies grown on LB plate, containing appropriate antibiotic(s), for the presence of the correct insert. The initial strenuous denaturing condition (95 °C) of PCR results in the cell to lyse and release of plasmid DNA into the PCR mixture. The PCR mixture contained forward and reverse primers flanking the insert of interest, and a robust *Thermus aquaticus* (*Taq*) polymerase that enabled the specific amplification of the insert of interest. A single colony was inoculated into the PCR reaction mixture, instead of template DNA, and the PCR protocol was followed as shown in Figure 2.2.

Analysis of the PCR reaction was done by agarose gel electrophoresis (refer to section 2.3.4). The strategy involved a forward primer from either the 3' end of GST gene or the 5' forward primer used to amplify the gene fragment of interest. The reverse primer used was the 3' reverse primer for the insert of interest.

2.3.10 Plasmid preparations and DNA sequencing

Plasmids preparations were carried out using the QIAprep® Spin Miniprep Kit following the instruction manual provided. For larger plasmid preps the QIAGEN® Plasmid Maxi Kit was utilized. Following individual plasmid preps, restriction digestion and DNA sequencing were used to confirm the presence of the correct gene fragment of interest. The oligonucleotides used for the sequencing reactions either annealed to the DNA sequence of interest or to the vector DNA flanking the MCS (refer to Table 2.1 and 2.2). Sequencing was done at the National Research Council (NRC) of Canada Plant Biotechnology Institute DNA sequencing facility. Experimental conditions were based upon request of the NRC instructions. Analysis of the sequencing data was with the Translate tool on the ExPASY Proteomics Server (Swiss Institute of Bioinformatics, http://ca.expasy.org/tools/dna.html).

2.3.11 Antibiotic selection

Inclusion of antibiotics in the growth media help to ensure the pGEX-6P-3 vector encoding the gene of interest was maintained and minimize contamination by unwanted bacteria. The pGEX-6P-3 vector encodes an ampicillin resistance gene providing use of ampicillin in cell media as a selective marker (refer to Fig. 2.1). *E. coli* Rosetta 2 cells encode an additional plasmid, which provides resistance for the antibiotic chloramphenicol. The final concentration for ampicillin was 100 μg/mL and for chloramphenicol the final concentration used was 34 μg/mL.

2.4 Bacterial strains and media

2.4.1 Bacterial strains

Table 2.3: Bacterial strains used in experiments.

Bacterial strain	Characteristics	Source
E. coli XL-1	Deficient of recA and endA, hsdR mutation of EcoK endonuclease	Stratagene®
E. coli BL-21(DE3)	T7 polymerase upon IPTG induction, deficient of lon and omp-t proteases	Novagen®
E. coli Rosetta™	Derivative of BL-21 cell line, F^- ompT $hsdS_B$ ($r_B^- m_B^-$) gal dcm, pRARE (argU, argW, ilex, glyT, leuW, proL)	Novagen®

2.4.2 Media type

LB broth was prepared in the laboratory by the following procedure. One litre of LB broth contained 10 g of Bacto™ Tryptone (Becton, Dickinson and Company, BD), 5 g of Bacto™ Yeast Extract (BD), and 10 g NaCl. The solution was sterilised at 121 °C for 30 min using an autoclave. For larger preparations, the length of sterilization was extended to 1 hr. Antibiotic(s) were added once the solution cooled to approximately 50 °C. The other growth medias were made according to Athena Enzyme Systems™ instructions (Table 2.4).

To make 1 L of LB agar, 15 g of Difco™ Granulated Agar was added to 1 L of LB broth. The solution was sterilised at 121 °C for 30 min using an autoclave. The resulting solution was cooled to approximately 50 °C before the addition of desired antibiotic. Then 15 mL was poured into sterile petri dishes. Once LB agar plates were set they could be stored at 4 °C for 21 days.

2.5 Recombinant proteins methods

2.5.1 Protein expression trials

The optimal protein expression conditions for the recombinant GST-fusion proteins were determined for each recombinant proteins studied. To determine optimal expression six experimental variables were altered, including: cell type (Rosetta 2 or BL-21), cell media (Table 2.4), optical density at 600 nm (OD_{600}) of cells before induction (from 0.4 to 1), isopropyl β-D-1-thiogalactopyranoside (IPTG, Ultrapure™ IPTG Invitrogen Life Technologies) concentration (from 0.05 mM – 0.5 mM), temperature (from 16 – 37 °C), and length of induction (4 – 16 hrs).

2.5.2 Expression of GST-fusion proteins and harvesting cells

The first step for small-scale and large-scale preparations of recombinant proteins was transformation of a competent protein expression cell line with the desired recombinant plasmid. A single colony was then inoculated into a 75 mL LB starter culture, containing appropriate antibiotic(s), which was grown at 37 °C for 16 hrs using an Innova 4330 Refrigerated Incubator Shaker, at 250 rpm (New Brunswick Scientific). From the starter culture, 10 mL of solution was inoculated into 2 L of sterile growth media, in a 6 L Erlenmeyer flask used to ensure proper aeration of the culture.

Table 2.4: Summary of growth media used in experiments.

Growth media	Source
Luria Bertani Broth	-----
Hyper Broth™	Athena Enzyme Systems™
Power Broth™	Athena Enzyme Systems™
Superior Broth™	Athena Enzyme Systems™
Turbo Broth™	Athena Enzyme Systems™

This culture was grown at 37 °C and 215 rpm in the shaker to a desired OD_{600} (measured with a Beckman Coulter DU 640 Spectrophotometer). At the correct OD_{600}, protein expression was induced by the addition of IPTG at the appropriate concentration, and the temperature of the shaker was adjusted to the desired expression temperature for the appropriate growth time. In order to harvest the cells and the expressed protein they produced, liquid cultures were centrifuged in a Beckman Coulter J2-HS Centrifuge with a JA-20 rotor for 10 min at 3,870 x g (r_{max} 4,000 rpm) and 4 °C. The subsequent supernatant was decanted and the cell pellets were frozen at -80 °C until required.

2.5.3 Lysis of cells

The first step in the cell lysis procedure was to resuspend the cell pellet in ice cold cleavage buffer (50 mM Tris-HCl [pH 8.0 at 25 °C], 150 mM NaCl, 1 mM EDTA). A 3X volume of lysis buffer was added to the pellet mass and homogenized with an RW16 Basic (IKA®-WERKE) cell homogenizer. For example a 10 g pellet was resuspended into approximately 30 mL of 1X cleavage buffer resulting in a homogeneous solution of approximately 35 mL. The following compounds were added to a final concentration of: 5 mM DTT (Sigma), 1 mg/mL hen egg white lysozyme (10 mg/mL) (Sigma), 0.002 mg/mL Pepstatin A (1 mg/mL) (Sigma), 0.05 mg/mL Leupeptin (5 mg/mL) (Sigma), 0.001 M PMSF (0.2 M) (Sigma), 250 µL/litre of cells DNaseI (10 mg/mL), and 2 µL (100 U) Benzonase (Invitrogen). The homogenized cell solution was then incubated at 4 °C for 25 min. DNase I was prepared in a reaction buffer (10 mM Tris-HCl [pH 7.5 at 25 °C], 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, 20 % (v/v) glycerol), to 10 mg/mL and stored at -20 °C. For small-scale preparations the 1 – 1.5 mL cell solution was lysed by sonication using a Fisher Scientific Sonic Dismembrator Model 500. Sonication was carried out at 70 % amplitude for 5 s pulses and 10 s delays for a total pulse time of 60 s. The sonicator tip used was a micro 1/8 inch. For large-scale preparations the cell solution was lysed by two consecutive passes through a French Press at 750 psi, attempting to maintain a consistent flow. Both a colour change and a decrease in the solution's viscosity were used as indications that lysis was achieved. The resultant lysate was centrifuged at 33,600 x g (r_{max} 10,000 rpm) for 60 min at 4 °C using a JA – 20 rotor and a Beckman Coulter J2-HS Centrifuge. Centrifugation was done to separate the soluble fraction of the lysate from un-lysed cells and cell debris. The supernatant was decanted and saved for

further purification. Centrifugation was repeated if the resultant supernatant was opaque. To remove any colloids within the supernatant, the solution was passed through glass wool packed in a 30 cc syringe. The resultant solution was next purified by Glutathione-Sepharose (GS) affinity purification.

2.6 Protein purification

2.6.1 Glutathione sepharose affinity purification

The first stage of GST-fusion protein purification was Glutathione-Sepharose (GS) affinity chromatography using an ÄKTA FPLC and a HR 16/5 column (Amersham Biosciences, now GE Life Science) packed with 10 – 15 mL of Glutathione Sepharose™ 4B (GE Life Science). The column was equilibrated with 1X Phosphate Buffered Saline (PBS: 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4 [pH 7.3 at 25 °C]). To apply the supernatant of the cell lysate to the column, a 50 mL Superloop™ (GE Life Science) was used. A flow rate of 1 mL/min was used, ensuring a pressure below 0.6 MPa because the GS resin is damaged at higher pressures. To remove unwanted material the column was washed with two times the volume loaded with 1X PBS at 1 – 2 mL/min. Bound GST-fusion protein was eluted with an glutathione elution buffer (50 mM Tris [pH 8.0 at 25 °C], 150 mM NaCl, and 10 mM reduced L-glutathione). Detection of the desired GST-fusion protein was accomplished by UV absorbance at 280 nm by the UV flow cell monitor within the FPLC. Fractions were collected using the FPLC fraction collector with 10 mL glass test tubes. To store the GS column, it was washed with 5 column volumes (CV) of H_2O and 3 CV 20 % ethanol, and then stored at 4 °C. After 3 – 5 purifications a more thorough cleaning was done. This included 3 CV H_2O wash to remove the ethanol, then 1 CV 0.5 – 1 M NaOH, and then 5 CV of H_2O. The column could then be used for protein purification or washed into 20 % ethanol for storage.

2.6.2 Buffer exchange by dialysis

Protein solutions were exchanged into the appropriate buffer through dialysis. Dialysis was achieved by using BioDesign Dialysis Tubing™ with a 3,500 molecular weight cut-off (MWCO). Dialysis tubing was cut to an appropriate length and soaked in the dialysis buffer for 15 min, after which one end of the dialysis tubing was double knotted. The protein solution was poured into the tubing and the remaining end was sealed with a clip. The protein solution

was dialysed in 1 – 4 L of the appropriate buffer, slowly stirred with a magnetic stir bar, for 12 – 16 hrs at 4 °C.

2.6.3 PreScission protease cleavage of GST-fusion proteins and removal of the GST tag

The analyte from the 1st GS affinity column containing the GST-fusion protein solution was dialysed into 4 L of PreScission protease cleavage buffer (50 mM Tris [pH 8 at 25 °C] , 150 mM NaCl, 4 mM DTT, and 1 mM EDTA). Cleavage of the GST-fusion protein was carried out by the addition of 20 μL (80 U) PreScission protease. The solution was incubated for 16 hrs at 4 °C. For optimal cleavage to occur, it was found that the solution should not be stirred or shaken. Post cleavage, insoluble particles were removed by filtration using a 0.45 μM Acrodisc® Syringe filter (Pall Corporation). Purification of the recombinant protein from the GST tag was done by repeating GS affinity chromatography. The recombinant protein, no longer containing the GST tag, will not interact with the Glutathione Sepharose™ 4B resin eluting from the column in the flow-through fractions. The unwanted GST tag bound to the stationary phase was eluted with glutathione elution buffer and then discarded. The column was cleaned and stored as previously described (refer to section 2.6.1). The eluant was dialysed into buffer appropriate for the next stage of purification.

2.6.4 Ammonium sulphate precipitation

Ammonium sulphate precipitation was used to remove contaminating nucleic acids and protein impurities from a partially purified protein solution. An ammonium sulphate precipitation protocol was optimized by determining the minimal percentage of ammonium sulphate that resulted in the precipitation of the protein which was to be further purified. The ammonium sulphate concentration was raised from 0 % to 25 % over a 1 hour period at 4 °C and incubated for 1 hour at 4 °C. The resulting precipitate was collected by centrifugation at 33,600 x g (r_{max} 10,000 rpm) for 30 minutes at 4 °C in the JA-20 rotor. The pelleted precipitate was gradually re-suspended into 100 mM Tris-HCl [pH 8 at 25 °C, 1 M NaCl, and 5 mM DTT buffer by gentle mixing on a Speci-Mix Aliquot Mixer (Barnstead International) for 16 hrs at 4 °C.

2.6.5 Ion exchange chromatography

Anion exchange chromatography employed a Source^{TM} 15 Q ion exchange media (GE Life Science) packed in an HR 16/5 column (GE Life Science) with a bed volume of 7.5 mL. In anion exchange chromatography, negatively charged proteins interact selectively with the positively charged stationary phase while positively charged impurities flow-through the column. Typically, the specific protein was prepared in a buffer at a pH above its isoelectric point. Cation exchange chromatography employed a Source^{TM} 15 S ion exchange media (GE Life Science) packed in an HR 16/5 column (GE Life Science) with a bed volume of 7.5 mL. Cation exchange chromatography utilizes the same principles as anion exchange however positively charged proteins selectively interact with the negatively charged stationary phase, and usually at a pH below the proteins' isoelectric point.

The ionic exchange column was equilibrated by washing with 5 CV of the appropriate buffer (Table 2.5), into which the relevant protein to be purified had been equilibrated. The protein solution was then applied to the column using a 50 mL Superloop^{TM} (GE Life Science) at a 0.5 – 1 mL/min flow rate, ensuring a column pressure below 0.5 MPa. To remove unwanted material bound by non-specific interactions the column was washed with 5 CV of buffer A. Elution of the proteins bound to the stationary phase was with a 0 % - 100 % concentration gradient of 1 M NaCl. The elution buffer used was the same as buffer A, but contained 1 M NaCl. Source Q and Source S columns were used at a flow rate of 1.0 mL/min. At the end of the elution, the column was regenerated by washing the column with 3 CV of buffer A, then 5 CV of H_2O, followed by 3 CV of 20 % ethanol. The cleaned column was stored at 4 °C.

Table 2.5: Ion exchange chromatography strategies employed for *Drosophila* recombinant protein purification. The isoelectric point (pI), type of ion exchange resin, the running buffer, and the pH for the individual recombinant proteins are shown.

Protein	pI	Ion Exchange	Running Buffer	pH
dMSL1$_{820 - 1039}$	5.5	Cation	Tris-HCl	8.0
dMSL3$_{185 - 512}$	5.3	Anion	Tris-HCl	8.0
dMOF$_{370 - 827}$	8.5	Anion	Bis-tris propane	7.0

2.6.6 Size exclusion chromatography

Size exclusion chromatography utilized either a Superdex^{TM} 75 10/300 (GE Life Science) or a Superdex^{TM} 200 10/300 (GE Life Science) analytical gel filtration column (bed volume 24 mL). The two columns differ in the average size of the pores within the matrix. This enabled optimal purification of a desired protein based on its respective size. For proteins or complexes with an approximate size between 70 kDa and 2 kDa the Superdex^{TM} 75 10/300 column was used, and for proteins or complexes between 400 kDa to 7 kDa the Superdex^{TM} 200 10/300 column was used. Size exclusion columns were pre-equilibrated in 50 mM Tris-HCl [pH 8.0 at 25 °C], 150 mM NaCl, and 5 mM DTT. The protein sample was applied to the column using either a 500 µL or a 2 mL sample loop (GE Life Science). The column was run at a flow rate of 0.5 mL/min ensuring the column pressure did not exceed 1 MPa. The column was cleaned by 3 CV of H_2O and 3 CV of 20 % ethanol, and stored at room temperature.

2.6.7 Concentration of proteins

Protein solutions were concentrated by use of either an Ultracel regenerated cellulose Amicon® Ultra-15 centrifugal filter device (3,500 MWCO) (Millipore), for large volumes (> 15 mL), and an Ultracel regenerated cellulose Amicon® Ultra 3.5k Dalton cut-off centrifugal filter device (3,500 MWCO) (Millipore), for small volumes (< 15 mL). Centrifugal filter devices were pre-washed with H_2O or protein buffer. Centrifugation was through the use of a Sorvall® Biofuge primoR (Mandel) at 2,150 x g (r_{max} 4,000 rpm) and 4 °C. Centrifugation was carried out for increments of times depending on the desired final volume.

2.7 Protein visualization techniques

2.7.1 Glycine sodium dodecyl sulphate – polyacrylamide gel electrophoresis

Glycine sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) was utilized for visualization and subsequent analysis of purified recombinant proteins through-out over-expression and purification. SDS-PAGE was performed using the Laemmli method (Laemmli, 1970). All SDS-PAGE gels were cast with a 4 % acrylamide stacking gel (25 µL 10 % (w/v) SDS, 0.62 mL 0.5 M Tris-HCl [pH 6.8 at 25 °C], 0.25 mL 40 % acrylamide, 50 µL 10 % APS, 10 µL TEMED made up to 2.5 mL with H_2O) and a 12.5 % acrylamide resolving gel (50 µL 10 % (v/v) SDS, 1.25 mL 1.5 M Tris-HCl [pH 8.8 at 25 °C],

1.4 mL 40 % acrylamide, 50 µL 10 % APS, 10 µL TEMED made up to 4.5 mL with H_2O) using a Hoefer SE245 Mighty Small Dual Gel Caster (GE Life Science), 0.75 mm spacers and 10 well combs. A 15 % acrylamide resolving gel was also used, which was made by adjusting the amount of 40 % acrylamide and H_2O. Samples were mixed with sample loading buffer (62.5 mM Tris-HCl [pH 6.8 at 25 °C], 10 % glycerol, 2 % SDS, 5 % (w/v) β mercaptoethanol, and 1 % (w/v) Bromophenol blue), then boiled for 2 min, and then loaded on the gel. The capacity of an individual gel well was approximately 20 µL, but usually 5 – 15 µL samples were loaded. SDS-PAGE was typically done at a 115 V for approximately 1.5 hrs or until the dye front had run off the bottom of the gel. The running buffer utilized consisted of 1X SDS buffer (25 mM Tris-HCl [pH 8.2 at 25 °C], 200 mM glycine, 0.1 % (w/v) SDS) diluted from a 10X stock solution. Gels were stained with Coomassie staining solution (40 % methanol, 10 % v/v glacial acetic acid, 0.1 % w/v Coomassie Brilliant Blue) for 30 mins and destained for 2 to 4 hrs in 7.5 % v/v glacial acetic acid, and 10 % v/v methanol. For staining and destaining, gels were gently agitated on an IKA® VIBRAX VXR Basic Orbital Shaker. Photographs of SDS-PAGE gels were taken with a Gel Logic 200 Scientific Imaging Systems (Eastman Kodak Company) and analysed with Kodak Molecular Imaging Software.

2.7.2 Tricine SDS-PAGE

Tricine SDS-PAGE was also used for analysing the molecular weight of recombinant proteins. Tricine SDS-PAGE was carried out using the method proposed by Schägger and von Jagow (Schagger and von Jagow, 1987). Tricine SDS-PAGE gels were cast with a 3.8 % acrylamide stacking gel (for 6.35 mL: 0.6 mL acrylamide solution [40 % (w/v) acrylamide and 1.25 % (v/v) N,N'-methylene-bis-acrylamide], 1.55 mL gel buffer [3 M Tris-HCl [pH 8.45 at 25 °C], 0.3 % (w/v) SDS], 5 µL TEMED, 50 µL 10 % APS, and 4.2 mL H_2O) and a 10 % acrylamide resolving gel (for 10.9 mL: 2.7 mL acrylamide solution, 3.3 mL gel buffer, 2.1 mL glycerol, 3.3 µL TEMED, 33 µL 10 % APS, and 2.8 mL H_2O) using a Hoefer SE245 Mighty Small Dual Gel Caster (GE Life Science), 0.75 mm spacers and 10 well combs. Samples were mixed with 2x sample buffer (4 % (w/v) SDS, 20 % (w/v) glycerol, 100 mM Tris-HCl [pH 6.8 at 25 °C], 4 % (v/v) β-mercaptoenthanol). The 1X anode buffer (0.2 M Tris-HCl, pH 8.9) and the 1X cathode buffer (0.1 M Tris-HCl [pH 8.25], 0.1 M Tricine, 0.1 % (w/v) SDS) were diluted from a 10X stock solution. Gels were stained with Coomassie staining solution for 30

mins and destained for 2 to 4 hrs in 10 % v/v glacial acetic acid, and 10 % v/v methanol. For staining and destaining gels were gently agitated on an IKA® VIBRAX VXR Basic Orbital Shaker. Photographs of Tricine SDS-PAGE gels were taken with a Gel Logic 200 Scientific Imaging Systems (Eastman Kodak Company) and analysed with Kodak Molecular Imaging Software.

2.8 Protein analysis

2.8.1 Protein concentration determination

A Far-UV absorbance spectrum from, 300 nm to 200 nm, was utilized to quantify the amount of protein and nucleic acid in a protein solution. Both a Beckman Coulter DU 640 Spectrophotometer and a Thermo Scientific Nanodrop Spectrophotometer were used to conduct a Far-UV absorbance spectrum. A quartz cuvette (Bio-Rad quartz spectrophotometer cell semi micro 9-Q-10 mm) was used. A quick and easy method for estimating protein concentration was carried out using the absorbance calculated at 278 nm based on the percentage of tyrosine, tryptophan, and cysteine in the protein sequence, and the corresponding extinction coefficient at 278 nm (http://www.expasy.ch/tools/protparam.html). The theoretical extinction coefficient was determined by the equation: Extinction coefficient (ε) = (# of Tyr)(ε Tyr) + (# of Trp)(ε Trp) + (# of Cys)(ε Cys), where ε Tyr equals 1490 M^{-1} cm^{-1}, ε Trp equals 5500 M^{-1} cm^{-1}, and ε Cys equals 125 M^{-1} cm^{-1} at 278 nm measured in water (Gill and von Hippel, 1989; Pace, 1990).

Nucleic acids absorb strongly in the near UV with a maximum at 260 nm whereas the aromatic amino acids tyrosine and tryptophan within proteins absorb maximally at 278 nm. For a more accurate quantification of the protein concentration it must be remembered that nucleic acids posses a strong shoulder extinction in the 280 nm region. The protein concentration in mg/mL was crudely calculated by the equation: Concentration (mg/mL) = $((1.55 \times Abs_{280}) - 0.76 \times Abs_{260})$, which takes into account the influence of the absorbance at 280 nm, but does not account for the influence of Tyr, Phe, and Trp at 260 nm (Waddell, 1956).

To calculate a more accurate protein concentration of a homogenous protein solution, the Bradford protein assay was carried out. A standard curve of bovine serum albumin (BSA) protein, with concentrations between 0 and 10 mg/mL, was constructed. Bradford protein

assays were carried out using Bradford reagent (Bio-rad), which has a maximum absorbance at 595 nm. From the graphed BSA standard curve the protein concentration of a desired protein solution could then be extrapolated.

2.8.2 Quaternary structure and molecular weight determination

The Superdex 75 and the Superdex 200 size exclusion columns were calibrated against a Low Molecular Weight and a High Molecular Weight Gel Calibration kit (GE Life Science) respectively (Lane *et al*., 2006). The low molecular weight kit contained Albumin (67 kDa), Ovalbumin (43 kDa), Chymotrypsinogen A (25 kDa), and Ribonuclease A (13.7 kDa). The high molecular weight kit consisted of Blue Dextran (void volume), Thyroglobulin (669 kDa), Ferritin (440 kDa), Catalase (232 kDa), and Aldolase (158 kDa). The calibration curves were plotted with K_{av} versus the log molecular weight, which permitted extrapolation of the molecular weight of a recombinant protein. $K_{av} = [(V_e - V_o) / (V_c - V_o)]$, where V_o equals void volume, V_e equals elution volume, and V_c equals geometric column volume. Therefore, the molecular weight and the quaternary structure could be estimated with the calibrated size exclusion columns. Extrapolated molecular weights are assumed to be reasonable estimates of the actual molecular size if the proteins being evaluated have an approximately spherical shape.

2.9 Protein crystallization

2.9.1 Hanging drop vapour diffusion

Following size exclusion chromatography, recombinant proteins were buffered in 25 mM Tris-HCl [pH 8.0 at 25 °C], 150 mM NaCl, and 5 mM DTT, which was exchanged for the crystallization buffer (25 mM Tris-HCl [pH 8.0 at 25 °C], 150 mM NaCl, 1 mM TCEP, and 10 % glycerol). Protein crystallization experiments were performed using the hanging drop vapour diffusion technique (Chayen, 1998). The protein concentrations used were between 8 – 10 mg/mL, determined with a Thermo Scientific Nanodrop Spectrophotometer (refer to section 2.8.1). Individual hanging drops were equilibrated over a 500 μL reservoir solution containing 0.5 μL protein solution mixed with 0.5 μL of the unique reservoir solution. Crystal formation is favoured by the droplet containing purified protein, buffer, and precipitants all equilibrating with a larger reservoir containing similar buffers and precipitants in higher concentration. Hence crystallization was carried out in a closed system at a stable temperature and pH.

Variables in crystallizations include initial protein concentration, the effective pH of the reservoir/protein droplet, type of precipitant, additives such as salts, and the type of buffer.

2.10 Protein-protein interaction studies

2.10.1 GST pull-downs

Protein-protein interactions were tested using GST pull-downs assays (Brymora *et al*., 2004). Experiments consisted of a GST-fusion protein (the bait) mixed with an excess of the recombinant protein (the prey). The prey protein is incubated with the bait protein bound to Glutathione Sepharose™ 4B resin, and then extensively washed with buffer. If some prey protein remained associated with the bait after washing, it is presumed that the prey and bait proteins interact. All GST pull-down experiments were done using 60 - 100 μg of protein for the prey, the bait, and the GST control. The assay had four controls: Glutathione Sepharose™ 4B resin (GS beads) only, purified GST protein plus GS beads, prey protein plus GS beads, and bait protein plus GS beads. All experiments were made with a consistent final volume of 500 μL. The pull-down experiment consisted of initial binding of the bait to the GS beads, done by adding 60 μg of bait to approximately 15 μL of GS beads (70 % solution), which was incubated for 1 hour at 4 °C and gently mixed on a Speci-Mix Aliguot Mixer (Barnstead International). GS beads were pelleted by centrifugation using a Microfuge® 18 Centrifuge (Beckman Coulter™) at 2,000 rpm for 1 min. The resulting supernatant was aspirated off the pellet. Non-specific interactions and impurities were removed by washing the GS-beads 4X with 1X PBS pH 7.3. Next the prey, 60 μg – 100 μg, was incubated with the bait bound to the GS beads for 2 hour at 4 °C on a Speci-Mix Aliguot Mixer (Barnstead International). Non-specific interactions and impurities were removed by washing the GS beads 4 times with 1X PBS pH 7.3. The supernatant was removed and the remaining pellet was mixed with 15 μL of sample loading buffer (62.5 mM Tris-HCl [pH 6.8 at 25 °C], 10 % glycerol, 2 % SDS, 5 % (w/v) β mercaptoethanol, and 1 % (w/v) Bromophenol blue) boiled for 2 min and then analysed by SDS-PAGE (refer to section 2.7.1).

2.10.2 Analysis by size exclusion

Size exclusion chromatography enabled protein interaction studies based on the formation of a stable stoichiometric complex with a characteristic molecular weight. Protein

interaction studies could be assessed with size exclusion chromatography because of the inert nature of the stationary phase of the column. Proteins to be tested were concentrated (refer to section 2.6.7) and then mixed together. The resulting protein solution if needed was concentrated and then analysed by size exclusion chromatography (refer to section 2.6.6). Protein interactions or the lack there of could be inferred from the elution profile. Individual elution peaks were visualized by SDS-PAGE (refer to section 2.7.1) to ascertain the protein make-up.

3. Results

3.1 Cloning the C-terminal domain of *Drosophila melanogaster* MSL1

In a study aimed at defining the protein domains of *Drosophila* MSL1, it was found that a MSL1 C-terminal construct, consisting of amino acid residues 705 – 1039, directly interacted with both MOF and MSL3 (Scott *et al*., 2000). Moreover, residues 766 – 939 of MSL1 were determined to interact with MOF and residues 973 – 1039 of MSL1 the region for interaction with MSL3 (Morales *et al*., 2005; Morales *et al*., 2004). Utilization of Protein-Protein BLAST (BLAST-P) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) enabled identification of conserved amino acids within the C-terminus of both *Drosophila* species and other higher eukaryotes. BLAST-P was used to search the non-redundant protein sequences databank held at NCBI. Amino acid sequence conservation between homologues (e.g. MSL1 sequences from different species) is a useful tool in defining domain boundaries in proteins. Therefore, the amino acid residues encompassing a conserved segment of the C-terminus of MSL1 were deemed to be part of a protein domain or a linker region enabling gene constructs to be designed to encompass a complete C-terminal domain. Statistical E-values from BLAST-P searches less than 0.005, were considered a significant match. This essentially means that aligned sequences yielding scores less than 0.005 are very unlikely to arise simply by chance. The E values for *Danio rerio*, *Mus musculus*, and *Homo sapiens* MSL1 were determined to be using the BLAST-P algorithm $3x10^{-9}$, $6x10^{-9}$, and $5x10^{-9}$ respectively. Position-Specific Iterated BLAST (PSI-BLAST) did not improve the effectiveness of the search compared to BLAST-P as no new homologous sequences were discovered. The protein sequence of *Drosophila melanogaster* MSL1 was aligned against the best hits from the search based on E-value criteria and obvious conservation of MSL1 domain structure (Fig. 3.1). BLAST-P and PSI-BLAST consistently identified candidate MSL1 proteins from numerous *Drosophila* species including *simulans*, *sechelia*, *erecta*, and *yakuba* all of which scored a 98 % or above sequence identity, and *Drosophila* species *ananassae*, *pseudoobscua*, *willstoni*, *mojavensis*, and *virulus* all were a sequence identity of around 50 %. Evident from the sequence alignment between MSL1 and its higher eukaryotic homologues is the conservation of the PEHE domain, suggesting its necessity for proper interaction with both MSL3 and/or MOF (Fig 3.2) (Morales *et al*., 2004; Scott *et al*., 2000).

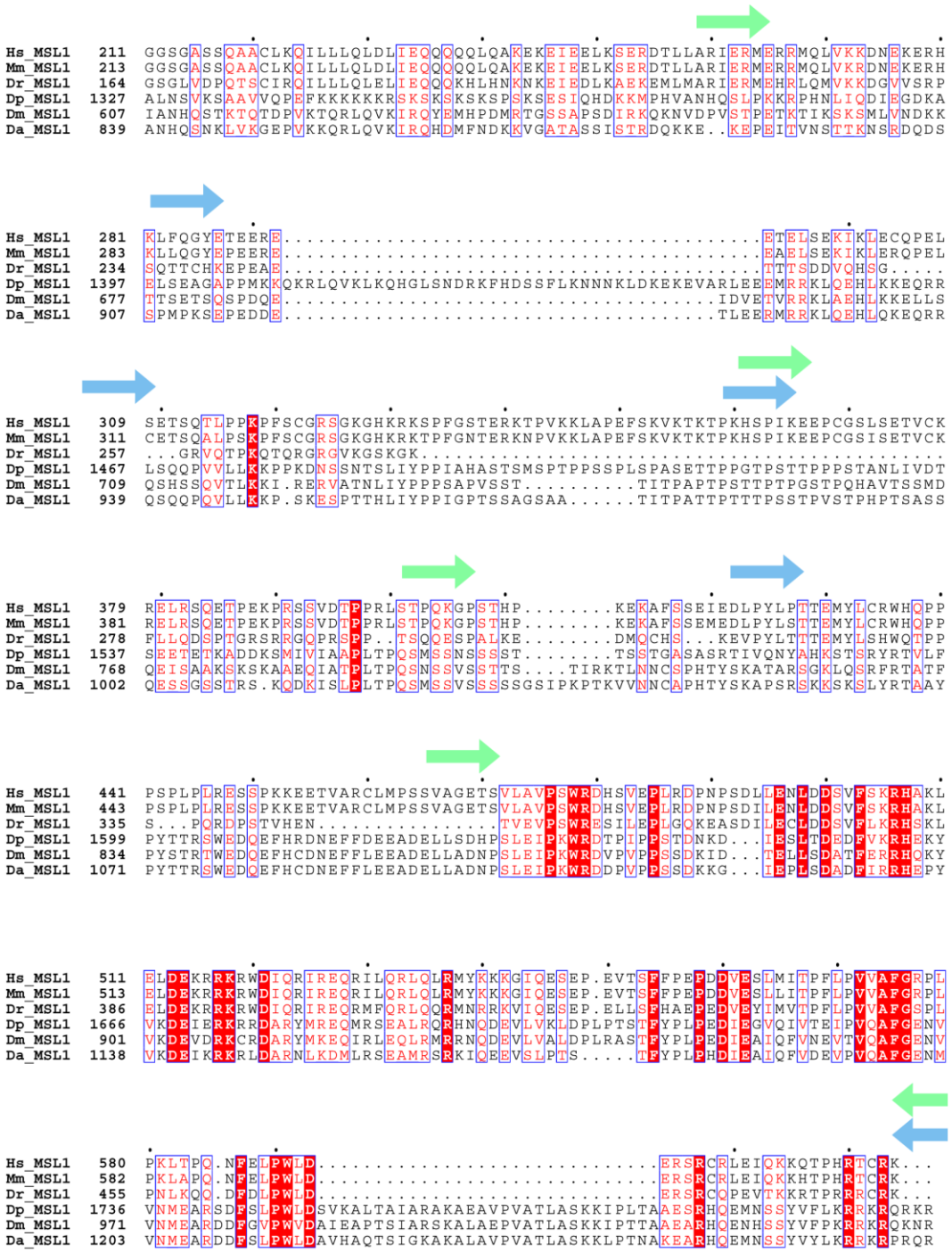


Figure 3.1: Multiple sequence alignment and cloning strategy for MSL1 proteins. Hs: *Homo sapiens*, Mm: *Mus musculus*, Dr: *Danio rerio*, Dp: *Drosophila psuedoobscua*, Dm: *Drosophila melanogaster*, Da: *Drosophila ananassae*. The blue arrows indicate the primers corresponding to the cloned dMSL1 constructs (refer to Table 2.1). The green arrows indicate the primers corresponding to the cloned hMSL1 constructs (refer to Table 2.2). ClustalW2 and ESPript 2.2 were used for the multiple sequence alignment (Gouet *et al*., 1999; Larkin *et al*., 2007).

Four unique constructs encompassing the identified PEHE domain were designed from the *Drosophila melanogaster* MSL1 cDNA sequence. The first construct comprised 657 base pairs or 219 amino acids in length, termed $dMSL1_{820\text{-}1039}$ (Fig. 3.1 & 3.2). The second construct encoded 855 base pairs or 285 amino acids in length, termed $dMSL1_{754\text{-}1039}$. The third construct was 981 base pairs or 327 in length, termed $dMSL1_{712\text{-}1039}$. Lastly the fourth construct was 1071 base pairs or 357 amino acids, termed $dMSL1_{688\text{-}1039}$. Increasing the overall size of the construct may improve stability by means of affecting the subsequent secondary structure of the protein. The constructs designed are listed below; with the intent of future protein interaction studies all contained the PEHE domain (865 – 1004), and regions previously determined to be required for interaction with MOF (766 – 939) and MSL3 (973 – 1039) (Morales *et al*., 2004; Scott *et al*., 2000).

All of the amplified *Drosophila* MSL1 PCR fragments were cleaved with *Bgl*II and *Eco*RI generating compatible ends for the pGEX-6P-3 vector also cleaved with *Bam*HI and *Eco*RI. *Bgl*II was used to cleave the amplified gene fragments because the MSL1 PEHE domain contains an internal *Bam*HI restriction site (amino acid residues 754 – 755). Cutting the vector and the gene of interest with *Bam*HI and *Bgl*II respectively destroyed the *Bam*HI restriction site within the MCS of the vector; therefore, analysis of the resulting chimeras were carried out using colony PCR and DNA sequencing. *Bgl*II was utilized because it generates compatible ends with *Bam*HI. Proper insertion of the fragment of interest requires verification of correct orientation, absence of unwanted mutations, and ablity to express the recombinant protein in-frame. Colony PCR was performed for all four MSL1 constructs and analyzed by agarose gel electrophoresis. The desired gene fragment of interest was verified for all four MSL1 constructs because the respective PCR product was the same size as the known amplified insert (Fig. 3.3).

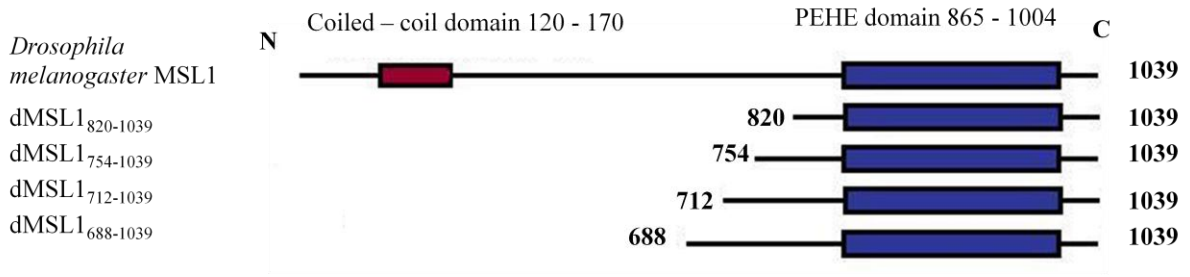


Figure 3.2: Schematic drawing of *Drosophila melanogaster* MSL1 recombinant proteins investigated. The numbers at the beginning and end of the constructs identify the amino acids of MSL1 and indicate the fragment of MSL1 expressed. The C-terminal dMSL1 constructs cloned are shown below full length MSL1.

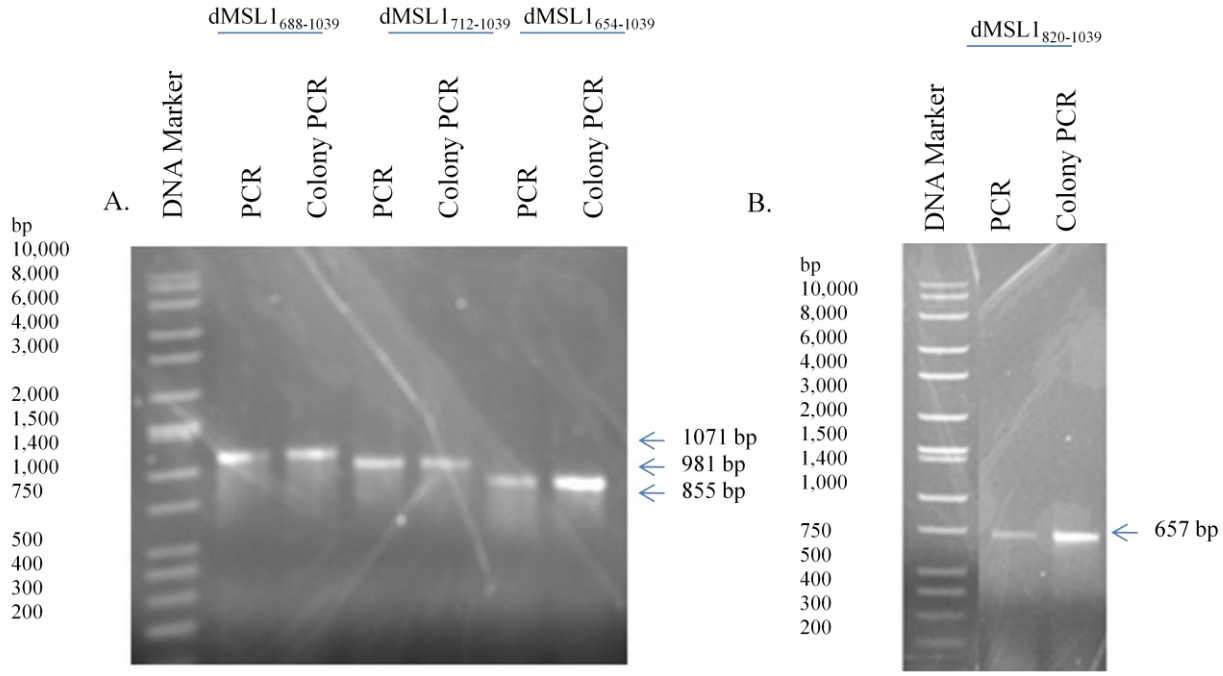


Figure 3.3: Colony PCR of candidate transformants for the four *Drosophila* MSL1 constructs cloned. (A). Agarose gel of the cloned *Drosophila* MSL1 gene fragments and verification of correct ligation and transformation into XL-1 *E. coli* cells with colony PCR. DNA marker (Wide Range), PCR amplified dMSL1$_{688-1039}$, Colony PCR product dMSL1$_{688-1039}$, PCR amplified dMSL1$_{712-1039}$, Colony PCR product dMSL1$_{712-1039}$, PCR amplified dMSL1$_{754-1039}$, Colony PCR product dMSL1$_{754-1039}$. (B). Agarose gel of the cloned *Drosophila* MSL1 gene fragments and verification of correct ligation and transformation into XL-1 *E. coli* cells with colony PCR. DNA marker (Wide Range), PCR amplified dMSL1$_{820-1039}$, Colony PCR amplified dMSL1$_{820-1039}$.

3.2 Optimization of protein expression and purification of dMSL1$_{820\text{-}1039}$

Expression and purification of the four unique MSL1 C-terminal domain constructs were attempted in bacterial cells. Utilization of *E. coli* provides a method for fast, inexpensive, and easy over-expression of recombinant proteins. However, various issues in over-expressing recombinant proteins in bacteria include: can the protein be reconstituted in bacteria, is it expressed as a soluble, stable, and correctly folded protein, can it be purified to homogeneity, and will how will the lack of post-translational modification affect the recombinant protein. For expression purposes, *E. coli* BL-21 DE3 (BL-21) cells were freshly transformed with the desired plasmid, and plated on LB Agar Amp^{100}. A single colony was inoculated into 75 mL of LB Amp^{100} and grown overnight in a shaking incubator at 210 rpm and 37 °C. For small-scale preps a pre-warmed 75 mL of LB Amp^{100} was inoculated with 1 mL of overnight culture; for large-scale preparations pre-warmed 4 L of LB Amp^{100} was inoculated with 20 mL of overnight culture. Media containing the appropriate bacterial culture were grown to the desired cell density measure by OD at 600nm (OD_{600}) in a shaking incubator at 210 rpm and 37 °C, once the OD_{600} was achieved protein expression was induced by addition of IPTG to a desired final concentration. The incubator was changed to the desired temperature and the culture was incubated for an additional 16 hrs at 210 rpm.

The initial screening of optimal protein expression conditions were carried out through the use of small-scale batch preps; harvested cells were lysed by sonication, and then the lysed cells were separated into soluble and insoluble fractions by centrifugation. Insoluble proteins aggregate forming inclusion bodies, typically mis-folded cellular proteins are found in inclusion bodies. In order to compare protein expression and solubility required a consistent sonication protocol. To aid in sonication the solution was incubated with lysozyme; moreover, experience with sonication was required to properly assess complete lysis of the cells. The resulting supernatant was subjected to small-scale GST pull-downs. The soluble fraction containing the soluble GST-fusion protein was incubated with approximately 35 µL of GS beads (70 %) slurry, which was washed 5 times with 1X PBS, for 2 hrs on a rocker at 4 °C. The GS beads were gently pelleted by centrifugation and then washed 5 times with 1X PBS to remove non-specific interactions. The GST pull-downs were analysed by SDS-PAGE.

Numerous protein expression conditions were tried with variation in the type of broth, optical density (OD_{600}), incubation temperatures, and IPTG concentrations to determine

optimal over-expression. To determine the optimal conditions for recombinant protein expression both BL-21 and Rossetta 2 cells were tried. In both types of cells three OD_{600} 0.4, 0.6 and 0.8 (or 1.0) were carried out. The different induction temperatures tested were 16 °C, 24 °C, and 37 °C and were attempted under different lengths of induction 15 hrs or 4 hrs. To then improve upon protein over-expression expression trails carried out with different designed protein expression broths (i.e Hyper or Turbo broth) and various IPTG concentrations (0.05, 0.1, 0.5, and 1 mM). Expression trials of the three recombinant proteins $dMSL1_{688-1039,}$ $dMSL1_{712-1039,}$ and $dMSL1_{754-1039}$ failed, as these three recombinant proteins were not expressed as soluble protein and found to be in the cell pellet.

An example of two small-scale GST-$dMSL1_{820-1039}$ protein expression trials illustrates the difference of soluble protein over-expressed solely due to change of the OD_{600} (Fig. 3.4). Growth of cells containing GST-$dMSL1_{820-1039}$ resulted in approximately five times more soluble recombinant protein when the OD_{600} of induction was chosen to be 0.6 AU versus 0.4 AU.

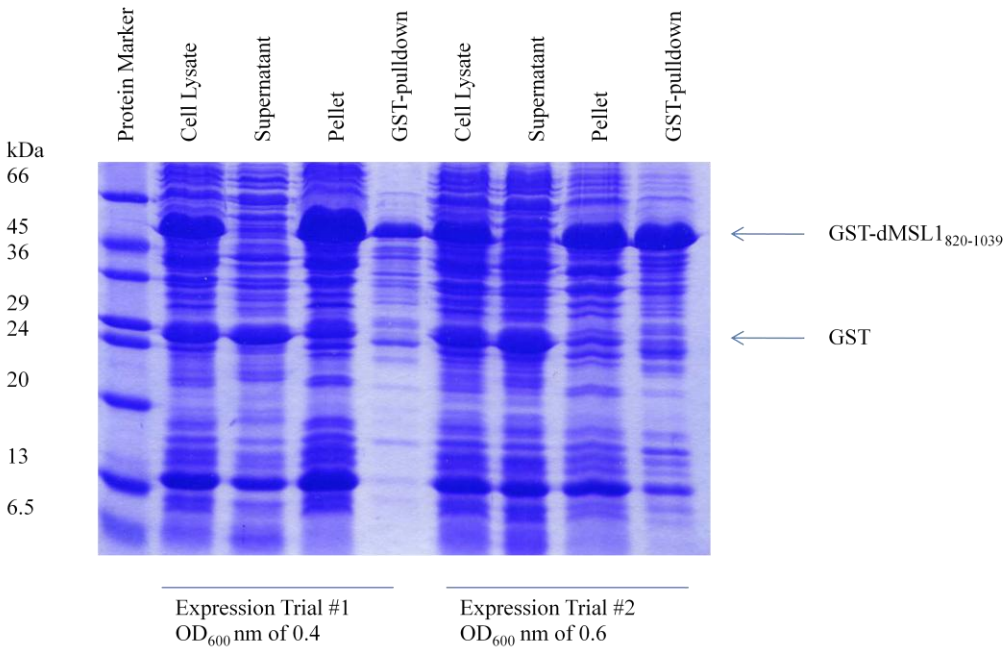


Figure 3.4: Analysis of two unique small-scale protein expression trials of $dMSL1_{820-1039}$. SDS-PAGE: protein marker, cell lysate, supernatant (soluble fraction/load GS column), pellet (insoluble fraction), GST pull-down, cell lysate, supernatant (soluble fraction/load GS column), pellet (insoluble fraction), GST pull-down.

Optimal over-expression of recombinant GST-dMSL1$_{820\text{-}1039}$ was empirically determined to be at the following conditions: LB broth, OD_{600} of 0.6 AU, 0.1 mM IPTG, at 16 °C for 14 hrs (Fig. 3.4). A large portion of over-expressed dMSL1 GST-fusion protein was found to be insoluble under all conditions tested. This is most likely due to the difficulty in expressing a eukaryotic protein in bacteria. Moreover, over-expression of a heterologous protein likely overloads the *E. coli* protein synthesis machinery, resulting in mis-folded proteins which aggregate into inclusion bodies. However, a significant level of the over-expressed protein was also found to be soluble enabling further large-scale preparations and purification of the dMSL1 C-terminal domain.

Large-scale purifications were carried out using 4 – 6 litres of LB media under the optimal conditions determined for over-expression of GST-dMSL1$_{820\text{-}1039}$. Large-scale preparations consistently resulted in a cell pellet weight of 6 g per L of cell culture. The cells were lysed using a French press, and centrifuged to separate the soluble fraction from the insoluble fraction. The lysate supernatant containing the soluble GST-dMSL1$_{820\text{-}1039}$ protein was purified by FPLC GS affinity chromatography. The molecular weights of recombinant GST-dMSL1$_{820\text{-}1039}$ and dMSL1$_{820\text{-}1039}$ were calculated with ExPASy ProPram tool to be 49.5 kDa and 25.5 kDa respectively. The correct GST-fusion protein was over-expressed and verified by SDS-PAGE, as the protein band migrated at its predicted molecular weight of approximately 50 kDa (Fig. 3.4). The chromatogram of the first GS affinity column verified the desired interaction of the GST tagged protein with the stationary phase, which enabled purification from the unwanted whole cell extract (Fig. 3.5).

To remove the GST tag, the purified protein solution of the first GS affinity column was dialysed into cleavage buffer: 50 mM Tris-HCl, 150 mM NaCl, and 1 mM EDTA (pH 7.0) for 16 hrs. Proteolytic cleavage of the dialysed GST-dMSL1$_{820\text{-}1039}$ protein was performed by the addition of 80 units of PreScission protease. Prior to proteolytic cleavage the fusion protein was diluted by approximately 30 percent to prevent precipitation of the resultant protein once the GST tag was removed, as it has been found in the laboratory that the GST-fusion proteins are generally more soluble than the proteolytically liberated fragments. The GST tag was purified away by a second passage over the GS affinity column (Fig 3.6 and 3.7). Cleavage and the subsequent removal of the GST tag were conducted carefully because the stability of the recombinant protein of interest can be altered. Therefore, protease inhibitors (Leupeptin,

Pepstatin A, and PMSF) were included during the PreScission protease cleavage, and the experiment was carried out at 4 °C. To aid in complete removal of the GST tag, the cleavage reaction was empirically found to work best without any shaking, rocking or stirring of the solution, likely because oxidation may reduce the effectiveness of cleavage reaction (PreScission protease contains an active site sulfhydryl group). Lastly, in order to enhance separation of the GST tag and the recombinant protein, the protein solution was dialysed into 1X PBS before the second GS affinity column. Dialysis permitted removal of the free L-glutathione that was interacting with the GST tag enabling the GST tag to once again strongly interact with the stationary phase of the GS affinity column.

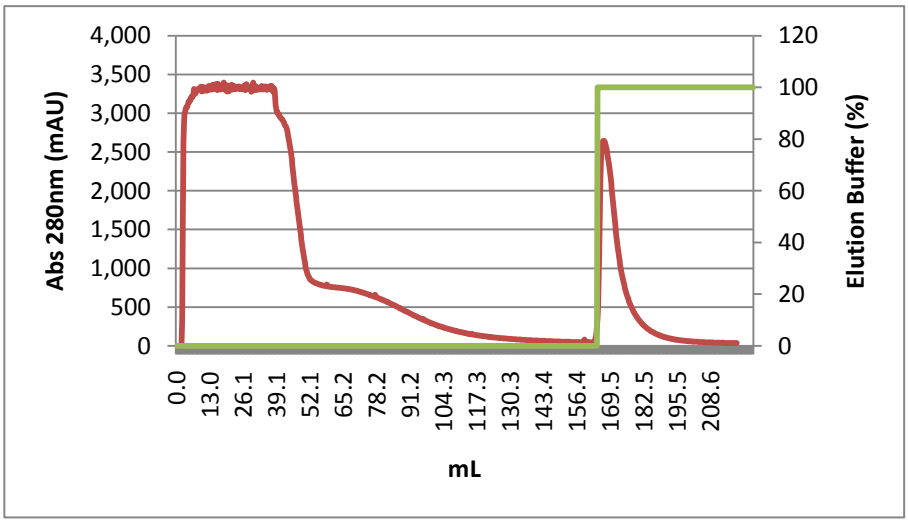


Figure 3.5: Purification of GST-dMSL1$_{820\text{-}1039}$ by GS affinity chromatography. The chromatogram of the 1st GS affinity column illustrates both the flow-through of the crude lysate and the elution of GST-dMSL1$_{820\text{-}1039}$. Four litres of cell culture had a wet weight of 24 grams, equalling approximately 50 mL of soluble fraction (cell lysate) applied to the column. The single fraction eluted by elution buffer (50 mM Tris [pH 8.0], 150 mM NaCl, and 10 mM reduced L-glutathione) was 25 mL. Absorbance (280 nm) is displayed in maroon and percentage of elution buffer in green.

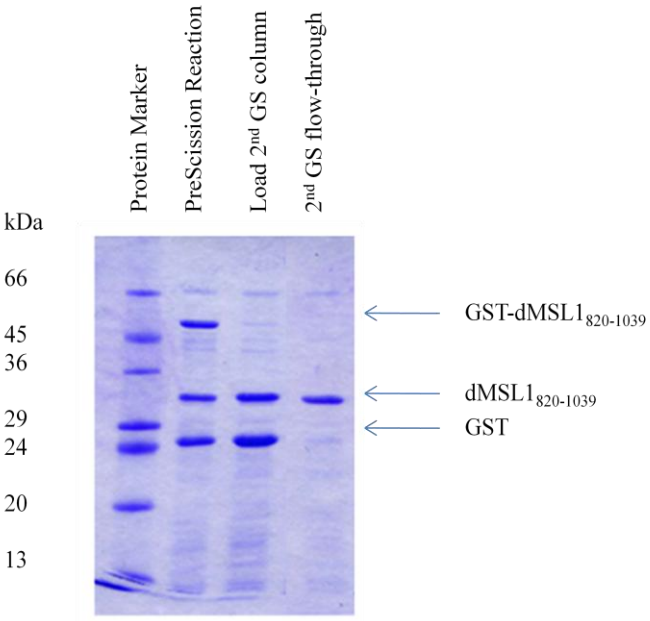


Figure 3.6: GST-dMSL1$_{820-1039}$ PreScission protease experiment and 2nd GS affinity chromatography. SDS-PAGE: protein marker, PreScission protease reaction (purified protein solution eluted from 1st GS column dialyzed into 1X cleavage buffer plus PreScission protease, load 2nd GS column (after centrifugation of PreScission protease digestion), 2nd GS flow-through.

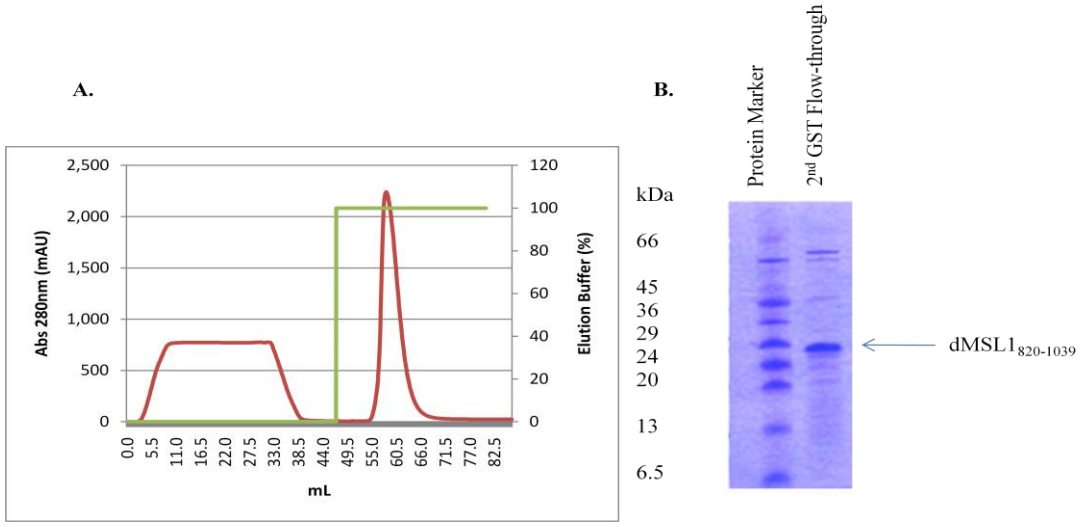


Figure 3.7: dMSL1$_{820-1039}$ purification with a 2nd GS affinity column. (A). Chromatogram of the 2nd GS affinity column of dMSL1$_{820-1039}$. The dialyzed protein dMSL1 protein solution equalling 35 mL was applied to the column. The wanted flow-through fraction was 35 mL and the unwanted fraction eluted by elution buffer 50 mM Tris [pH 8.0], 150 mM NaCl, and 10 mM reduced L-glutathione) was 10 mL. Absorbance (280 nm) is displayed in maroon and percentage of elution buffer in green. (B). Tricine-SDS-PAGE of 2nd GS affinity column. Protein markers, 2nd GS flow-through (2nd GS affinity column).

The initial two GS affinity columns illustrated excellent and robust purification of $dMSL1_{820-1039}$ (Fig 3.4 & 3.6). Analysis of the chromatogram of the second GS affinity column showed that the GST-fusion tag was correctly cleaved from the recombinant protein and two products were successfully purified from one another (Fig. 3.7). However, the $dMSL1_{820-1039}$ that eluted from the second GS affinity column visualized by SDS-PAGE was not at the expected size; the corresponding band indicated its apparent molecular weight to be approximately 34 kDa. Tricine-SDS-PAGE is often used to improve separation of proteins in the low mass range, typically less than 100 kDa (Schagger, 2006). Eluted protein from the second GS affinity column was analysed by a tricine gel, which revealed the expected size of the predict $dMSL1_{820-1039}$ band at 26 kDa (Fig. 3.7).

The eluted protein from the second GS affinity column was analyzed by Far-UV absorbance spectroscopy, ranging from 300 nm to 200 nm, which revealed the presence of significant nucleic acid impurities based on the ratio of Abs at 280 nm to Abs 260 nm (Table 3.1). Therefore, nucleic acids had co-purified with recombinant MSL1 protein. The addition of DNase I and Benzonase® Nuclease was found to be insufficient to remove the nucleic acid bound to $dMSL1_{820-1039}$. DNaseI is an endonuclease, which cleaves non-specifically dsDNA and ssDNA, while Benzonase® Nuclease is a non-specific endonuclease, which degrades all forms of DNA and RNA. Co-purification of nucleic acid is likely because it forms a complex with $dMSL1_{820-1039,}$ which effectively protects the nucleic acid from digestion. The high level of absorbance at 260 nm required the removal of the nucleic acid co-purifying with $dMSL1_{820-1039}$ for further purification of the recombinant protein. Hydroxyapatite chromatography was first attempted as it is often used in the purification of both nucleic acids and DNA binding proteins (Broadhurst, 2001; Schroder *et al*., 2003). The column used was a Bio-Scale^{TM} Mini ceramic hydroxyapatite type I (BIO-RAD). The end result of the hydroxyapatite chromatography was a protein solution which remained unchanged relative to to the sample loaded. Nucleic acid impurities may be removed by their differential affinity for the stationary phase as compared to the protein; however, this did not occur. The ratio of Abs_{280} nm compared to Abs_{260} nm was not increased, indicating that the interaction between the nucleic acids and MSL1 was not disrupted by the column matrix (data not shown).

Ammonium sulphate precipitation was used in an attempt to remove the unwanted nucleic acids. Initial findings showed that increasing the ammonium sulphate from 0 % to 40

% resulted in the complete precipitation of dMSL1$_{820-1039}$ (Fig. 3.8). Furthermore, increasing the percent of ammonium sulphate from 40 % to 70 % and then 70 % to 100 % resulted in precipitation of various impurities, one being GST. The Abs$_{280}$ to Abs$_{260}$ ratio was found to increase by the 40 % ammonium sulphate precipitation, indicating the removal of unwanted nucleic acids (Table 3.1). Therefore, ammonium sulphate precipitation effectively worked as a purification step and acted as a concentration step (Fig. 3.8). Likely due to the increase in concentration ammonium sulphate precipitation also revealed other significant protein impurities with dMSL1$_{820-1039}$. Both optimal purification of dMSL1$_{820-1039}$ and the capability for resuspension of the resulting precipitant was determined to be at 25 % ammonium sulphate. The resulting precipitant was readily resuspended into 100 mM Tris-HCl (pH 8.0), 1 M NaCl, and 5 mM DTT. Furthermore, comparison of precipitation by 25 % to 40 % ammonium sulphate revealed an improvement in removal of the nucleic acid contaminant (Table 3.1).

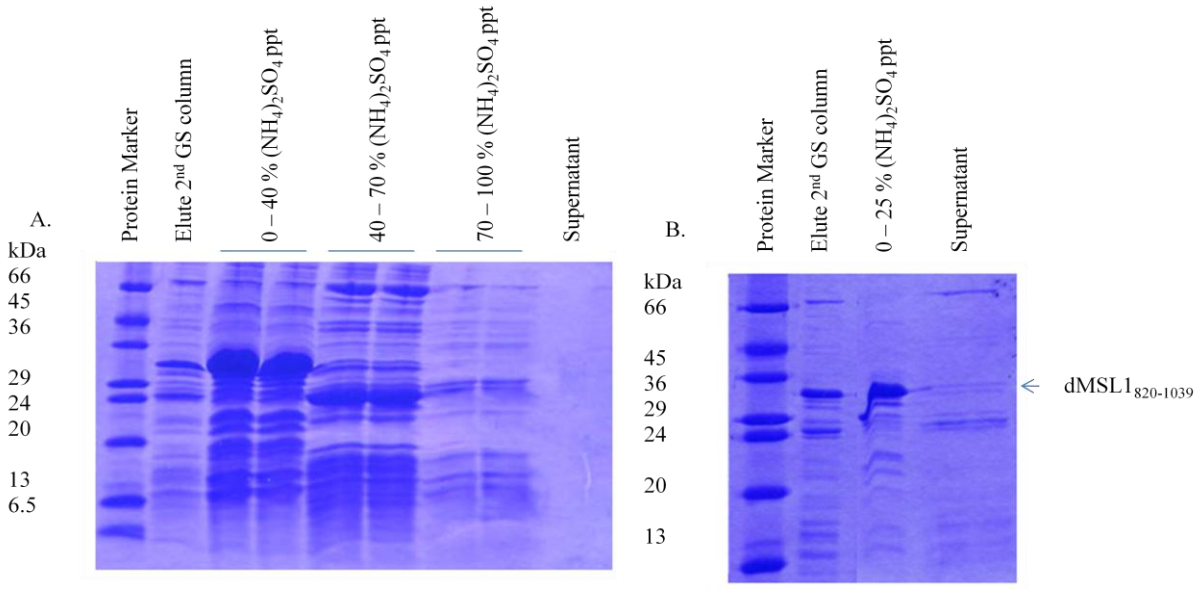


Figure 3.8: Ammonium sulphate precipitation of partially purified dMSL1$_{820-1039}$. (A). SDS-PAGE of the precipitants from an ammonium sulphate gradient from 0 – 40 %, 40 – 70 %, 70 – 100 %, and the final supernatant. Protein marker, Elute 2nd GST analyte, 0 – 40 % $(NH_4)_2SO_4$ precipitation, 0 – 40 % $(NH_4)_2SO_4$ precipitation, 40 - 70 % $(NH_4)_2SO_4$ precipitation, 40 - 70 % $(NH_4)_2SO_4$ precipitation, 70 – 100 % $(NH_4)_2SO_4$ precipitation, 70 – 100 % $(NH_4)_2SO_4$ precipitation, supernatant (final solution). (B). SDS-PAGE of optimized ammonium sulphate precipitation of dMSL1$_{820-1039}$ protein solution. Protein markers, elute 2nd GS affinity column, 0 – 25 % $(NH_4)_2SO_4$ precipitation, supernatant (final solution). Note $(NH_4)_2SO_4$ ppt stands for the precipitant resulting from ammonium sulphate precipitation.

Table 3.1: $dMSL1_{820-1039}$ ammonium sulphate precipitation.

	Original solution	40 % $(NH_4)_2SO_4$ ppt (resuspended)	40 % $(NH_4)_2SO_4$ supernatant	Original solution	25 % $(NH_4)_2SO_4$ ppt (resuspended)	25 % $(NH_4)_2SO_4$ supernatant
Abs_{280} nm	7.0	1.4	5.2	6.0	1.4	4.6
Abs_{260} nm	8.0	0.68	7.0	9.0	0.38	8.8
Abs_{280} nm / Abs_{260} nm ratio	1:1	2:1	0.8:1	0.7:1	4:1	0.5:1

Note $(NH_4)_2SO_4$ ppt stands for the precipitant resulting from ammonium sulphate precipitation. Samples were diluted tenfold to obtain accurate measurements.

The theoretical pI for $dMSL1_{820-1039}$ is predicted to be 5.5; consequently, anion exchange chromatography (Source Q) should be the preferred type of purification. Numerous experiments on a Source Q anion exchange column were attempted, varying the type of buffer, the pH, and the amount of NaCl in buffer A (equilibrating / loading buffer). In each case, both the protein and the Source Q column were equilibrated in identical buffers. It was determined that $dMSL_{820-1039}$ could not be further purified by anion exchange chromatography as it did not strongly interact with the stationary phase. Moreover, the analyte from Source Q columns typically produced a dilute unstable protein solution (data not shown). For these reasons cation exchange chromatography (Source S) was attempted. Source S equilibrated with 25 mM Tris-HCl (pH 8.0), 25 mM NaCl, and 5 mM DTT was found to further purify $dMSL1_{820-1029}$. Tris-HCl buffer (pH 8.0), is not an ideal buffer for Source S or a typical pH because the ammonium group of Tris might interact with the negative charge of the stationary phase. The chromatogram of the cation exchange chromatography showed that $dMSL1_{820-1039}$ interacted with the stationary phase and eluted as a single fraction at a unique elution buffer percentage, 20 % - 25 % of 1.0 M NaCl or 27.6 mS/cm – 32 mS/cm (Fig. 3.9). The results indicated that the dMSL1 C-terminal domain had been purified to homogeneity as shown by SDS-PAGE. Furthermore, the Abs_{280} to Abs_{260} ratio was found to improve from 1.4:1 to 4:1 (Table 3.2).

Size exclusion chromatography (Superdex 75) was utilized for a final cleaning of $dMSL1_{820-1039}$ and determination of the quaternary structure. Following cation exchange chromatography, $dMSL1_{820-1039}$ protein solution was dialysed into 20 mM Tris-HCl (pH 8.0),

100 mM NaCl, and 5 mM DTT and then concentrated to 8 mg/mL. dMSL1$_{820\text{-}1039}$ was determined to have an elution volume (V_e) of 12.0 mL, which corresponds to the expected size of a dMSL1$_{820\text{-}1039}$ monomer, approximately 25 kDa (Fig. 3.10). SDS-PAGE verified that dMSL1$_{820\text{-}1039}$ was purified to homogeneity, enabling protein crystallization trials (Fig. 3.10). Overall the purification of dMSL1$_{820\text{-}1039}$ led to an improvement of the Abs$_{280}$ to Abs$_{260}$ ratio, which was mirrored with the step-wise removal of protein impurities and the resultant homogenous dMSL1$_{820\text{-}1039}$ protein solution (Table 3.2 and Fig. 3.11). The In addition to SDS-PAGE, a UV spectrum enables a means to determine the quality of the protein purification.

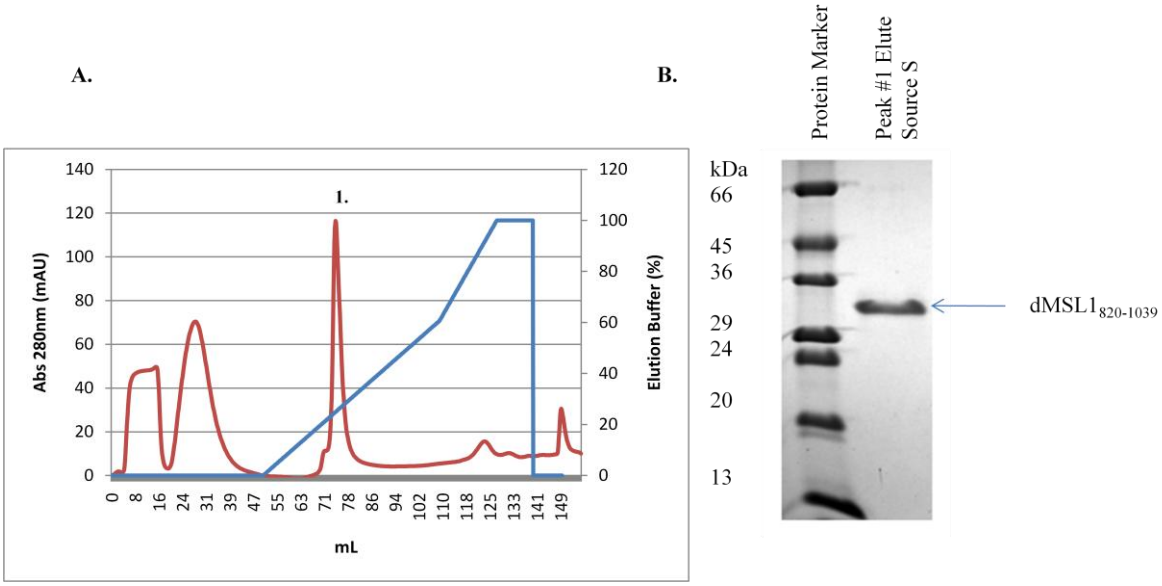


Figure 3.9: dMSL1$_{820\text{-}1039}$ purification by Source S cation exchange. (A). Source S cation exchange chromatogram of dMSL1$_{820\text{-}1039}$. Dialyzed dMSL1$_{820\text{-}1039}$ protein solution (25 mM Tris-HCl [pH 8.0], 25 mM NaCl, and 5 mM DTT) applied to the ion exchange column equalled 20 mL. The collected fraction was 6 mL eluting at 20 % - 25 % of 1.0 M NaCl. Elution conditions were a concentration gradient, 0 % to 50 %, of 25 mM Tris (pH 8.0), 1 M NaCl, and 5 mM DTT Absorbance (280 nm) is displayed in maroon and the elution buffer percent in blue. (B). SDS-PAGE of dMSL1$_{820\text{-}1039}$ purification by Source S. Protein marker, peak #1 eluted Source S (volume 70 – 78 mL).

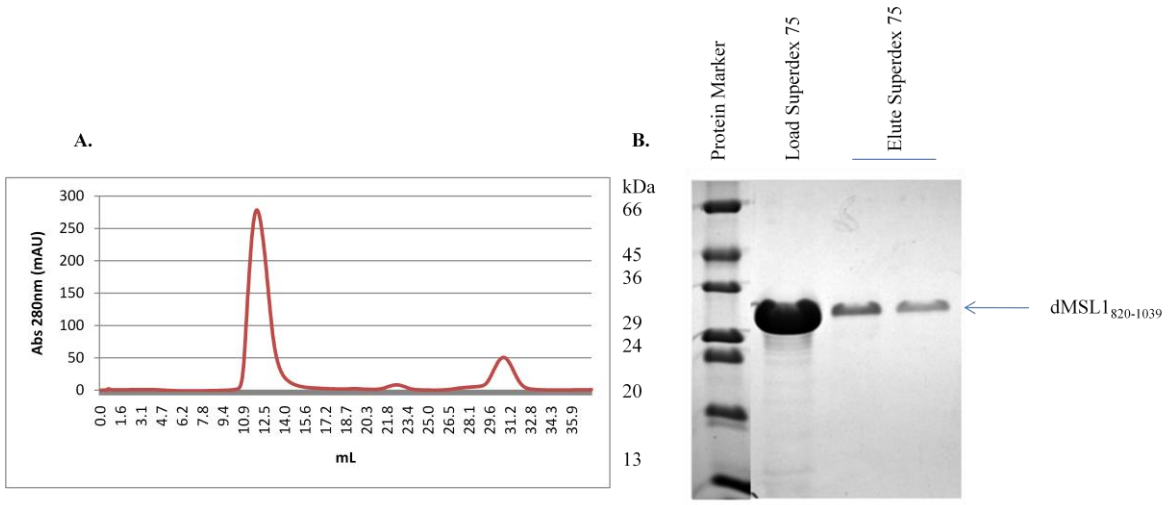


Figure 3.10: dMSL1$_{820\text{-}1039}$ purification by Superdex 75 size exclusion chromatography. (A). Superdex 75 Size Exclusion chromatogram of dMSL1$_{820\text{-}1039}$. The loaded protein solution was 1 mL (4 mg/mL). The collected fraction was 3.0 mL. Absorbance (280 nm) is displayed in maroon. The void volume (V_o) was 7.84 mL, with the total column volume (V_t) being 24 mL. (B). SDS-PAGE of dMSL1$_{820\text{-}1039}$ purification by Superdex 75. Protein marker, load Superdex 75 (4 mg/mL), elute Superdex 75 (V_e of 11.0 – 11.5), elute Superdex 75 (V_e 11.5 – 14.0 mL).

Table 3.2: Spectroscopic measurements through-out the purification of dMSL1$_{820\text{-}1039}$ from a single 4 litre recombinant protein preparation.

Identification	Abs280nm	Abs260nm	280 / 260 ratio
Elute from 1st GS affinity column	3.1	2.5	1.2:1
Elute from 2nd GS affinity column	1.8	0.9	2:1
Eluted GST from 2nd GS column	1.8	2.1	0.9:1
25 % $AmSO_4$ ppt (resuspended)	0.7	0.5	1.4:1
25 % $AmSO_4$ supernatant	2.3	2.5	1:1
Load Source S	0.7	0.5	1.4:1
Concentrated elute of Source S	1.5	0.4	4:1
Elute of Superdex 75	0.7	0.3	2.2:1

Note samples were diluted tenfold to obtain accurate measurements.

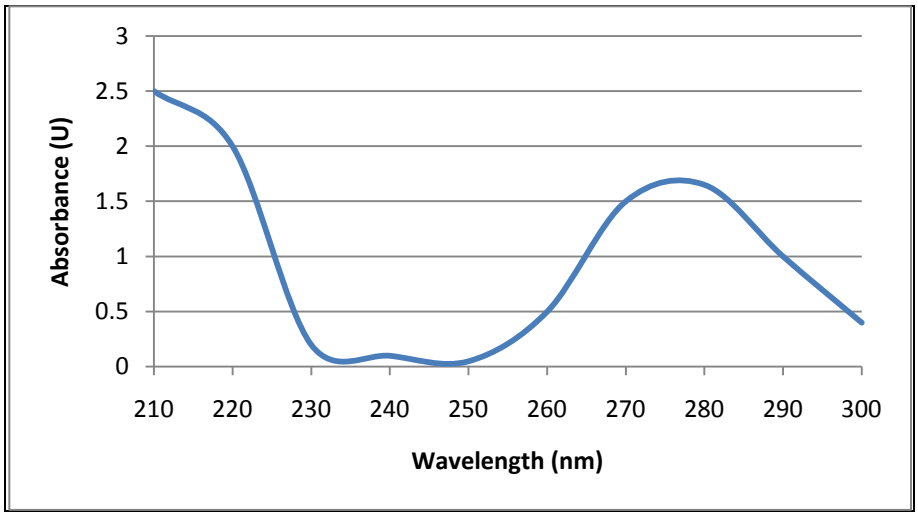


Figure 3.11: Far-UV spectrum of dMSL1$_{820\text{-}1039}$. Superdex 75 size exclusion column analyte was 1.5 mg/mL. The Far-UV spectrum was from 210 – 300 nm.

The Far-UV spectrum for dMSL1$_{820\text{-}1039}$ illustrated it was a homogenous protein solution, because a single peak at 280 nm which lacked a significant shoulder at 260 nm was observed. From a 4 litre large-scale protein preparation 20 mg of purified soluble dMSL1$_{820\text{-}1039}$ protein was produced.

Numerous crystallization conditions were tried; several commercially available sparse matrix crystallization kits were tried they included Wizard I and II, JCSG+ Suite, HR2-110 Suite, and the PACT Suite. The protein concentration used for crystallization was approximately 8 mg/mL. The crystallization buffer used was 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM TCEP, and 10 % glycerol. Glycerol was added to help maintain protein stability (Sousa, 1995). So far no protein crystals have been obtained, with the major problem being related to the instability of the dMSL1$_{820\text{-}1039}$ protein. Storage of purified dMSL1$_{820\text{-}1039}$ at 4 °C for longer than 48 hrs consistently resulted in almost complete degradation of the original protein band as seen with SDS-PAGE (data not shown). Moreover, preliminary tests of secondary structure and stability, by circular dichroism and limited proteolysis (trypsin digestion) indicted an unstable protein as both tests failed (data not shown). However, promising precipitants and phase separations have been found indicating the possibility of obtaining a protein crystal. The most promising precipitant found was 2-methyl-2,4-pentanediol commonly known as MPD.

3.3 Cloning C-terminal construct of *Homo sapiens* MSL1

Four unique constructs were designed for the C-terminal domain of *Homo sapiens* MSL1 (hMSL1) based on the MSL1 sequence alignment (Fig 3.1). The first construct encoded 149 amino acids and was termed $hMSL1_{465-614}$ (Fig 3.1 & Fig 3.12). The second construct was 210 amino acids in length, termed $hMSL1_{404-614}$. The third construct encoded 250 amino acids in length, termed $hMSL1_{364-614}$. The last was 352 amino acids in length, termed $hMSL1_{262-614}$.

Individual constructs were generated through the use of unique synthetic primers. All of the PCR products were designed to have a *Bam*HI restriction site and a *Sal*I restriction site. Cleavage of the PCR products and the pGEX-6P-3 vector with *Bam*HI and *Sal*I generated compatible ends, facilitates insertion of the inserts of interest and correct orientation. Clean double digestion with *Bam*HI and *Sal*I was difficult to achieve, but this was overcome by first digesting with *Sal*I for 4 hrs and then *Bam*HI for 16 hrs. The subsequent ligation experiments were also difficult, as single colonies bearing the recombinant plasmid were rarely observed. Ligation reaction mixtures were incubated with XL-1 competent cells, previously determined to possess a transformation efficiency of 4.7×10^5 colonies per μg of DNA, at 4 °C for 45 min. Transformation was found not to be the problem, which was determined using positive and negative controls. The positive controls used were transformation of XL-1 competent cells with uncut pGEX-6P-3 vector, and the negative control was XL-1 competent cells alone.

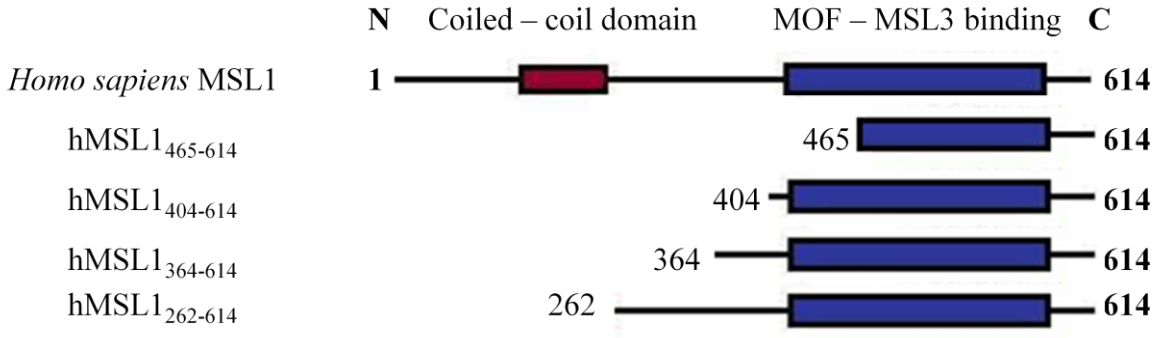


Figure 3.12: Schematic drawing of *Homo sapiens* MSL1 recombinant protein constructs that were studied. The full length hMSL1 protein is illustrated at the top. The numbers at the beginning and end of the constructs identify the amino acids of hMSL1 and indicate the hMSL1 fragment that was over-expressed and purified.

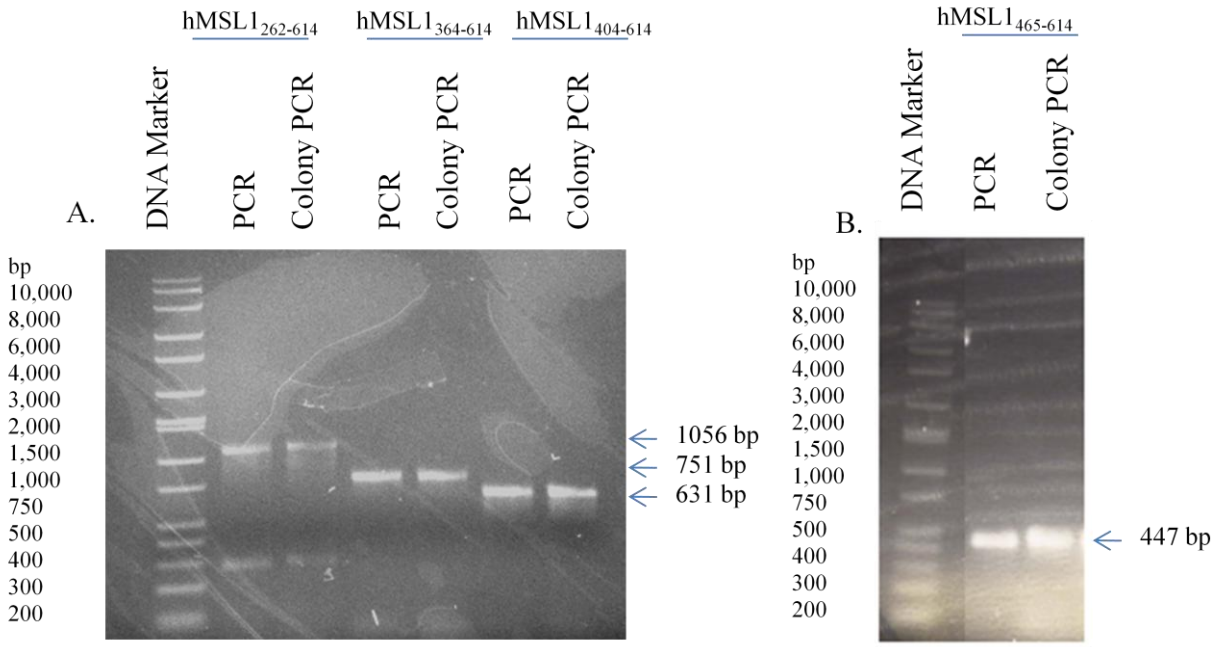


Figure 3.13: Colony PCR on the four *Homo sapiens* MSL1 constructs cloned. (A). Agarose gel of the cloned *Homo sapiens* MSL1 gene fragments and verification of correct ligation and transformation into XL-1 *E. coli* cells with colony PCR. DNA marker (Wide Range), PCR amplified $hMSL1_{465-614}$, Colony PCR product $hMSL1_{465-614}$, PCR amplified $hMSL1_{404-614}$, Colony PCR product $hMSL1_{404-614}$, PCR amplified $hMSL1_{364-614}$, Colony PCR product $hMSL1_{364-614}$. (B). Agarose gel of the cloned *Homo sapiens* MSL1 gene fragments and verification of correct ligation and transformation into XL-1 *E. coli* cells with colony PCR. DNA marker (Wide Range), PCR amplified $hMSL1_{262-614}$, Colony PCR product $hMSL1_{262-614}$.

Eventually, however, successful ligation/transformation was achieved for each of the four constructs. Verification of the presence of the desired gene fragment in each of the plasmids cloned was carried out by colony PCR (Fig. 3.13). All four constructs were subsequently verified by DNA sequencing.

Small-scale protein expression trials were attempted for each of the hMSL1 constructs cloned. All of the protein expression trials failed; none of the trials resulted in a soluble recombinant protein that would bind to GS beads. The most promising result was found for the construct $hMSL1_{465-614}$, which was found to be expressed, however, only as an insoluble protein (data not shown).

3.4 Optimization of protein expression and purification of dMSL3$_{186-512}$

The *Drosophila* MSL3 (dMSL3) construct was designed to contain the entire MRG domain, and encompass a stable C-terminal domain (Fig. 3.15) (Morales *et al*., 2005; Zhang *et al*., 2006a). The dMSL3 construct was 981 base pairs or 327 amino acids, termed dMSL3$_{186-512}$ (Fig 3.14 & 3.15). *Drosophila melanogaster* MSL3 homologues were identified by BLAST-P (3.19). The *Homo sapiens* homologue had a sequence identity of 25 % to dMSL3$_{186-512}$ and an E value of $1x10^{-11}$, and the *Mus musculus* homologue had a sequence identity of 24 % to dMSL3$_{186-512}$ and an E value of $2x10^{-9}$. Various *Drosophila* species reported 90 % and above in sequence identity including *simulans* and *yakuba*; however, the majority of the *Drosophila* species including *ananassae* and *pseudoobscura* had a sequence identity between 60 – 70 %. The paralogue of dMSL3 MRG15 was determined to have a sequence identity of 30 % and an E value of $1x10^{-4}$. Alignment of the protein sequences of *Drosophila* MSL3 homologues, paralogues, and orthologues revealed conservation of the MRG domain (Fig. 3.15). Moreover, the secondary structure of the solved *Homo sapiens* MRG15 structure was overlaid with the aligned proteins' sequences, which showed the conservation of the reported secondary structure of the MRG domain (refer to section 1.5.1). The gene fragment of dMSL3$_{186-512}$ was successful cloned into the pGEX-6P-3 vector; both the PCR amplified gene fragment and the expression vector were double digested with *Bam*HI and *Eco*RI, generating compatible ends between the vector and the insert. Colony PCR verified the correct presence of the desired insert (Fig. 3.16). DNA sequencing further confirmed the correct orientation of the insert and that the recombinant protein to be over-expressed was in-frame, and no unwanted mutations occurred.

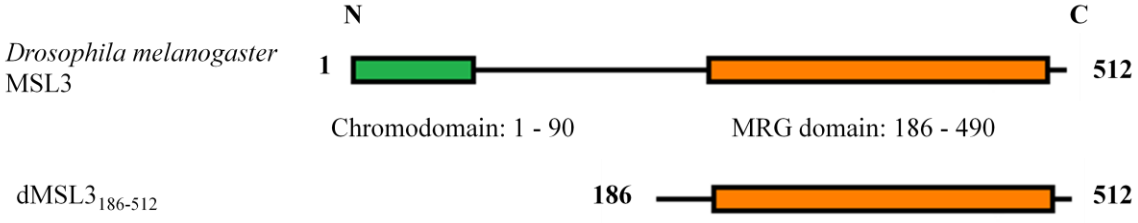


Figure 3.14: Schematic drawing of *Drosophila melanogaster* MSL3 domain structure. The recombinant protein construct that was studied is shown below the illustration of the wild-type protein. The numbers at the beginning and end of the constructs identify the amino acids of dMSL3 and indicate the dMSL3 fragment that was over-expressed and purified.

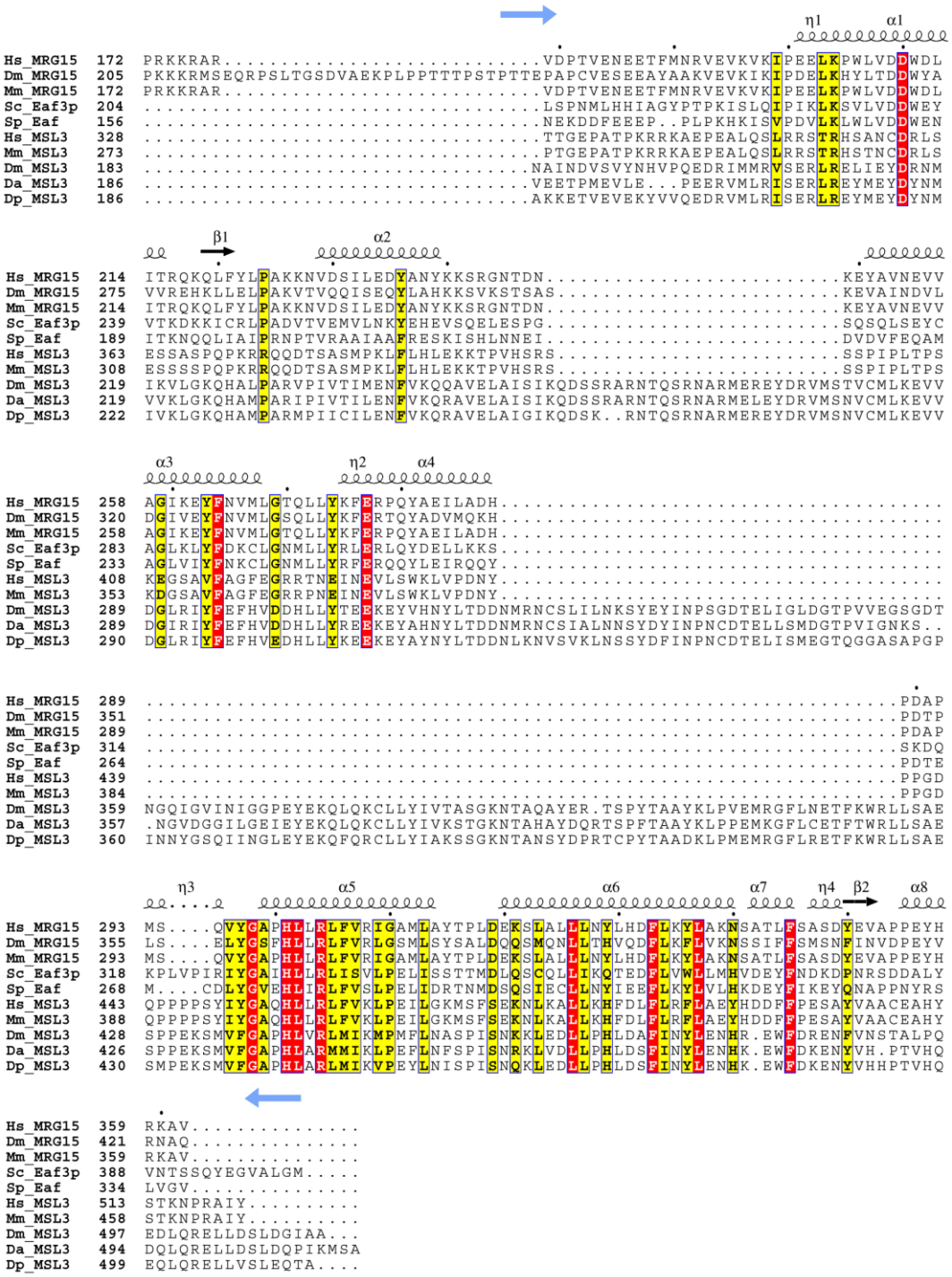


Figure 3.15: Multiple sequence alignment of *Drosophila* MSL3. Da: *Drosophila ananassae*, Dm: *Drosophila melanogaster*, Dp: *Drosophila psuedoobscura*, Hs: *Homo sapiens*, Mm: *Mus musculus*, Sc: *Saccharomyces cerevisiae*, Sp: *Schizosaccharomyces pombe*. The blue arrows indicate the primers corresponding to the cloned MSL3 constructs (Table 2.1). ClustalW2 was used for the multiple sequence alignment (Larkin *et al*., 2007). The PDB accession identification of *Homo sapiens* MRG15 is 2AQL. ESPript 2.2 was used for the alignment of the secondary protein structure against the sequence alignment (Gouet *et al*., 1999).

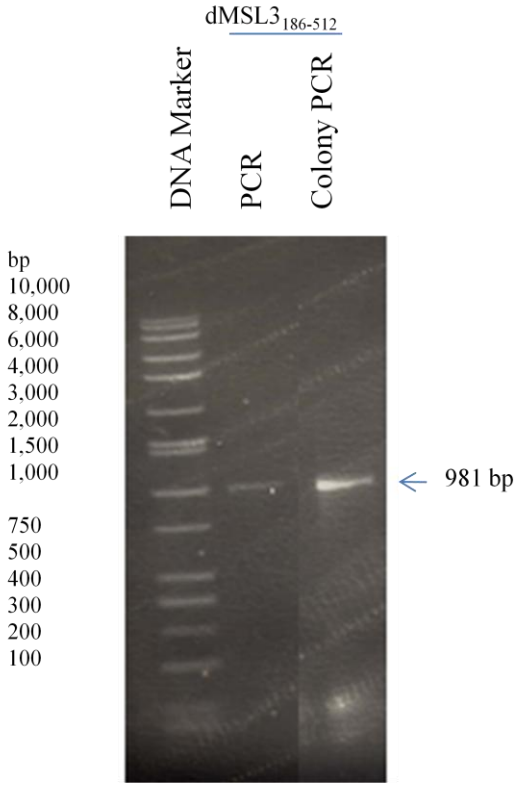


Figure 3.16: Colony PCR of *Drosophila* MSL3 constructs cloned. Agarose gel of the cloned *Drosophila* MSL3 gene fragment and verification of its correct ligation and transformation into XL-1 *E. coli* cells with colony PCR. DNA marker (Wide Range), PCR amplified $dMSL3_{186-512}$ gene fragment, Colony PCR product of transformed XL-1 cells possessing plasmid encoding the $dMSL3_{186-512}$ gene fragment.

Small-scale GST pull-downs were performed to identify optimal recombinant protein over-expression conditions. Recombinant $dMSL3_{186-512}$ protein possessing an N-terminal GST tag was successfully over-expressed as a soluble protein (Fig. 3.17). Optimal protein expression of recombinant GST-$dMSL3_{186-512}$ was determined to be under the following conditions: *E. coli* BL-21 cells, LB broth, OD_{600} of 0.8 AU, 0.1 mM IPTG, at 14 °C. Even at the optimized conditions over half of the expressed protein was found to be insoluble, indicating the difficulties of expressing a eukaryotic protein in *E. coli*. Subsequent large-scale purifications were carried out using 4 litres of LB media; the cell culture consistently resulted in a cell pellet weight of 3 g per 1 L of cell culture. The predicted molecular weight of the $dMSL3_{186-512}$ was determined to be 37.6 kDa, and the GST-$dMSL3_{186-512}$ fusion protein to be 64 kDa.

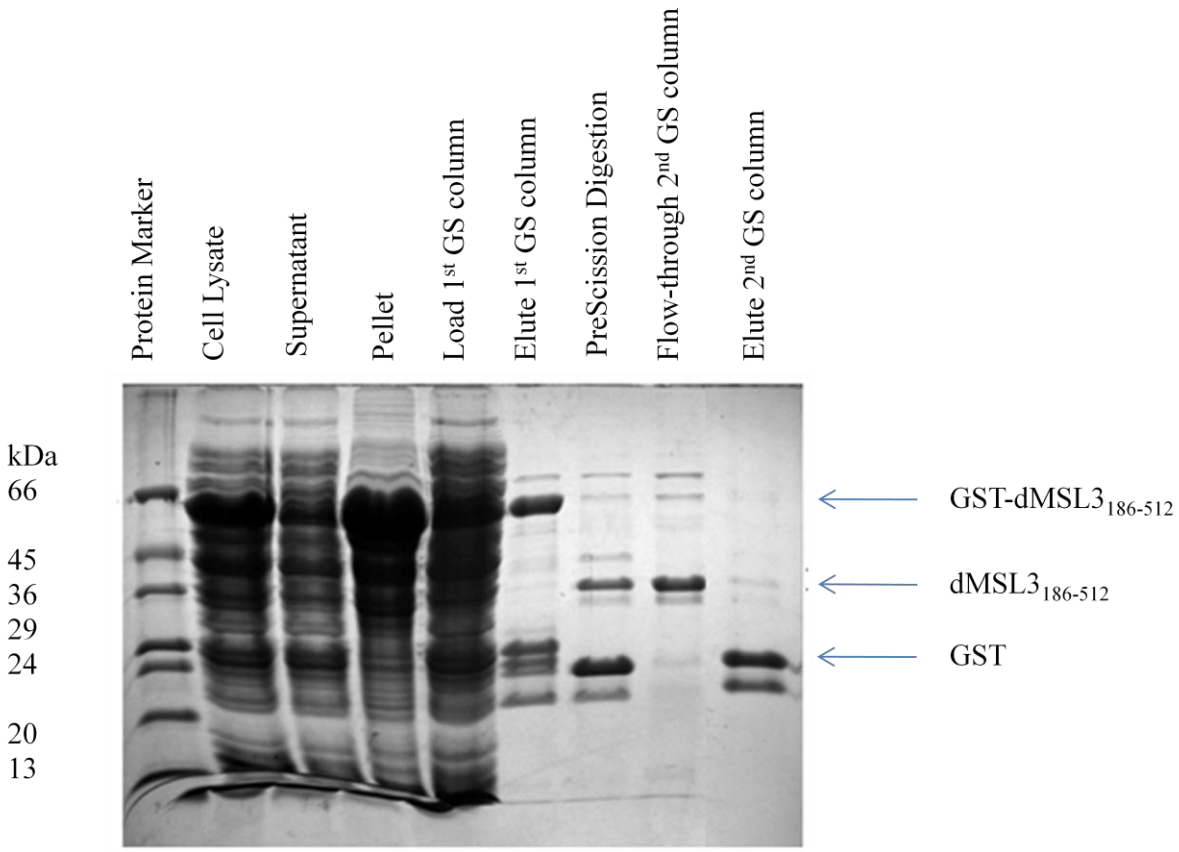


Figure 3.17: Expression and crude purification $dMSL3_{186-512}$ by affinity chromatography. SDS-PAGE of $dMSL3_{186-512}$ protein expression and purification with GS affinity chromatography. Protein was expression under the following conditions: BL21 (D3) cells grown in LB Broth, OD_{600nm} = 0.8, 0.1 mM IPTG, grown at 16 °C for 14 hrs: protein marker, cell lysate, supernatant (soluble fraction), pellet (insoluble fraction), load 1st GS column, elute 1st GS column, PreScission digestion (load 2nd GS column), flow-through 2nd GS column, elute 2nd GS column.

Utilization of batch cultures enabled large-scale purification of $dMSL3_{186-512}$. GS affinity chromatography was used to capture the GST-fusion protein and thereby facilitate purification. Hence, the GST-$dMSL3_{186-512}$ protein and the cleaved $dMSL3_{186-512}$ protein were effectively purified by GS affinity chromatography (Fig. 3.17).

The theoretical pI for $dMSL3_{186-512}$ was determined to be 5.3; therefore, anion exchange chromatography with a equilibrating / loading buffer of 25 mM Tris (pH 8.0), 25 mM NaCl, and 5 mM DTT was utilized. $dMSL3_{186-512}$ protein was found to strongly interact with the stationary phase of the Source Q anion exchange column eluting as a single peak between 24 % – 27 % of 1.0 M NaCl or 23 mS/cm – 25.8 mS/cm (Fig. 3.18)

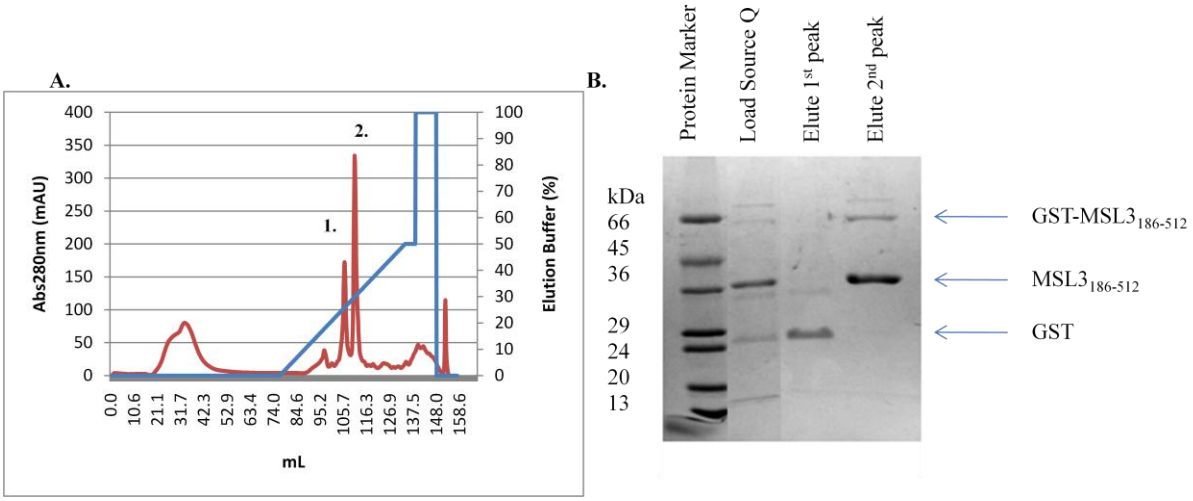


Figure 3.18: dMSL3$_{186-512}$ purification by Source Q anion exchange. (A). Source Q anion exchange chromatography of dMSL3$_{186-512}$. Dialyzed dMSL3$_{186-512}$ protein solution (25 mM Tris-HCl [pH 8.0], 25 mM NaCl, and 5 mM DTT) applied to the ion exchange column equalled 10 mL. The collected fraction was 4 mL eluting at 24 % - 27 % of 1.0 M NaCl. Elution conditions were a concentration gradient, 0 % to 50 %, of 25 mM Tris (pH 8.0), 1 M NaCl, and 5 mM DTT. (B). SDS-PAGE of dMSL3$_{186-512}$ anion exchange chromatography. Protein markers, load Source Q, elute 1st peak (volume 102 – 104 mL), elute 2nd peak (volume 106 – 110 mL).

The dMSL3$_{186-512}$ protein solution was concentrated to 2 mg/mL for size exclusion chromatography (Superdex 200). Size exclusion chromatography verified that MSL3$_{186-512}$ has a monomeric quaternary structure, a V_e of 16.0 mL, which correlated to a MW of 40 kDa (Fig. 3.19). However, the protein partially precipitated during concentration therefore crystallization trials could not be attempted.

A Far-UV spectrum revealed that the protein solution did not contain significant amounts of unwanted nucleic acid (Fig. 3.20). Therefore, the recombinant dMSL3$_{186-512}$ protein was determined to be soluble, purified to homology, a prerequisite for subsequent studies. Moreover, recombinant dMSL3$_{186-512}$ was found to be stable, as the protein could be stored at 4 °C for over 120 hrs without degradation (data not shown). In summary, the 4 litre large-scale protein prep produced 16 mg of purified soluble dMSL3$_{186-512}$ protein.

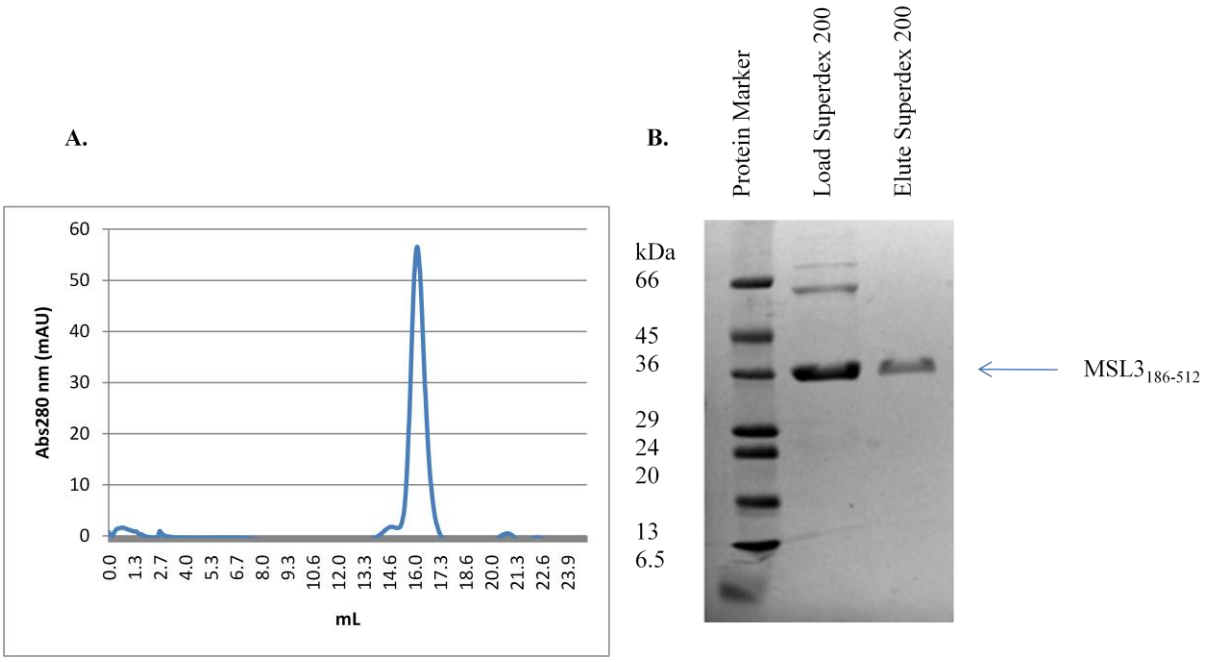


Figure 3.19: dMSL3$_{186-512}$ purification by Superdex 200 size exclusion. (A). Chromatogram of size exclusion chromatography of dMSL3$_{186-512}$ (Superdex 200 10/300). The loaded protein solution was 1 mL (2 mg/mL). The collected fraction was 1.5 mL. Absorbance (280 nm) is displayed in maroon. The void volume (V_o) was 8.04 mL, with the total column volume (V_t) being 24 mL. (B). SDS-PAGE of dMSL3$_{186-512}$ size exclusion chromatography. Protein marker, load Superdex 200, elute Superdex 200 (V_e of 15.5 – 17.0 mL).

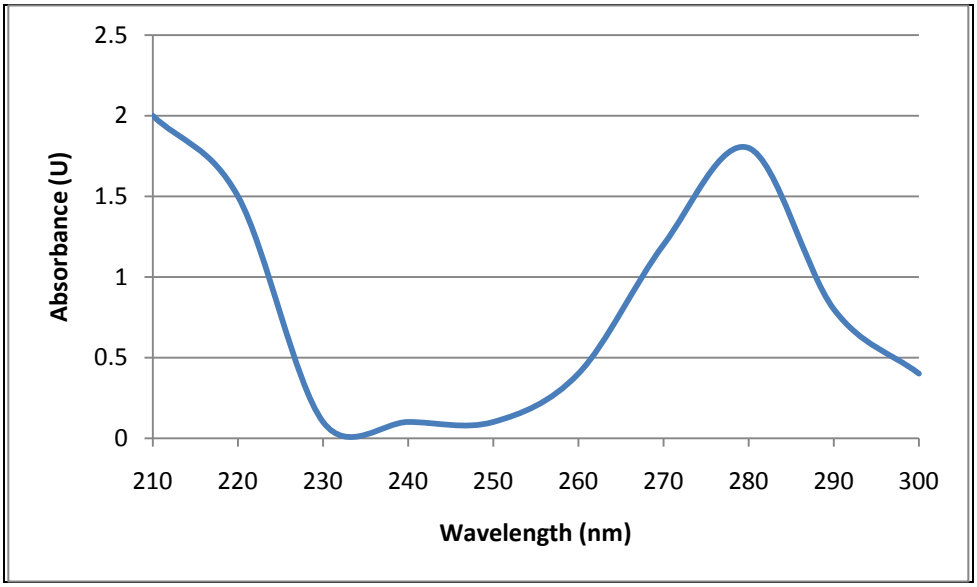


Figure 3.20: Far-UV spectrum of dMSL3$_{186-512}$. The spectrum was of the dMSL3$_{186-512}$ protein solution purified by anion exchange chromatography, and was determined to be 1.8 mg/mL.

3.5 Optimization of protein expression and purification of $MOF_{371-827}$

The cloning of the construct *Drosophila* $MOF_{371-827}$ was previously completed by Ewa Kerc (University of Saskatchewan) prior to the start of my masters' degree. The recombinant $dMOF_{371-827}$ construct contained the chromodomain, the CCHC zinger finger domain, and the MYST family HAT domain (Fig. 3.21). The CCHC zinc finger segment of MOF is thought to be involved in the protein interaction with MSL1 (Akhtar and Becker, 2000). The construct also was designed to contain the known domains necessary for acetyltransferase activity, which permitted activity assays to ascertain where or not a catalytically active C-terminal domain was generated. The amino acid sequence of $dMOF_{371-827}$ was aligned against the reported MYST HAT family members from both *Homo sapiens* and *Saccharomyces cerevisiae* (Fig. 3.22). PSI-BLAST revealed *Homo sapiens* MOF had a sequence identity of 54 % and an E value of $1x10^{-135}$ compared to the $dMOF_{371-827}$ construct. *Homo sapiens* MOZ had a sequence identity of 51 % and an E value of $1x10^{-81}$, *Homo sapiens* Tip60 had a sequence identity of 43 % and an E value of $5x10^{-81}$, and *S. cerevisiae* Esa1 had a sequence identity of 45 % and an E value of $2x10^{-73}$.

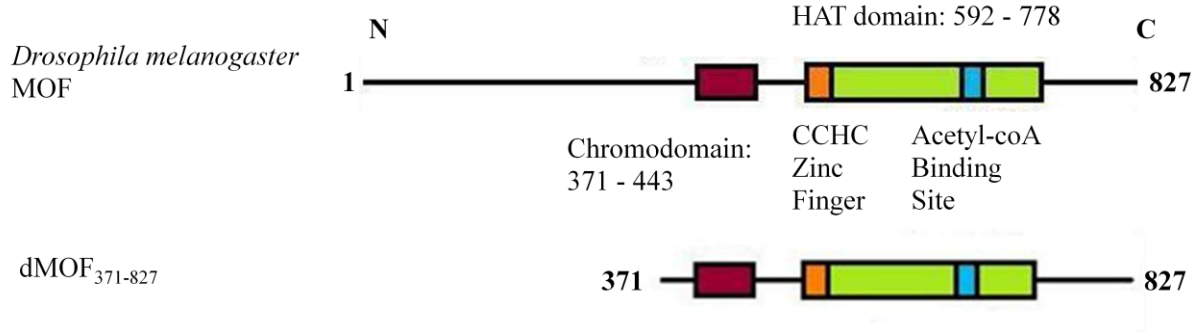


Figure 3.21: Schematic drawing of *Drosophila melanogaster* MOF domain structure. The recombinant protein construct that was studied is shown below the wild-type protein. The numbers at the beginning and end of the constructs identify the amino acids of dMOF and indicate the dMOF fragment that was over-expressed and purified.

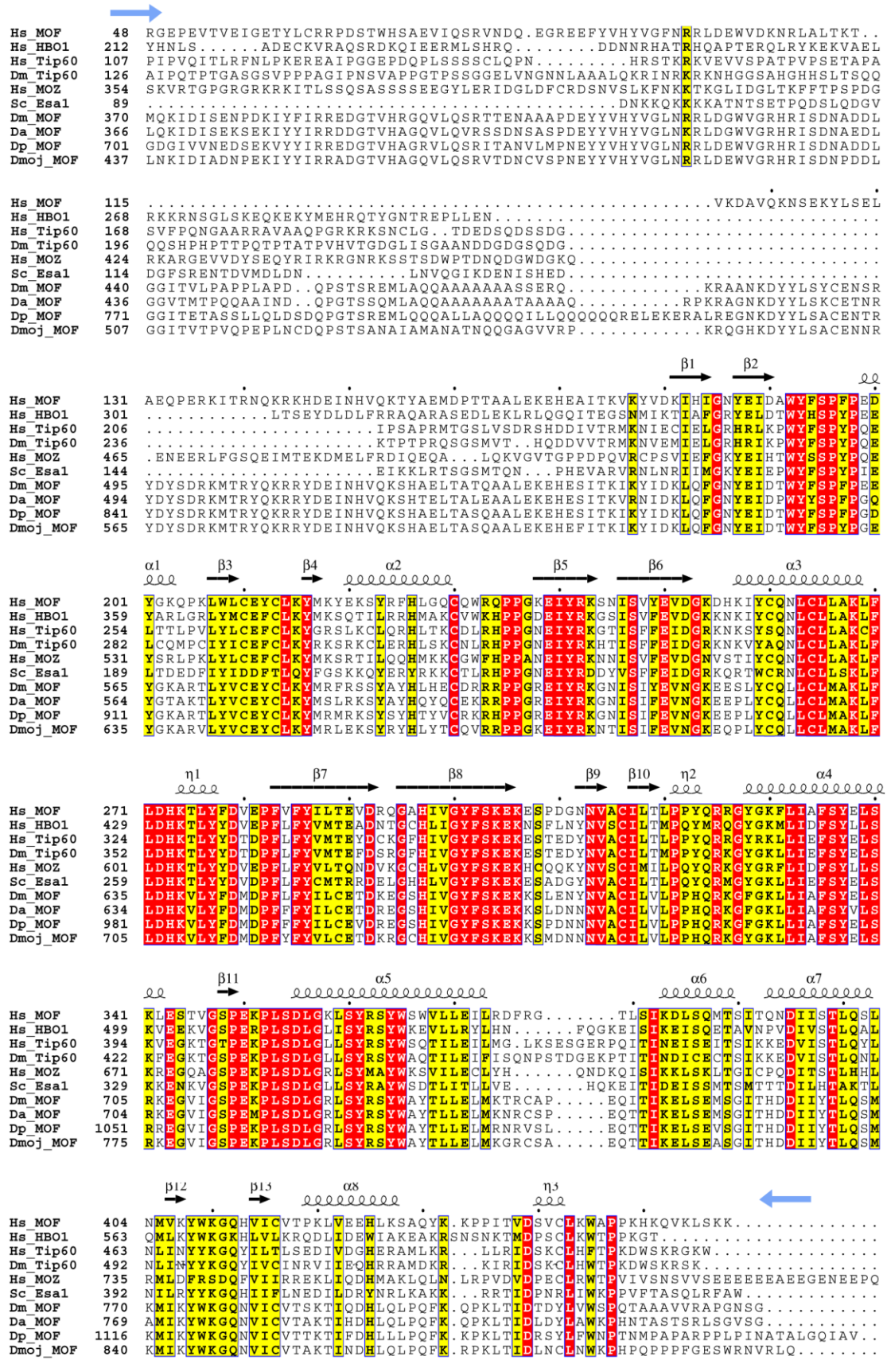

Hs_MOF 48 RGEPEVTVEIGETYLCRRPDSTWHSAEVIQSRVNDQ.EGREEFYVHYVGFNRRLDEWVDKNRLALTKT..
Hs_HBO1 212 YHNLS.....ADECKVRAQSRDKQIEERMLSHRQ.....DDNNRHATRHQAPTERQLRYKEKVAEL
Hs_Tip60 107 PIPVQITLRFNLPKEREAIPGGEPDQPLSSSSCLQPN.....HRSTKRKVEVVSPATPVPSETAPA
Dm_Tip60 126 AIPQTPTGASGSVPPPAGIPNSVAPPGTPSSGGELVNGNNLAAALQKRINRKRKNHGGSAHGHHSLTSQQ
Hs_MOZ 354 SKVRTGPGRGRKRKITLSSQSASSSSEEGYLERIDGLDFCRDSNVSLKFNKKTKGLIDGLTKFFTPSPDG
Sc_Esa1 89 .....DNKKQKKKATNTSETPQDSLQDGV
Dm_MOF 370 MQKIDISENPDKIYFIRREDGTVHRGQVLQSRTTENAAAPDEYYVHYVGLNRRLDGWVGRHRISDNADDL
Da_MOF 366 LQKIDISEKSEKIYYIRRDDGTVHAGQVLQVRSSDNSASPDEYYVHYVGLNKRLDGWVGRHRISDNAEDL
Dp_MOF 701 GDGIVVNEDSEKVYYIRREDGTVHAGRVLQSRITANVLMPNEYYVHYVGLNRRLDEWVGRHRISDNADDL
Dmoj_MOF 437 LNKIDIADNPEKIYYIRRADGTVHAGQVLQSRVTDNCVSPNEYYVHYVGLNRRLDEWVGRHRISDNPDDL
Hs_MOF 115 .....VKDAVQKNSEKYLSEL
Hs_HBO1 268 RKKRNSGLSKEQKEKYMEHRQTYGNTREPLLEN.....
Hs_Tip60 168 SVFPQNGAARRAVAAQPGRKRKSNCLG..TDEDSQDSSDG.....
Dm_Tip60 196 QQSHPHPTTPQTPTATPVHVTGDGLISGAANDDGDGSQDG.....
Hs_MOZ 424 RKARGEVVDYSEQYRIRKRGNRKSSTSDWPTDNQDGWDGKQ.....
Sc_Esa1 114 DGFSRENTDVMDLDN.....LNVQGIKDENISHED.....
Dm_MOF 440 GGITVLPAPPLAPD..QPSTSREMLAQQAAAAAAASSERQ.....KRAANKDYYLSYCENSR
Da_MOF 436 GGVTMTPQQAAIND..QPGTSSQMLAQQAAAAAAATAAAAQ.....RPKRAGNKDYYLSKCETNR
Dp_MOF 771 GGITETASSLLQLDSDQPGTSREMLQQQALLAQQQQILLQQQQQQRELEKERALREGNKDYYLSACENTR
Dmoj_MOF 507 GGITVTPVQPEPLNCDQPSTSANAIAMANATNQQGAGVVRP.....KRQGHKDYYLSACENNR
β1
β2
Hs_MOF 131 AEQPERKITRNQKRKHDEINHVQKTYAEMDPTTAALEKEHEAITKVKYVDKIHIGNYEIDAWYFSPFPED
Hs_HBO1 301 .....LTSEYDLDLFRRAQARASEDLEKLRLQGQITEGSNMIKTIAFGRYELDTWYHSPYPEE
Hs_Tip60 206 .....IPSAPRMTGSLVSDRSHDDIVTRMKNIECIELGRHRLKPWYFSPYPQE
Dm_Tip60 236 .....KTPTPRQSGSMVT..HQDDVVTRMKNVEMIELGRHRIKPWYFSPYPQE
Hs_MOZ 465 .ENEERLFGSQEIMTEKDMELFRDIQEQA...LQKVGVTGPPDPQVRCPSVIEFGKYEIHTWYSSPYPQE
Sc_Esa1 144 .....EIKKLRTSGSMTQN...PHEVARVRNLNRIIMGKYEIEPWYFSPYPIE
Dm_MOF 495 YDYSDRKMTRYQKRRYDEINHVQKSHAELTATQAALEKEHESITKIKYIDKLQFGNYEIDTWYFSPFPEE
Da_MOF 494 YDYSDRKMTRYQKRRYDEINHVQKSHTELTALEAALEKEHESITKVRNIDKLQFGNYEIDPWYYSPFPGQ
Dp_MOF 841 YDYSDRKMTRYQKRRYDEINHVQKSHAELTASQAALEKEHESITKIKYIDKLQFGNYEIDTWYFSPFPGD
Dmoj_MOF 565 YDYSDRKMTRYQKRRYDEINHVQKSHAELTASQAALEKEHEFITKIKYIDKLQFGNYEIDTWYFSPYPGE
α1
β3
β4
α2
β5
β6
α3
Hs_MOF 201 YGKQPKLWLCEYCLKYMKYEKSYRFHLGQCQWRQPPGKEIYRKSNISVYEVDGKDHKIYCQNLCLLAKLF
Hs_HBO1 359 YARLGRLYMCEFCLKYMKSQTILRRHMAKCVWKHPPGDEIYRKGSISVFEVDGKKNKIYCQNLCLLAKLF
Hs_Tip60 254 LTTLPVLYLCEFCLKYGRSLKCLQRHLTKCDLRHPPGNEIYRKGTISFFEIDGRKNKSYSQNLCLLAKCF
Dm_Tip60 282 LCQMPCIYICEFCLKYRKSRKCLERHLSKCNLRHPPGNEIYRKHTISFFEIDGRKNKVYAQNLCLLAKLF
Hs_MOZ 531 YSRLPKLYLCEFCLKYMKSRTILQQHMKKCGWFHPPANEIYRKNNISVFEVDGNVSTIYCQNLCLLAKLF
Sc_Esa1 189 LTDEDFIYIDDFTLQYFGSKKQYERYRKKCTLRHPPGNEIYRDDYVSFFEIDGRKQRTWCRNLCLLSKLF
Dm_MOF 565 YGKARTLYVCEYCLKYMRFRSSYAYHLHECDRRPPPGREIYRKGNISIYEVNGKEESLYCQLLCLMAKLF
Da_MOF 564 YGTAKTLYVCEYCLKYMSLRKSYAYHQYQCEKRRPPGREIYRKGNISIFEVNGKEEPLYCQLLCLMAKLF
Dp_MOF 911 YGKARTLYVCEYCLKYMRMRKSYSYHTYVCRKRHPPGREIYRKGNISIFEVNGKEEPLYCQLLCLMAKLF
Dmoj_MOF 635 YGKARVLYVCEYCLKYMRLEKSYRYHLYTCQVRRPPGKEIYRKNTISIFEVNGKEQPLYCQLLCLMAKLF
η1
β7
β8
β9
β10
η2
α4
Hs_MOF 271 LDHKTLYFDVEPFVFYILTEVDRQGAHIVGYFSKEKESPDGNNVACILTLPPYQRRGYGKFLIAFSYELS
Hs_HBO1 429 LDHKTLYYDVEPFLFYVMTEADNTGCHLIGYFSKEKNSFLNYNVSCILTMPQYMRQGYGKMLIDFSYLLS
Hs_Tip60 324 LDHKTLYYDTDPFLFYVMTEYDCKGRHIVGYFSKEKESTEDYNVACILTLPPYQRRGYRKLLIEFSYELS
Dm_Tip60 352 LDHKTLYYDTDPFLFYVMTEFDSRGFHIVGYFSKEKESTEDYNVACILTMPPYQRKGYGKLLIEFSYELS
Hs_MOZ 601 LDHKTLYYDVEPFLFYVLTQNDVKGCHLVGYFSKEKHCQQKYNVSCIMILPQYQRKGYGRFLIDFSYLLS
Sc_Esa1 259 LDHKTLYYDVDPFLFYCMTRRDELGHHLVGYFSKEKESADGYNVACILTLPQYQRMGYGKLLIEFSYELS
Dm_MOF 635 LDHKVLYFDMDPFLFYILCETDKEGSHIVGYFSKEKKSLENYNVACILVLPPHQRKGFGKLLIAFSYELS
Da_MOF 634 LDHKVLYFDMDPFFFYILCETDREGSHIVGYFSKEKKSLDNNNVACILVLPPHQRKGFGKLLIAFSYVLS
Dp_MOF 981 LDHKVLYFDMDPFFFYILCEVDREGSHIVGYFSKEKKSPDNNNVACILVLPPHQRKGFGKLLIAFSYELS
Dmoj_MOF 705 LDHKVLYFDMDPFYFYVLCETDKRGCHIVGYFSKEKKSMDNNNVACILVLPPHQRKGYGKLLIAFSYELS
β11
α5
α6
α7
Hs_MOF 341 KLESTVCSPEKPLSDLGKLSYRSYWSWVLLEILRDFRG.....TLSIKDLSQMTSITQNDIISTLQSL
Hs_HBO1 499 KVEEKVCSPERPLSDLGLISYRSYWKEVLLRYLHN.....FQGKEISIKEISQETAVNPVDIVSTLQAL
Hs_Tip60 394 KVEGKTGTPEKPLSDLGLLSYRSYWSQTILEILMG.LKSESGERPQITINEISEITSIKKEDVISTLQYL
Dm_Tip60 422 KFEGKTGSPEKPLSDLGLLSYRSYWAQTILEIFISQNPSTDGEKPTITINDICECTSIKKEDVISTLQNL
Hs_MOZ 671 KREGQACSPEKPLSDLGRLSYMAYWKSVILECLYH.....QNDKQISIKKLSKLTGICPQDITSTLHHL
Sc_Esa1 329 KKENKVGSPEKPLSDLGLLSYRAYWSDTLITLLVE.....HQKEITIDEISSMTSMTTTDILHTAKTL
Dm_MOF 705 RKEGVIGSPEKPLSDLGRLSYRSYWAYTLLELMKTRCAP.....EQITIKELSEMSGITHDDIIYTLQSM
Da_MOF 704 RKEGVIGSPEMPLSDLGRLSYRSYWAYTLLELMKNRCSP.....EQTTIKELSEMSGITHDDIIYTLQSM
Dp_MOF 1051 RREGVIGSPEKPLSDLGRLSYRSYWAYTLLELMRNRVSL.....EQTTIKELSEVSGITHDDIIYTLQSM
Dmoj_MOF 775 RKEGVIGSPEKPLSDLGRLSYRSYWAYTLLELMKGRCSA.....EQTTIKELSEASGITHDDIIYTLQSM
β12
β13
α8
η3
Hs_MOF 404 NMVKYWKGQHVICVTPKLVEEHLKSAQYK.KPPITVDSVCLKWAPPKHKQVKLSKK.....
Hs_HBO1 563 QMLKYWKGKHLVLKRQDLIDEWIAKEAKRSNSNKTMDPSCLKWTPPKGT.....
Hs_Tip60 463 NLINYYKGQYILTLSEDIVDGHERAMLKR...LLRIDSKCLHFTPKDWSKRGKW.....
Dm_Tip60 492 NLINYYKGQYIVCINRVIIEQHRRAMDKR...KIRIDSKCLHWTPKDWSKRSK.....
Hs_MOZ 735 RMLDFRSDQFVIIRREKLIQDHMAKLQLN.LRPVDVDPECLRWTPVIVSNSVVSEEEEEEAEEGENEEPQ
Sc_Esa1 392 NILRYYKGQHIIFLNEDILDRYNRLKAKK...RRTIDPNRLIWKPPVFTASQLRFAW.....
Dm_MOF 770 KMIKYWKGQNVICVTSKTIQDHLQLPQFK.QPKLTIDTDYLVWSEQTAAAVVRAPGNSG.....
Da_MOF 769 AMIKYWKGQNVICVTAKTIHDHLQLPQFK.QPKLTIDLDYLAWKPHNTASTSRLSGVSG.....
Dp_MOF 1116 KMIKYWKGQNVICVTTKTIFDHLLLPQFK.KPKLTIDRSYLFWNPTNMPAPARPPLPINATALGQIAV..
Dmoj_MOF 840 KMIKYWKGQNVICVTAKTINDHLQLPQFK.RPKLTIDLNCLNWKPHPQPPPFGESWRNVRLQ.....

Figure 3.22: Multiple sequence alignment of *Drosophila* MOF homologues. Hs: *Homo sapiens*, Sc: *Saccharomyces cerevisiae*, Da: *Drosophila ananassae*, Dm: *Drosophila melanogaster*, Dmoj: *Drosophila mojavensis*, and Dp: *Drosophila pseudoobscura*. The blue arrows indicate the primers corresponding to the cloned MSL3 constructs (Table 2.1). ClustalW2 was used for the multiple sequence alignment (Larkin *et al*., 2007). The PDB accession identification of *Homo sapiens* MOF is 2PQ8. ESPript 2.2 was used for the alignment of the secondary protein structure against the sequence alignment (Gouet *et al*., 1999).

Recombinant GST-dMOF$_{371-827}$ was successfully over-expressed when BL-21 cells were used and the cell culture was grown in Hyper Broth with protein expression conditions of OD_{600} of 0.6 AU, 0.1 mM IPTG, 16 °C for 14 hrs. Protein expression with the Hyper Broth growth media was determined to be enhanced compared to LB broth, Power Broth, Superior Broth, and Turbo Broth. Large-scale purifications were done using 4 – 6 litres of LB cell culture, which consistently resulted in a cell pellet weight of 15 g per litre of cell culture. The predicted molecular weight of GST-dMOF$_{371-827}$ was determined to be 79.4 kDa and the cleaved product dMOF$_{371-827}$ was calculated to be 53 kDa. Crude purification of GST-dMOF$_{371-827}$ was effectively achieved with GS affinity chromatography (Fig. 3.23). Analysis of the fractions from first GS affinity column verified the capture and purification of GST-dMOF$_{371-827}$. Moreover, SDS-PAGE of the eluate from the GS affinity column revealed that the protein solution contained a significant portion of both GST-dMOF$_{371-827}$ and dMOF$_{371-827}$ (Fig. 3.23). Cleavage of the GST tag by means of PreScission protease combined with a second GS affinity column resulted in successful purification of dMOF$_{371-827}$ and effective removal of the GST tag (Fig. 3.23).

dMOF$_{371-827}$ was determined to have a theoretical pI of 8.5; further purification, however, was determined to be by means of anion exchange chromatography and not cation exchange chromatography. Anion exchange chromatography (Source Q) with a loading / equilibrating buffer consisting of 25 mM bis-Tris propane (pH 7.0), 25 mM NaCl, and 5 mM DTT resulted in dMOF$_{371-827}$ to effectively bind to the stationary phase and elute as a single peak at 40 – 46.5 % 1.0 M NaCl (Fig. 3.24). SDS-PAGE analysis of the anion exchange chromatography revealed removal of numerous impurities (Fig. 3.24). Interestingly, spectrophotometric analysis of the different purification stages illustrated a significant amount

of nucleic acid present all through purification (Table 3.3 & Fig. 3.25). Furthermore, none of the different purification techniques removed the nucleic acids.

Histone acetyltransferase assays were carried out on GST-dMOF$_{371-827}$ after the first GS affinity column, on dMOF$_{371-827}$ protein solution eluted from the second GS affinity column, and MOF$_{371-827}$ purified by Source Q anion exchange column (Fig. 3.26). The HAT assays were all completed by Dr. Vikki Weake (Stowers Research Institute, University of Kansas). The results demonstrated that the purified recombinant GST-MOF$_{371-827}$ and dMOF$_{371-827}$ possessed high acetyltransferase activity in the presence of HeLa core histones 21,104 and 14,655 cpm respectively. The recombinant dMOF$_{371-827}$ further purified with ion exchange (Source Q) was found to have lost its HAT activity, 859 cpm. The two negative controls GST alone and core histone proteins alone were both determined to low activity 1785 and 530 cpm respectively. The level of auto-acetylation was found to be negligible as the HAT activity was never above the measured level for core histone proteins alone.

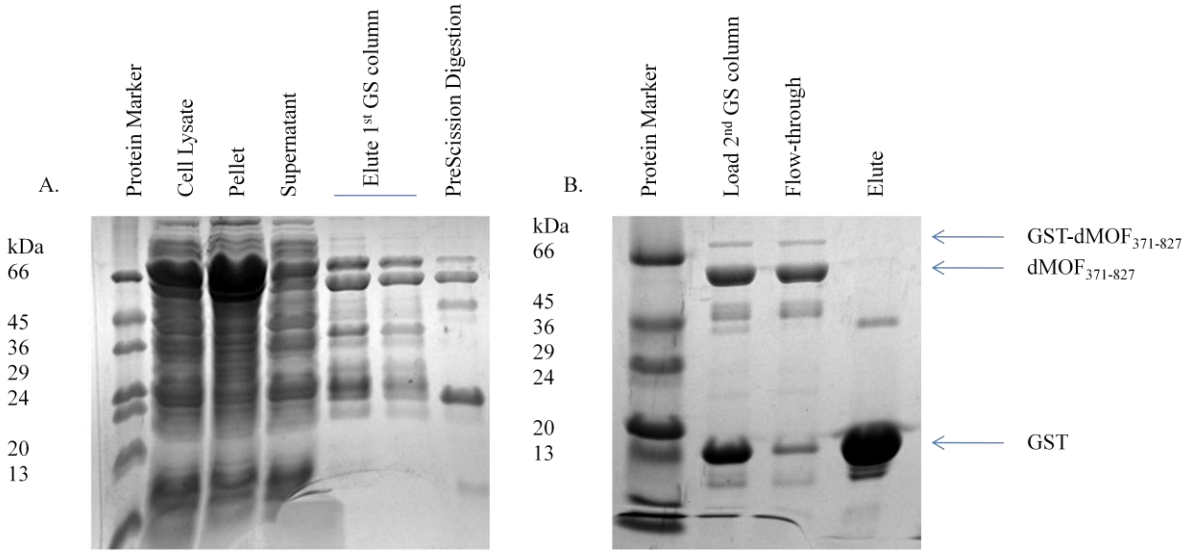


Figure 3.23: dMOF$_{371-827}$ protein expression and purification by affinity chromatography. (A). SDS-PAGE of dMOF$_{371-827}$ protein expression and purification with the 1st GS affinity column. Protein was expression under the following conditions: BL21 (D3) cells grown in Hyper Broth^{TM}, OD$_{600nm}$ = 0.6, 0.1 mM IPTG, grown at 16 °C for 14 hrs. Protein marker, cell lysate, pellet (insoluble fraction), supernatant (soluble fraction/load GS column), elute 1st GS column, elute 1st GS column, PreScission protease digestion. (B). SDS-PAGE of 2nd GS affinity column. Protein marker, load 2nd GS column, flow-through (2nd GS column), elute (2nd GS column).

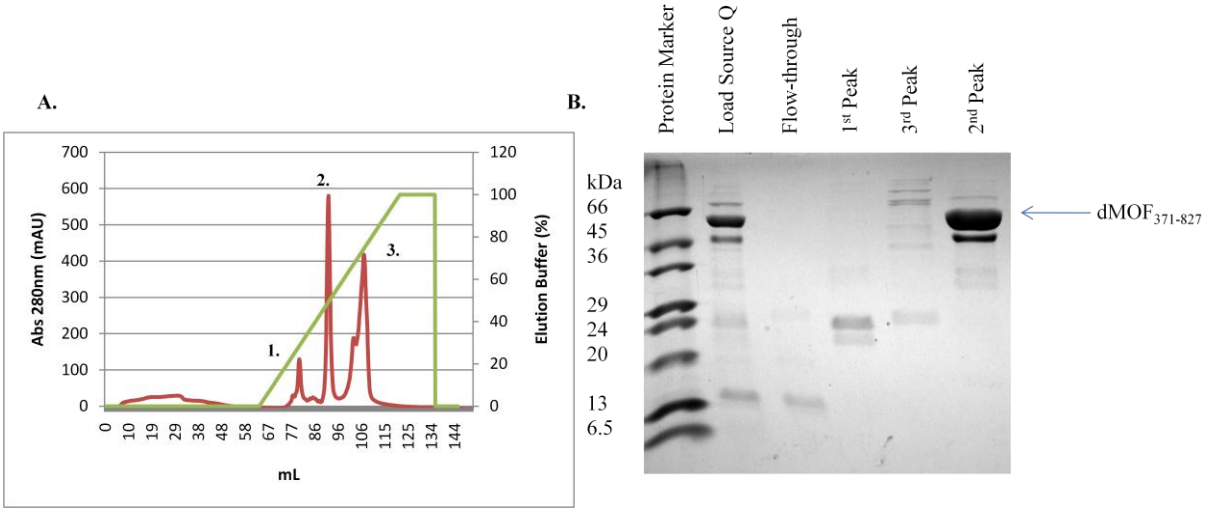


Figure 3.24: $dMOF_{371-827}$ purification by Source Q anion exchange chromatography. (A). Chromatogram of $MOF_{371-827}$ purification by Source Q. Dialyzed $dMOF_{371-827}$ protein solution (25 mM bis-Tris [pH 7.0], 25 mM NaCl, and 5 mM DTT) applied to the ion exchange column equalled 15 mL. The collected fraction was 4 mL eluting at 40 % - 46.5 % of 1.0 M NaCl. Elution conditions were a concentration gradient, 0 % to 100 %, of 25 mM bis-Tris (pH 8.0), 1 M NaCl, and 5 mM DTT. (B). SDS-PAGE of $dMOF_{371-827}$ purification by Source Q. Protein markers, load Source Q, flow-through (8 – 50 mL), 1st peak (80 – 84 mL), 3rd peak (100 – 112 mL), 2nd peak (90 – 94 mL).

Table 3.3: Spectroscopic measurements through-out the purification of $dMOF_{371-827}$ from a single 4 litre recombinant protein preparation.

Identification	Abs260 nm	Abs280 nm
Elute from 1st GS affinity column	2.1	2.2
Elute from 2nd GS affinity column	0.4	0.3
Eluted GST sepharose from 2nd GS column	0.7	0.8
Elute Source Q	0.3	0.2

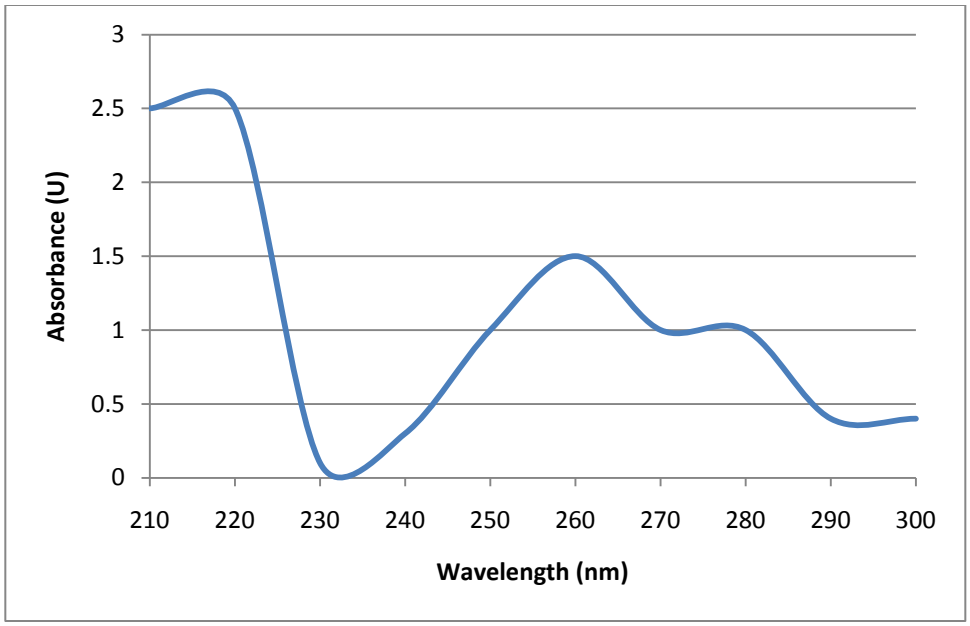
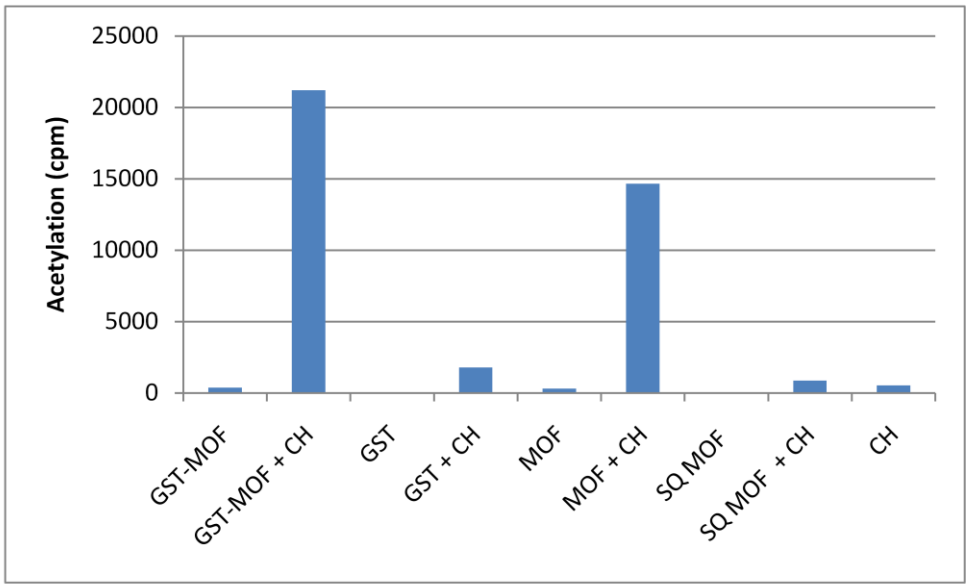


Figure 3.25: Far-UV spectrum of $dMOF_{371-827}$. The spectrum was of the $dMOF_{371-827}$ protein solution purified by anion exchange chromatography, and was determined to be 1.5 mg/mL.

Moreover, in the absence of core histones minimal acetyltransferase activity was measured indicating very low levels of auto-acetylation. Therefore, the majority of the measured acetyltransferase activity in the presence of core histones was directed towards the histone proteins and not $dMOF_{371-827}$. GST alone and core histones alone were the negative controls verifying acetyltransferase activity was via $dMOF_{371-827}$. Surprisingly, $dMOF_{371-827}$ purified by anion exchange chromatography no longer possessed acetyltransferase activity, indicating an unwanted disruption of $dMOF_{371-827}$ occurred.

The $dMOF_{371-827}$ protein solution was typically concentrated to 2.0 mg/mL for size exclusion chromatography (Superdex 200). The majority of $dMOF_{371-827}$ eluted at a V_e of 10.5 mL, which corresponds to approximately 230 kDa. This suggested that the recombinant $dMOF_{371-827}$ protein had a tetrameric quaternary structure (Table 3.4). Numerous trials to attempt to remove the quaternary structure by means of detergents and other additives failed to result in a change in the state of the protein. The two main additives tried were zinc (for the CCHC zinc finger) and acetyl-CoA (for the MYST HAT domain). Moreover, purification by size exclusion chromatography drastically reduced the overall protein concentration and subsequently the total amount of the purified protein (Fig. 3.27).

A.



B.

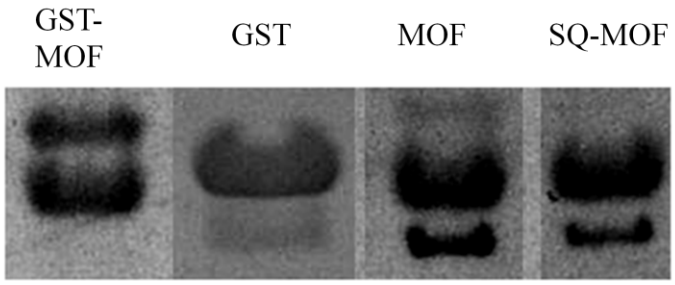


Figure 3.26: Recombinant $dMOF_{371-827}$ specifically acetylates histones. (A). Histone acetyltransferase assays were carried out with HeLa core histones and various purified recombinant $dMOF_{371-827}$ proteins. GST-MOF indicates the concentrated protein analyte from the 1st GS affinity column (6 mg/mL). GST demarks the concentrated protein analyte from the 2nd GS affinity column (4 mg/mL). MOF is the concentrated protein analyte of the flow-through from the 2nd GS affinity column (4 mg/mL). SQ MOF corresponds to the concentrated protein analyte of anion exchange chromatography (2 mg/mL). CH demarks the presence of core histone extracted from HeLa cells. The enzyme activity assays were completed by Dr. Vikki Weake (Stowers Research Institute, University of Kansas). (B). Major bands of a SDS-PAGE of the recombinant proteins tested for acetylation activity. GST-MOF (6 mg/mL), GST (4 mg/mL), MOF (4 mg/mL), SQ-MOF (2 mg/mL).

The reduction of the overall protein amount permitted only protein-protein interaction studies, but not crystallization trials. In summary, the 4 litre large-scale protein prep produced 24 mg of purified soluble $dMOF_{371-827}$ protein.

Table 3.4: Summary of the analysis of the quaternary structure of $dMOF_{371-827}$.

Running Buffer	pH	Additives to the Running Buffer	Additional Info	Results
20 mM Tris, 150 mM NaCl 5 mM DTT	8	-----	-----	Elution at 10.5 mL indicating a tetramer
20 mM sodium citrate 100 mM NaCl 5 mM DTT	5.5	-----	-----	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	0.5 mM zinc chloride, and 0.3 mM Acetyl-CoA	-----	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	7	0.5 µM zinc chloride	-----	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	0.5 mM zinc chloride, and 0.3 mM Acetyl-CoA	Zinc chloride was present post PreScission protease	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	0.5% (w/v) CHAPS detergent	Incubated with 2% (w/v) CHAPS detergent prior to column	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	0.5% (w/v) CHAPS detergent	Incubated with 1.2 % (w/v) CHAPS detergent prior to column	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	0.5% (w/v) CHAPS detergent	Incubated with 0.5 % (w/v) Octyl β-D glucopyranoside detergent prior to column	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	0.5% (w/v) CHAPS detergent	Incubated with 0.5 % (w/v) N-Dodecyl-β-D maltoside detergent prior to column	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	1 M Urea	Incubated with 1 M Urea prior to column	Elution at 10.5 mL indicating a tetramer
20 mM Tris 150 mM NaCl 5 mM DTT	8	2 M Urea	Incubated with 2 M Urea prior to column	Elution at 10.5 mL indicating a tetramer

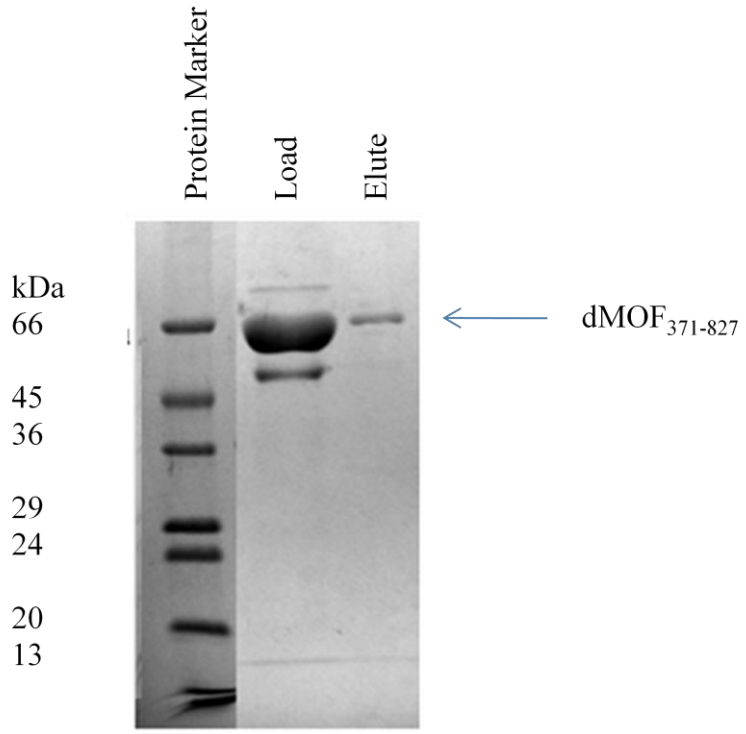


Figure 3.27: Representative purification of $dMOF_{371-827}$ by Superdex 200 size exclusion column. SDS-PAGE: protein marker, load (2 mg/mL), elute (V_e of 9.5 – 10.5 mL).

3.6 Protein interaction studies: $dMSL1_{820-1039}$ and $dMOF_{371-827}$

The initial GST pull-down assays to test the interaction between the recombinant proteins $dMSL1_{820-1039}$ and $dMOF_{371-827}$ revealed that the interaction was not above background levels as the amount of $dMOF_{371-827}$ in lanes 2 – 4 (GST-dMSL1 + $dMOF_{371-827}$ + GS beads) was the same as in the control, lane 7 ($dMOF_{371-827}$ + GS beads) (Fig. 3.28). In an attempt to reduce the background level of dMOF interacting with GS beads the non-ionic detergent TWEEN®20 was added; the results from the addition of detergent indicated a reduced level of background but also the complete removal of dMOF-dMSL1 interaction.

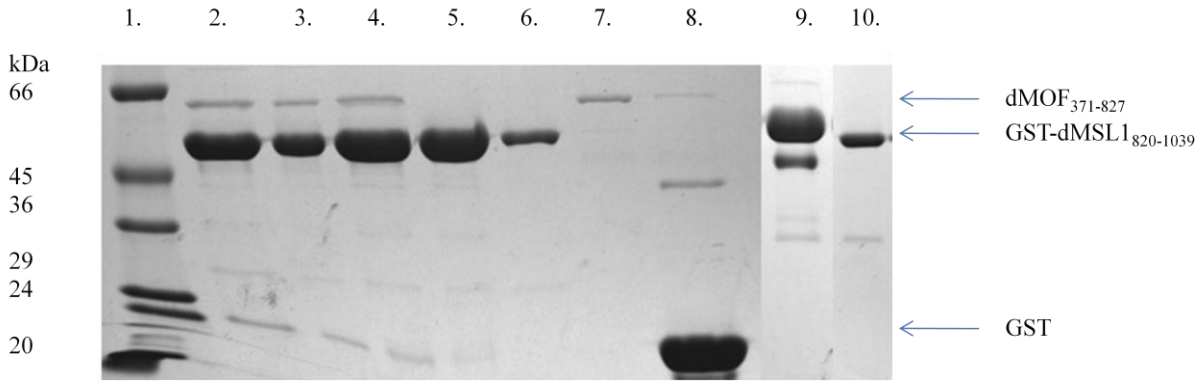


Figure 3.28: dMSL1$_{820-1039}$ and dMOF$_{371-827}$ GST pull-down assay. 12.5 % SDS-PAGE, Lane 1: protein markers, Lane 2: 60 µg GST-dMSL1$_{820-1039}$ and 60 µg dMOF$_{371-827}$, Lane 3: 60 µg GST-dMSL1$_{820-1039}$ and 60 µg dMOF$_{371-827}$, Lane 4: 60 µg GST-dMSL1$_{820-1039}$ and 60 µg dMOF$_{371-827}$, Lane 5: 60 µg GST-dMSL1$_{820-1039}$ incubated with GS beads, Lane 6: 8 µg GST-dMSL1$_{820-1039}$ incubated with GS beads, Lane 7: 10 µg dMOF$_{371-827}$ incubated with GS beads, Lane 8: 60 µg purified GST incubated with GS beads, Lane 9: 10 µg dMOF$_{371-827}$, Lane 10: 8 µg dMSL1$_{820-1039}$.

Additional interaction studies were performed using size exclusion chromatography; where the respective proteins were incubated together at unique stages in their purification, and with a variety of additives. Size exclusion chromatography permits studying protein-protein interactions by means of separating proteins based on their size in an inert environment. The first of the interaction studies combined purified dMSL1$_{820-1039}$ and purified dMOF$_{371-827}$ (Table 3.5). Zinc and acetyl-CoA were added to promote the secondary protein structure of dMOF$_{371-827}$.

Table 3.5: Experimental setup for the dMSL1$_{820-1039}$ and dMOF$_{371-827}$ protein-protein interaction study.

Proteins	Additives
Approx. 0.3 mg/mL dMSL1$_{820-1039}$ and Approx. 0.25 mg/mL dMOF$_{371-827}$	0.002 mg/mL Pepstatin A, 0.05 mg/mL 0.001 M PMSF, 0.5 mM DTT, and 0.5 µM $ZnCl_2$

Analysis of the size exclusion chromatography by SDS-PAGE revealed that the $dMSL1_{820-1039}$ and $dMOF_{371-827}$ co-eluted with an elution volume (V_e) of 13 mL. The calibrated size exclusion column revealed that a V_e of 13 mL equates to a molecular weight of approximately 100 kDa, indicating that the both $dMSL1_{820-1039}$ and $dMOF_{371-827}$ are in a monomeric state. The recombined molecular weight of two respective monomers of $dMSL1_{820-1039}$ and $dMOF_{371-827}$ would be 80 kDa. The discrepancy between the 100 kDa and 80 kDa may be explained by the impurities present. However, the overall protein levels were diminished greatly, to the extent that the results may not be significant.

Additional experiments directed towards establishing the $dMSL1_{820-1039}$ and $dMOF_{371-827}$ protein-protein interaction were performed by incubating the recombinant proteins: GST-$dMSL1_{820-1039}$ and GST-$dMOF_{371-827}$, where the GST-fusion tag remained on both proteins (Table 3.6). The GST fusion tag would aid in maintaining stability and provide additional secondary structure, in an effort to promote protein-protein interaction. The combined protein solution was incubated for 12 hours at 4 °C, after which PreScission protease was added to cleave the fusion tag, and the protein solution was incubated for an additional 16 hours. To identify the protein-protein interaction, the protein solution was first purified with GS affinity chromatography to remove the cleaved GST tag and then subsequent examination by means of size exclusion chromatography.

Table 3.6: Experimental setup for the GST-$dMSL1_{820-1039}$ and GST-$dMOF_{371-827}$ protein-protein interaction study

Proteins	Additives
Approx. 0.5 mg/mL GST-$dMSL1_{820-1039}$ and Approx. 0.5 mg/mL GST-$dMOF_{371-827}$	0.002 mg/mL Pepstatin A, 0.001 M PMSF, 5 mM DTT, and 0.5 μM $ZnCl_2$

The chromatogram of the size exclusion column indicated that $dMOF_{371-827}$ and $dMSL1_{820-1039}$ eluted as two distinct peaks (data not shown). The V_e of $dMOF_{371-827}$ ranged from 11.5 to 12.5 mL, identifying the MW as a 185 kDa to 155 kDa or a trimer quaternary structure. The V_e of $dMSL1_{820-1039}$ was 16 mL, which extrapolates to a MW of 30 kDa a monomer. Again, the interaction between MSL1 and MOF was observed to be very weak, where only a portion of $dMOF_{371-827}$ co-eluted with $dMSL1_{820-1039}$. In an effort to further aid in protein stability, the protein interaction between GST-$dMSL1_{820-1039}$ and GST-$dMOF_{371-827}$ was tried in the presence of 10% glycerol. The end result was a protein solution that failed to be properly cleaved by PreScission protease (data not shown).

Lastly, protein interaction between $dMOF_{371-827}$ and $dMSL1_{820-1039}$ was tested by means of anion exchange chromatography. Anion exchange chromatography was tried because the resin may mimic the overall positive charge of histone proteins. The positive charge of histones extends from the high percentage of basic amino acids. The chromatogram of the Source Q column combined with SDS-PAGE verified that both $dMOF_{371-827}$ and $dMSL1_{820-1039}$ recombinant proteins were loaded onto the column; however, only $dMSL1_{820-1039}$ eluted from the column (data not shown). Therefore, the *in vitro* protein interaction studies between $dMOF_{371-827}$ and $dMSL1_{820-1039}$ by means of GST-pulldowns, size exclusion chromatography, and ion exchange chromatography failed to demonstrate a strong stable interaction.

3.7 Protein interaction studies between $dMSL3_{186-512}$ and $dMOF_{371-827}$

Protein interaction studies between $dMSL3_{186-512}$ and $dMOF_{371-827}$ did not result in a stably bound complex. Both recombinant proteins were determined to be 0.5 mg/mL (Fig. 3.29B). Protease inhibitors (Leupeptin, PMSF, and Pepstatin A), reducing agent DTT, and zinc chloride were added to the incubation mixture. The size exclusion chromatogram indicates two distinct peaks. However, due to the proteins becoming too dilute SDS-PAGE can not verify the results (Fig. 3.29). Furthermore, it seems that the complex quaternary structure of $dMOF_{371-827}$ remained.

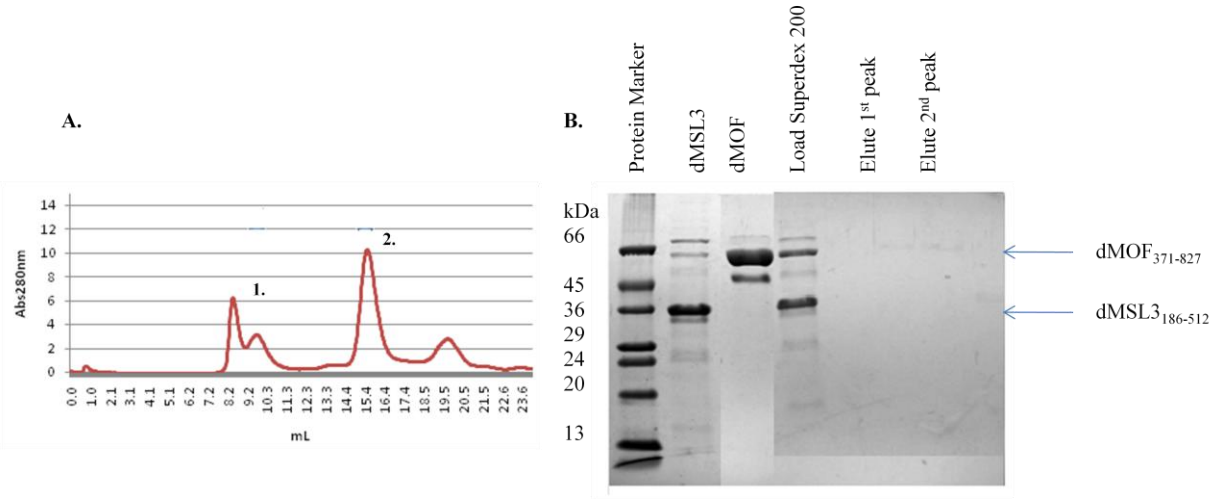


Figure 3.29: $dMOF_{371-827}$ and $dMSL3_{186-512}$ protein interaction study. (A). Chromatogram of size exclusion chromatography (Superdex 200) of $dMSL3_{186-512}$ and $dMOF_{371-827}$. Absorbance (280 nm) is displayed in maroon. The void volume (V_o) was 7.84 mL, with the total column volume (V_t) being 24 mL. (B). SDS-PAGE of $dMOF_{371-827}$ and $dMSL3_{186-512}$ protein interaction studies. Protein marker, $dMSL3_{186-512}$, $dMOF_{371-827}$, load Superdex 200 ($dMSL3_{186-512}$ and $dMOF_{371-827}$ protein solution combined), elute 1st peak (V_e 9.0 – 9.5 mL), elute 2nd peak (V_e 15.0 – 15.5 mL).

3.8 Protein interaction study: $dMSL1_{820-1039}$, $dMSL3_{186-512}$, and $dMOF_{371-827}$

In an effort to see if the three recombinant proteins when present together would form a stable complex all three recombinant proteins: $dMSL1_{820-1039}$, $dMSL3_{186-512}$, and $dMOF_{371-827}$ were individually purified and then incubated. The experimental design consisted of incubating purified $dMSL3_{186-512}$, GST-$dMSL1_{820-1039}$, and GST-$dMOF_{371-827}$. The volumes of the individual proteins were adjusted to attempt a 1 to 1 to 1 ratio. All three proteins had a final protein concentration of 0.25 mg/mL. Protease inhibitors (PMSF and Pepstatin A), reducing agent DTT, and zinc chloride were added to the incubation mixture. The interaction studies between $dMSL1_{820-1039}$, $dMSL3_{186-512}$, and $dMOF_{371-827}$ resulted in a protein solution loaded onto the size exclusion column which was too unstable and too dilute (Fig. 3.30). Therefore, the presence of all three proteins did not lead to a more stable protein complex.

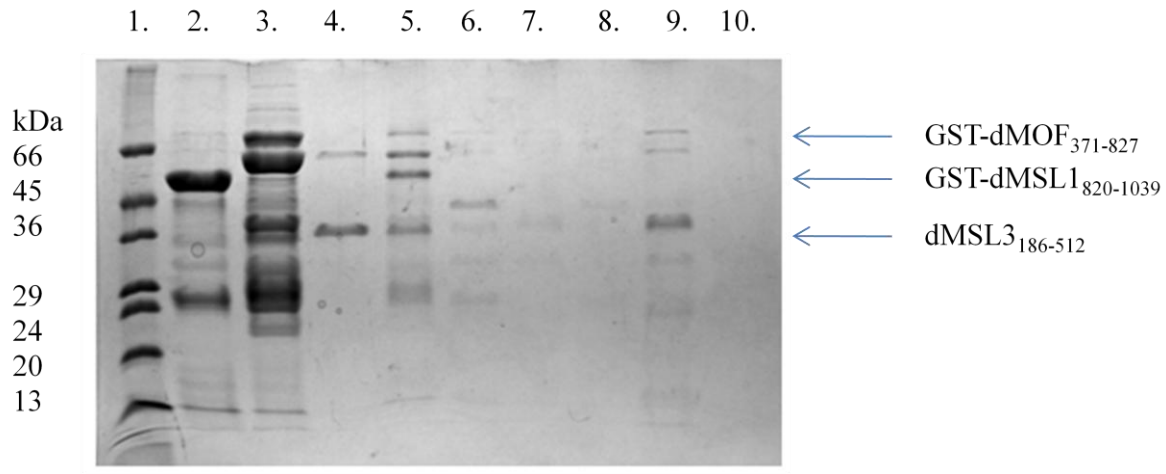


Figure 3.30: $dMSL1_{820-1039}$, $dMSL3_{186-512}$, and $dMOF_{371-827}$ protein interaction study. SDS-PAGE, Lane 1: protein markers, Lane 2: GST-$dMSL1_{820-1039}$, Lane 3: GST-$dMOF_{371-827}$, Lane 4: $dMSL3_{186-512}$, Lane 5: protein solution of all three recombinant proteins (pre-cleavage), Lane 6: load 2nd GS column (post PreScission protease cleavage), Lane 7: elute 2nd GS affinity column, Lane 8: elute GST only, Lane 9: load Superdex 200 10/300.

4. Discussion

Co-immunoprecipitation experiments on *Drosophila* larval cell extracts revealed that MSL1, MSL2, and MSL3 all interact with one another (Copps *et al.*, 1998). Yeast two-hybrid experiments demonstrated that the interaction between MSL1 and MSL3 was strong. From co-immunoprecipitation experiments in the presence of RNase from *Drosophila* cell extracts it was demonstrated that MSL1, MSL2, MSL3, and MOF remained as a stable complex, which suggested that the interactions were due to protein-protein interactions and were not RNA mediated (Akhtar *et al.*, 2000). It was further shown that FLAG-tagged MSL1 could co-purify MSL2, MSL3, and MOF from adult *Drosophila* cell extracts (Scott *et al.*, 2000). Moreover, a C-terminal construct of MSL1 comprised of amino acid residues 705 – 1039 ($MSL1_{705-1039}$) was able to specifically pull-down both MSL3 and MOF from *Drosophila* cell extracts (Scott *et al*., 2000). To confirm the protein interactions between $MSL1_{705-1039}$ plus HA-tagged MSL3 and $MSL1_{705-1039}$ plus HA-MOF recombinant proteins were made with *in vitro* translation (expression of protein in a cell free system); it was found that $MSL1_{705-1039}$ co-purified with HA-MSL3 but not with HA-MOF (Smith *et al*., 2000). This indicated that the C-terminal domain of MSL1 directly interacted with MSL3 but the interaction with MOF requires an additional factor (post-translational modification). MOF was determined to strongly interact with MSL1 when in the presence of MSL3 (Morales *et al*., 2004; Smith *et al*., 2000). It was also found that the *in vitro* translated MSL1, residues 973 – 1039, interacted with HA-MOF from *Drosophila* cell extracts (Morales *et al*., 2004). Furthermore, it was determined that the complete MRG domain, residues 185 – 490, of MSL3 was required for interaction with MSL1 as shown with *Drosophila* larvae expressing FLAG-tagged MSL3 constructs co-purified with MSL1 (Morales *et al.*, 2005). In order to conclusively establish the interactions between MSL1, MSL3, and MOF, and confirm the precise protein domains required for protein-protein interaction, all three proteins were over-expressed in an *E. coli* bacteria system. Conversion from studies in insect cells or *Drosophila* to bacteria enables over-expression of a select protein to levels required for structure determination by X-ray crystallography and also rules out the possibility of protein post-translational modifications contributing to the interactions being studied.

4.1 Cloning and recombinant protein expression

To date, the only reported domain within the C-terminus of *Drosophila* MSL1 is the PEHE domain, consisting of amino acids 865 – 1004 (Marin *et al.,* 2003). However, BLAST searches do not reveal any annotated protein domains within the C-terminal region. A *Drosophila* MSL1 C-terminal construct, consisting of amino acid residues 705 – 1039, directly interacted with both MOF and MSL3, suggesting a functional C-terminal domain exists (Scott *et al*., 2000). Consequently, this project attempted to clone a complete C-terminal domain; a complete domain is defined as a protein domain which could properly fold, akin to its native conformation. Moreover, protein constructs were designed to remove linker sequences, amino acids that connect two functional protein domains, as they are typically highly susceptible to proteolytic cleavage. In an effort to do so, an analysis of amino acid sequence conservation between *Drosophila* MSL1 and higher eukaryotes was completed. Sequence conservation did not confidently identify a clearly delineated MSL1 C-terminal domain. To improve the odds of over-expressing a stable C-terminal domain, four unique *Drosophila* MSL1 constructs -- $dMSL1_{688\text{-}1039}$, $dMSL1_{712\text{-}1039}$, $dMSL1_{754\text{-}1039}$, and $dMSL1_{820\text{-}1039}$ -- were cloned and protein expression trials were conducted. In conjunction with these studies on the C-terminus of *Drosophila* MSL1, four unique *Homo sapiens* MSL1 C-terminal constructs were cloned: $hMSL1_{262\text{-}614}$, $hMSL1_{364\text{-}614}$, $hMSL1_{404\text{-}614}$, and $hMSL1_{465\text{-}614}$. Minimal results have been reported on the protein domains of *Homo sapiens* MSL1; therefore, this study was directed also towards characterizing the *Homo sapiens* MSL1 C-terminal domain. Furthermore, to confirm the interaction between dMSL1 and dMSL3 a *Drosophila* $MSL3_{186\text{-}512}$ construct was cloned containing the MRG domain. The MRG domain is the reported domain for protein interaction with MSL1 (Morales *et al*., 2005; Zhang *et al*., 2006a). The structure of the MRG domain from dMSL3 remains unknown; therefore, another objective of this study was to generate a stable and soluble protein permitting structural studies.

In order to clone a eukaryotic gene permitting recombinant protein expression in bacteria, MSL cDNAs were utilized. cDNA was generated from mature mRNA from a eukaryotic cell., in this case either *Drosophila melanogaster* or *Homo sapiens*, that was comprised solely of exons as the introns were previously spliced.

As outlined in sections 3.1, 3.3, and 3.4, cloning was completed for four unique dMSL1 gene fragments, four unique hMSL1 gene fragments, and one dMSL3 gene fragment. Cloning

of the desired gene fragments was completed within the multiple cloning site (MCS) of the pGEX-6P-3 vector. Notably, the pGEX-6P-3 vector enables antibiotic selection because it possesses an Ampicillin resistance gene. As such, in the presence of Ampicillin *E. coli* that is transformed with a vector is pressured to keep the plasmid in order to survive. Ligation of the double digested gene fragment into the MCS of the vector proved to be the most difficult stage in cloning, particularly with the hMSL1 constructs. Experimentation with different ratios of vector to insert eventually produced single colonies, but even then not many colonies were typically recovered. Consequently, this indicates that *E coli* XL-1 cells had difficultly up-taking or maintaining the hMSL1 chimeras. *E. coli* XL-1 cells were used as they have been genetically engineered for cloning and plasmid purification, by improving their competency. Moreover, XL-1 cells are deficient of recombination (recA), and the endonuclease (endA), thereby improving insert stability and improving the quality of plasmid preparation, respectfully. Confirmation of cloning was determined by three means: firstly, by streaking single colonies produced from the transformation of ligated vector into XL-1 cells, onto LB agar plates possessing Ampicillin; secondly, using colony PCR, in which a single colony is used as a template as opposed to a purified DNA template; and thirdly, by DNA sequencing which aids in assessing the quality of cloning as well the accuracy of recombinant protein expression. It is important to note, that while colony PCR uses similar principles to that of PCR, the initial denaturing step causes the cells to burst exposing the plasmid encoding the insert.

To aid in expression of recombinant proteins all of the constructs were individually cloned into a pGEX-6P-3 vector. The pGEX-6P-3 vector enables expression of recombinant protein with an N-terminal GST tag. Through double digestion of both the vector and insert complementary restriction ends were generated and correct alignment of the gene fragment into the MCS was completed; therefore, the in-frame GST fusion protein was expressed. Over-expression of a GST-fusion protein can be achieved through the pGEX-6P-3 vector's chemically inducible *Ptac* promoter. By means of IPTG, chemical induction results in expression of T7 RNA polymerase, release of the *lac operon repressor* from the *Ptac* promoter enabling expression of the recombinant promoter. The strength of *Ptac* promoter directs the cell's attention towards expression of the recombinant protein slowing cell division.

Additionally, the *lac I* gene within the pGEX-6P-3 vector creates basal expression of the recombinant GST fusion protein continuously.

As opposed to yeast, insect, mammalian or other bacterial cell lines, *E. coli* was used as it allows the recombinant protein to be over-expressed to levels of 50 % of the total cellular protein. Unfortunately, however, recombinant eukaryotic protein expression in *E. coli* often results in insoluble protein and can be toxic to the bacteria. Moreover, *E. coli* often have difficulty translating and properly folding nascent protein, and incorporating post-translational modifications. To facilitate increased expression levels and protein assembly in *E. coli*, unique growth temperatures, ODs, type of media, and IPTG concentrations were tried in combination. Varying induction temperatures, ranging from 16 °C to 37 °C, were used to further retard cell division thereby enhancing the ability of bacteria to express recombinant protein. Notably, most eukaryotic protein post-translational modifications do not occur in bacteria.

To decipher which C-terminal dMSL1, hMSL1, or dMSL3 construct encoded a stable and soluble protein small-scale GST pull-downs were conducted. GST pull-downs are a fast and simple technique. As outlined in section 3.2, the soluble fraction is separated from the insoluble fraction by centrifugation; the GST of the GST-fusion binds to its substrate, the tri-peptide glutathione immobilized onto agarose beads, and captures the fusion protein. These glutathione-agarose beads are washed to remove any contaminating bacterial proteins and the desired GST-fusion protein are visualized with SDS-PAGE. Complete solubility of a protein often cannot be achieved since bacteria have problems with expressing eukaryotic proteins. Therefore, the goal is to maximize the amount of soluble protein. Notably, protein expressions of all four dMSL1 constructs were tried in both BL21 and Rosetta 2 cells (data not shown). BL21 cells are *E. coli* cells genetically designed for protein expression as they are deficient in both *Lon* and *ompT proteases* and thus able to aid in recombinant protein stability. Rosetta 2 cells, derived from BL21 cells, are designed to enhance eukaryotic protein expression by possessing seven rare tRNAs which are required for efficient translation of many eukaryotic genes.

As shown in section 3.1, the $dMSL1_{820\text{-}1039}$ construct transformed into BL21 cells resulted in a soluble GST-fusion protein. While $dMSL1_{688\text{-}1039}$, $dMSL1_{712\text{-}1039}$, and $dMSL1_{754\text{-}1039}$ were cloned, they were not successfully over-expressed as soluble protein (data not shown). Consequently, it can be concluded that a minimal C-terminal region consists of

residues 820 – 1039, which is largely comprised of the PEHE domain -- the region of highest conservation of amino acids between other eukaryotes --.

As determined by colony PCR (Fig. 3.13) and DNA sequencing (data not shown), the four *Homo sapiens* MSL1 constructs were all successfully cloned but the subsequent recombinant proteins failed to be expressed (Fig. 3.13). One possibility is that the pGEX-6P-3 vector used was altered during cloning and as the entire vector was not sequence the disruption was not detected. This possibility was strengthened from the lack of GST protein found in expression trials, typically over-expressed even when the recombinant protein was not. Another possibility is that the protein sequences did not accurately encompass a complete domain and thus did not encode stable, soluble proteins. While the difficulty of bacteria to express eukaryotic proteins cannot be ruled out, BL21 (DE3) cells and Rosetta 2 cells were utilized in protein expression.

Recombinant $dMSL3_{186-512}$ was successfully cloned as determined by colony PCR (Fig. 3.16) and DNA sequencing (data not shown). Over-expression of soluble protein to levels permitting further studies was observed by small-scale GST pull-downs (Fig. 3.17). The $dMSL3_{186-512}$ construct slightly differed from the two dMSL3 proteins, $MSL3_{141-512}$ and $MSL3_{181-512}$, found to interact with MSL1 (Morales *et al*., 2005); our lab altered the MRG domain, since residues eliminated were not well conserved with *Drosophila* MRG15 and MSL3 from other eukaryotes.

As outlined in section 3.5, $dMOF_{371-827}$ was cloned prior to the start of this project. Small-scale GST pulldowns were tried to assess both expression of a soluble protein and capture of the GST-fusion protein (Fig. 3.23). $dMOF_{371-827}$ was found to be expressed as a soluble protein to levels that allowed large scale preparations.

In an effort to study each individual protein and possible protein complexes, the recombinant proteins needed to be over-expressed, harvested, and purified to homogeneity. While consistent small-scale over-expression and harvesting of the recombinant proteins dMSL1, dMSL3, and dMOF allowed large-scale protein preparations and purifications to be examined, a future goal of this project is to verify each protein through crystallization as well as plausible protein interactions via co-crystallization. As such, purity and yield of the recombinant proteins were used to assess the quality of the purification protocol. Large-scale protein preparations and subsequent purification protocols for $dMSL1_{820-1039}$, $dMSL3_{186-512}$,

and $dMOF_{371\text{-}827}$ were tried with an ultimate goal of purifying each protein to a concentrated, homogeneous protein solution. The final stage of protein purification entails achieving 95 % homogeneity. This is reached when the desired protein is the only observable band identified with SDS-PAGE. The difficulty in purifying a protein of interest is separating it from the other components in the cell, particularly unwanted contaminating proteins, while simultaneously yielding large amounts of pure protein. Utilization of the pGEX-6P-3 vector allowed use of a GST tag, which aided in both solubility and stability of the recombinant protien. Moreover, the GST tag enables affinity chromatography to be used as the initial step in the purification of a recombinant GST-fusion protein thereby permitting a robust purification step to quickly and easiliy remove the majority of unwanted cellular proteins.

With the aim of maintaining protein structure and solubility temperature, pH, and ionic strength of a buffered solution must routinly be adjusted for each individual protein. Hydrogen bonds, hydrophobic interactions, and ionic interactions are important to protein stability and are also directly affected by the buffered protein solution. The unique properties required to purify a recombinant protein have to be carefully considered in an empirical manner as the protein sequence only provides limited estimates of the protein's isoelectic charge, molecular weight, and extinction coefficient. Protein purifaction is completed by exploiting the differences in properties between the protein of interest and the unwanted proteins within the mixture. Reducing the number of steps yeilds quicker preperation, lower protein losses, and lower costs.

Protein purification was performed using fast protein liquid chromatography (FPLC). FPLC is a form of column chromatography that enables protein purification using a variety of different stationary phases, including affinity, ionic, and hydrophobic chromatography. The flow-rate is commanded by an adjustable pump that drives the mobile phase containing the analyte through a select column. FPLC permits purification under constant pressure, whereby each protein can be purified from the column in a fast and efficient manner in an effort to decrease protein degradation. Purification under higher pressure also permits increased resolution when compared to gravity flow. A detector attached to the FPLC provides a way to characterize the retention time for the analyte and enables UV spectroscopic data on the analyte, allowing measurement of the yield, concentration, and purity of the recombinant protein.

To further prevent protein degradation, protease inhibitors were included in the buffered protein solutions. In doing so, protein integrity was maintained during the course of purification. Pepstatin A, PMSF, and Leupeptin were used to provide a broad inhibition of potential proteases. Pepstatin A inhibits aspartyl proteases along with a majority of other acid proteases, PMSF inhibits serine proteases and Leupeptin inhibits both serine and cysteine proteases.

To characterize the C-terminus of dMSL1 the recombinant $dMSL1_{820\text{-}1039}$ protein was successfully over-expressed and purified to homogeneity (Fig. 3.3 and 3.10). The conditions for over-expression were determined to be at an OD of 0.6, 0.1 mM IPTG, and an induction temperature of 16 °C. Analysis of the cell pellet weight was found consistently to be 6 g/L. Cell pellet weight is a quick assessment of the recombinant protein yield; however, its quality is limited as both the solubility and the final OD of the culture can greatly vary. The best assessment of quality is consistency among individual preparations and the efficacy of cellular growth to the desired OD.

Through utilization of two GS affinity columns, a progressive and robust purification of $dMSL1_{820\text{-}1039}$ was observed (Fig. 3.7 and 3.8). The pGEX-6P-3 vector provides a GST-fusion protein with a PreScission protease site between GST and the recombinant protein; this site allows for quick cleavage and removal of the unwanted GST fusion tag under very mild conditions. While the GST-fusion tag greatly facilitates purification, solubility, and stability, it can inadvertently interfere with crystallization; and was therefore removed by the second GS affinity column (refer to section 3.2). Notably, to prevent unwanted disulphide bonds between the two cysteine residues within $dMSL1_{820\text{-}1039}$ the reducing agent, DTT, was added to all buffered protein solutions after the initial GST affinity column.

As shown in section 3.2, further purification of $dMSL1_{820\text{-}1039}$ was attempted through both cation exchange (Source S) and anion exchange (Source Q) chromatography. Anion exchange (Source Q) did not aid in purification as the pI (isoelectric point) of the dMSL1 fragment was 5.5 and it consistently eluted in the flow-through (data not shown). Conversely, cation exchange (Source S) was successful in further purifying dMSL1. To maximize the purification by Source S, minimal levels of salt were used in the loading/equilibration buffer; specifically 25 mM NaCl and 25 mM Tris-HCl (pH 8.0) (Fig. 3.9). At pH 8.0 $dMSL1_{820\text{-}1039}$ is above its pI and should be negatively charged; the charge of the stationary phase in Source S at

pH 8.0 is also negative. Therefore, purification by Source S may be attributed to a positively charged region on the surface of dMSL1, as the pI is an estimate of the entire protein. The electrostatic potential on the surface of a protein can be different from other surface regions, and the interior of a protein, as proteins in solution are a globular structure and not uniform in shape and charge. Ion exchange permits fast and mild elution by a 25 mM – 1 M NaCl concentration gradient, as proteins elute based on the strength of their interaction with the stationary phase. Gel filtration (Superdex 200) enabled both assessment of the protein's quaternary structure as well as the purification of remaining contaminants based on size. Figure 3.10 illustrates that $dMSL1_{820\text{-}1039}$ is a monomer and that further purification was achieved as smaller protein bands were not observed by SDS-PAGE. Through the various stages of purification aggregates, minor contaminants, and degradation products were removed and the desired recombinant protein was transferred into a suitable buffer for structural studies.

Similar to dMSL1, recombinant $dMSL3_{186\text{-}512}$ protein was purified to homogeneity, establishing both over-expression and a purification protocol (Section 3.5). The conditions for over-expression were found to be at an OD of 0.8, 0.1 mM IPTG, and induction temperature of 16 °C. Compared to that of $dMSL1_{820\text{-}1039}$ the pelleted cell weight was consistently 5 g/L.

As with $dMSL1_{820\text{-}1039}$, $dMSL3_{186\text{-}512}$ was purified by the combination of two GS affinity columns, which resulted in the removal of the GST tag, capture of the fusion protein, and rapid removal of impurities (Fig. 3.17). Again DTT aided in stabilizing the recombinant protein by ensuring the buffered solution was kept in reduced state. Contrary to $dMSL1_{820\text{-}1039}$ purification, Source Q facilitated as the intermediate step in the purification of $dMSL3_{186\text{-}512}$. After screening different salt concentrations and buffer types (data not shown) purification was achieved under the following conditions: 25 mM Tris-HCl (pH 8), 25 mM NaCl, and 5 mM DTT (Fig. 3.18). To further purify the $dMSL3_{186\text{-}512}$ and determine its quaternary structure gel filtration (Superdex 200) was again performed and $dMSL3_{186\text{-}512}$ was determined to be a monomer (Fig. 3.19). Majority of the $dMSL3_{186\text{-}512}$ protein precipitated while concentrating the eluted protein solution. Further experiments need to be directed towards improving the stability of $dMSL3_{186\text{-}512}$ at high concentrations; it was determined that concentrating of $dMSL3_{186\text{-}512}$ to levels above 2 mg/mL resulted in protein precipitation (Fig. 3.19). Interestingly, the stability of $dMSL3_{186\text{-}512}$ estimated by the absence of degradation products

during purification and storage indicates that it is stable. To aid in further stability, different buffers with varying pH or the addition of 5 – 20 % glycerol may be tried.

As shown in section 3.6, recombinant $dMOF_{371-827}$ was successfully over-expressed and purified to homogeneity (Fig. 3.27). It was previously illustrated that full-length dMOF could be coupled with a GST tag and purified from *E. coli* (Akhtar *et al.*, 2000). This project further demonstrated that higher yields and purity could be achieved permitting dMOF structure determination. Over-expression was found to be improved with the use of Hyper Broth under the following conditions: 0.1 mM IPTG, OD 0.6, and an induction temperature of 16 °C. The pellet cell weight was determined to be 15 g/L; the difference in weight compared to dMSL1 and dMSL3 is likely attributed to the Hyper Broth. Protein inactivation was a key consideration during the purification of $dMOF_{371-827}$, as the directive of this study was to characterize how the complex of MSL1, MSL3, and MOF mediates MOF enzyme activity. Recombinant GST-fusion protein was again captured by using GS affinity chromatography, and after PreScission protease digestion a second GS affinity chromatography enabled a fast recovery of recombinant $dMOF_{371-827}$ (Fig. 3.23). Given that $dMOF_{371-827}$ contained 10 cysteine residues, the risk of disulphide bonds was high and DTT was used as an additive to all solutions. After screening Source Q ion exchange chromatography with various buffers and pH the following conditions were used: 25 mM Bis-tris propane (pH 7), 25 mM NaCl, and 5 mM DTT; this resulted in high resolution purification (Fig. 3.24). $dMOF_{371-827}$ eluted from Source Q was concentrated and polished by gel filtration (Superdex 200). Many difficulties were encountered upon gel filtration. Specifically, $dMOF_{371-827}$ did not elute efficiently from the column, as the fractions were too dilute to continue with crystallization or protein interaction studies. Furthermore, the eluted dilute $dMOF_{371-827}$ appeared to be a tetramer (Fig. 3.25 and Table 3.4). Often $dMOF_{371-827}$ eluted in the void volume, indicating that it is an aggregate with a weight above the maximum molecular weight that can be resolved by the column. High activity of $dMOF_{371-827}$ and GST tagged $dMOF_{371-827}$ indicates a correctly folded protein (Fig.3.26); however, the loss of activity of dMOF after ion exchange chromatography indicates a detriment to either overall protein structure or a possible removal of the zinc atom bound by the zinc finger. The possible removal of zinc atom in the CCHC zinc fingers is unlikely as dMOF was shown to possess acetylation activity in the presence of EDTA (Akhtar and Becker, 2000). Moreover, the inclusion of DTT in the purification of dMOF after the

second GS affinity column may result in the removal of the zinc atom of the zinc finger or an alteration of the zinc finger provoking the aggregation of $MOF_{371-827}$. If either occurs, an alteration of the protein structure may result. Another possibility was oxidation of the protein, which may cause aggregation as the half-life of DTT is short. PMSF a strong nucleophile may also be affecting the zinc finger through disruption of the interaction between the zinc atom and the sulphurs of the three cysteine residues. Understanding the activity assays of $dMOF_{371-827}$ remained problematic and due to time restrictions this problem was not fully explored. Interestingly, the Abs_{280} to Abs_{260} nm ratio, which measures the protein to nucleic acid quantity, was used and did not change throughout purification indicating that there was no detectable change in the protein structure. Furthermore, the limited elution of $dMOF_{371-827}$ during size exclusion chromatography and the analyte that does elute was measured to be a tetramer, supporting the idea that purification by anion exchange chromatography was problematic. Moreover, SDS-PAGE of soluble fraction of $dMOF_{371-827}$'s over-expression consistently revealed two protein bands at approximately 80 kDa and at 55 kDa respectively. Further purification of $dMOF_{371-827}$ by preScission protease and a GS affinity column SDS-PAGE indicated an accumulation of the protein band at 55 kDa; indicating that the protein band observed in the SDS-PAGE of the soluble fraction is the over-expressed $dMOF_{371-827}$ lacking the N-terminal GST tag (prior to cleavage of the tag). Interestingly the SDS-PAGE of ion exchange column revealed a protein band at approximately 45 kDa, which may be $dMOF_{371-827}$ without its chromodomain.

Ascertaining properly folded recombinant proteins can be done with enzyme activity assays, limited trypsin proteolysis, circular dichroism, and ^{1}H, ^{15}N HSQC 2D NMR. Given that $dMSL1_{820-1039}$ is not an enzyme; the secondary structure was determined by limited trypsin proteolysis and circular dichroism. Due to the possibility that the protein was susceptible to trypsin, a negative result was difficult to understand and thus the results of limited trypsin proteolysis were not shown. Similarly, the results of circular dichroism were also not shown as additional trials need to be done to ensure the results. As such, an assessment of whether MSL1 was folded correctly was inconclusive. Ultimately ^{1}H, ^{15}N HSQC 2D NMR should be utilized to confirm the state of folding of $dMSL1_{820-1039}$ (Dyson and Wright, 1998). However, due to time restrictions an assessment of $dMSL3_{186-512}$ and $dMOF_{371-827}$ was not attempted.

Activity assays on $dMOF_{371-827}$ were attempted, which provided insight that the recombinant protein was properly folded.

All three recombinant proteins dMSL1, dMSL3, and dMOF have been reported to interact with nucleic acid. dMOF has been shown to interact with RNA, DNA, and nucleosomes (Akhtar *et al*., 2000), dMSL3 was reported to interact with DNA and RNA and dMSL1 is known to interact with nucleosomes, suggesting it also associates with DNA and RNA (Morales *et al*., 2004). However, the C-terminal domains of MSL1, MSL3, and MOF not thought to be important for interacting with roX RNA or the chromosome. Therefore, in accordance to protein purification strategies all unwanted cellular particles should be removed; therefore, all purification protocols included the addition of DNase I and benzonase (RNase and DNase). During purification Far-UV spectra were also measured to observe the Abs_{280} to Abs_{260} ratio. Assessment of purification was tested by the Far-UV spectrum as a peak at 280 nm and a lack of a visible shoulder at 260 nm indicates homogeneity. It is important to note that through-out purification of the recombinant proteins the Abs_{280} to Abs_{260} ratio was attempted to be improved not the specific 280 nm and 260 nm measurements as the inconsistency between different protein batches.

The purification protocol for $dMSL1_{820-1039}$ resulted in a protein solution lacking appreciable amount of nucleic acid (Fig. 3.11). Evaluation of ammonium sulphate precipitation strongly indicated that contaminating the nucleic acids were largely removed (Table 3.1). Moreover, ethanol precipitation of the flow-through fractions of Source S cation exchange revealed material was purified away from $dMSL1_{820-1039}$ (Fig. 3.9). Interestingly, this material could not be visualized by ethidium bromide staining of native PAGE or agarose gels of the material (data not shown). It was also determined that $dMSL3_{186-512}$ was purified to eliminate contaminating nucleic acids (Fig. 3.20). The presence of nucleic acid as evidence by absorbance at 260 nm for $dMOF_{371-827}$ remains inconclusive due to the potential absorbance and interference by the CCHC zinc finger at 260 nm (Fig. 3.25). A summary of the 280 nm to 260 nm ratio from for the unique stages of $dMSL1_{820-1039}$'s protein purification protocol was provided in Table 3.2. Whereas shown in Table 3.1 is the 280 nm to 260 nm ratio from the initial ammonium sulphate precipitations.

4.2 Protein interaction studies

Quantification of interaction strength and propensity for the above proteins is difficult to ascertain from the literature as previous studies fail to include protein quantity within their reported data. Furthermore, *in vitro* interaction studies of recombinant protein expressed in bacterial systems are susceptible to problems due to mis-folded proteins and lack of post-translational modifications. This study used GST pull-down assays for testing direct interaction between a GST tagged bait protein and a prey protein, both of which have been purified to a desired homogeneity. The bait and prey protein can be interchangeable depending on which protein is the GST-tagged protein. Complexes recovered from GS beads are readily resolved by SDS-PAGE (Brymora *et al*., 2004). Size exclusion chromatography can be utilized to assess protein-protein interactions because it provides an inert environment which will not disrupt a pre-formed protein complex (Copps *et al*., 1998). Moreover, ion exchange chromatography can be employed to test protein-protein interactions.

Using GST pull-down assays (Fig 3.28) interaction between $dMSL1_{820-1039}$ and $dMOF_{371-827}$ did not exceed background levels. Furthermore, the adjustment of incubation periods, protein concentration, and the addition of non-ionic detergents did not improve interaction between $dMSL1_{820-1039}$ and $dMOF_{371-827}$. The problem with $dMOF_{371-827}$ found after ion exchange chromatography should not have been observed in the GST pull-down assays with GST-tagged $dMOF_{371-827}$. In addition, interaction studies by size exclusion chromatography and anion exchange chromatography established that the interaction between $dMSL1_{820-1039}$ and $dMOF_{371-827}$ is weak to non-existent as the proteins failed to elute together. These results are in accordance with previous studies with *in vitro* translated MSL1 (705 – 1039) and MOF (Smith *et al*., 2000). However, these results are contrary to studies with *in vitro* translated MSL1 (residues 973 – 1039) and HA-MOF from *Drosophila* cell extract (Morales *et al*., 2004). Interaction studies between $dMSL3_{186-512}$ and $dMOF_{371-827}$ showed no interaction, which coincides with the literature that MSL3 and MOF cannot directly interact (Fig. 3.29) (Morales *et al*., 2005; Morales *et al*., 2004). However, in *Homo sapiens* MSL3 and MOF were found to directly interact (Taipale *et al*., 2005). Contrary, to previous studies finding the interaction with MSL1 and MOF was strengthen in the presence of MSL3 the interaction studies with all three recombinant proteins $dMSL1_{820-1039}$, $dMSL3_{186-512}$, and $dMOF_{371-827}$ demonstrated no interaction (Fig. 3.30) (Morales *et al*., 2004; Smith *et al*., 2000).

In summary, all of the interaction studies proved to be inconclusive as no strong stable interactions were observed. The possibility that post-translation modifications are needed may be a problem, specifically MOF. Also, this study tested the interaction with a MSL1 construct ranging from 820 – 1039, previous studies tested the interaction with MSL3 and MOF with longer MSL1 constructs (705 – 1039 and 766 – 1039). Therefore, the removal of the amino acid residues 766 – 819 may be important for the protein interaction. Moreover, the difficulties may be due to the difficulties in $dMOF_{371-827}$ purification, inclusion of DTT, PMSF, and dMOF tetrameric quaternary.

4.3 Crystallization

As shown in section 3.2, crystallization trials were attempted on $dMSL1_{820-1039}$, with no crystals obtained. Commercial sparse matrix screens were used to scan conditions likely to form $dMSL1_{820-1039}$ crystals. If promising conditions were obtained, then conditions were explored in the vicinity of the promising hits. Various promising precipitants were observed, indicating the protein concentration, the precipitants, the salt concentration, and the type of buffer are having a positive effect on provoking $dMSL1_{820-1039}$ to crystallize. The stability of $dMSL1_{820-1039}$ remains to be problem for crystallization as after 48 hrs in storage significant degradation products are consistently observed (data not shown). Crystallization trials could not be performed with $dMSL3_{186-512}$ because its stability at concentrations above 2 mg/mL must be first improved. The crystal structure of the MRG domain of MRG15 from *Homo sapiens* has been solved, indicating that the MRG domain folds properly *in vitro*, and can be crystallized (Bowman *et al.*, 2006). The structure of the chromodomain of *Drosophila* MOF was solved (Nielsen *et al.*, 2005). Also, the structure of the *Homo sapiens* MOF MYST HAT domain has been published, which supports the value of determining $dMOF_{371-827}$, because the construct also includes the chromodomain. Crystallization trials could not be performed with $dMOF_{371-827}$ due to the heterogeneity in quaternary structure and the difficulty in purification by size exclusion chromatography.

5. Future Directions

The research presented summarizes cloning, expression, and purification to homogeneity of a *Drosophila* MSL1 C-terminal construct and a *Drosophila* MSL3 C-terminal construct consisting of the MRG domain. Moreover, the study has shown expression and purification of an enzymatically active *Drosophila* MOF construct comprised of the chromodomain, MYST HAT domain, and the CCHC zinc finger. Initial protein-protein interaction studies were attempted between the recombinant MSL1, MSL3, and MOF *Drosophila* proteins. Lastly, cloning of the *Homo sapiens* MSL1 C-terminus was described.

Additional experiments are required to improve the stability of *Drosophila* $MSL1_{820-1039}$, and to obtain proof of its secondary structure. In an effort to do so, $dMSL1_{820-1039}$ could be bound to either dMOF or dMSL3. In doing so, the complex would stabilize the individual proteins by both altering the conformation of the proteins and by protecting susceptible regions from proteases. Another practical approach could be to carefully remove or add 3 - 5 amino acids to the C-terminal construct. In this way, the folding (secondary structure) of the recombinant protein could be improved by the inclusion or removal of amino acids. Interestingly, the use of Tris buffer did not affect the interaction of recombinant $dMSL1_{820-1039}$ with the stationary phase, even though the recommended buffer for Source S is HEPES buffer. Both buffers optimally work at pH 8.0; however, Tris buffer is zwitterionic its positive charge may interact with the negative stationary phase, effectively inhibiting desired interaction because of a loss of binding sites.

Also, *Drosophila* $MSL3_{186-512}$ protein needs to be stabilized at a higher concentration in order for protein crystallization trials and interaction studies to be completed. Typically crystallization trials require a protein concentration between 8 – 20 mg/mL. The most direct method would be to try obtaining more soluble protein by varying the expression conditions. Also, the stability of protein can be greatly affected by the choice of buffer or the addition of additives such as 5 – 20 % glycerol. Furthermore, different growth media should be attempted in order to aid in the production of soluble protein.

Future studies may also benefit from addressing what exactly occurred to $dMOF_{371-827}$ because of, or during, anion exchange chromatography. Moreover, activity assays could be done in the presence and absence of zinc to quantify whether or not zinc effects activity at all stages of dMOF purification. Near-UV spectroscopy could be used to quantify the level of

zinc in dMOF protein during purification as zinc absorbs maximally at 500 nm (Futaki *et al.*, 2004). Size exclusion chromatography must be done more conclusively in order for crystallization trials. Size exclusion chromatography could also be done prior to ion exchange to determine whether the problem created during ion exchange is also problematic to size exclusion. $MOF_{371-827}$ possessing 13 cysteine residues requires the constant presence of a reducing agent; reduced L-glutathione could be substituted for DTT due to the possibility of problems arising because of DTT. Successful activity assays with recombinant MOF now permit mutation studies coupled with activity assays to provide insight into the mechanism of acetylation.

Successful expression and purification of recombinant dMSL1, dMSL3, and dMOF permits an assessment of the strength of their unique protein interactions. These unique protein interactions experiments would include: dMSL1 and dMSL3, dMSL1 and dMOF, dMSL3 and dMOF, and all three dMSL1, dMSL3, and dMOF. The dissociation constant (K_d) enables the measurement of affinity between two proteins. In order to measure the K_d of the unique protein interactions several techniques may be employed including: surface plasmon resonance, fluorescence spectroscopy (tryptophan fluorescence), isothermal titration calorimetry, and/or a Myc-tag permitting use of antibodies. Over-expression and purification of the recombinant proteins also permits future structure determination studies with either x-ray crystallography and or nuclear magnetic resonance. The possibility of co-crystallization trials requires further experiments directed towards establishing stable protein complexes.

Over-expression of recombinant $dMOF_{371-827}$ and establishment of its HAT activity on core histone proteins now grants further studies into its catalytic mechanism. The mechanism proposed for Esa1, MOF homologue and member of the MYST HAT family, can be tested (refer to section 1.4.2) (Yan *et al*., 2002). The two Esa1 mutations C304A and E338Q may be carried out MOF C680A and E714Q respectively. Additionally the affect on the previously reported MOF G691E mutation may be assessed (Akhtar and Becker, 2000).

6. Bibliography

Akhtar, A., and 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.

Akhtar, A., and Becker, P. B. (2001). The histone H4 acetyltransferase MOF uses a C2HC zinc finger for substrate recognition. EMBO Rep *2*, 113-118.

Akhtar, A., Zink, D., and Becker, P. B. (2000). Chromodomains are protein-RNA interaction modules. Nature *407*, 405-409.

Alekseyenko, A. A., Larschan, E., Lai, W. R., Park, P. J., and Kuroda, M. I. (2006). High-resolution ChIP-chip analysis reveals that the Drosophila MSL complex selectively identifies active genes on the male X chromosome. Genes Dev *20*, 848-857.

Alekseyenko, A. A., Peng, S., Larschan, E., Gorchakov, A. A., Lee, O. K., Kharchenko, P., McGrath, S. D., Wang, C. I., Mardis, E. R., Park, P. J., and Kuroda, M. I. (2008). A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. Cell *134*, 599-609.

Aratani, S., Kageyama, Y., Nakamura, A., Fujita, H., Fujii, R., Nishioka, K., and Nakajima, T. (2008). MLE activates transcription via the minimal transactivation domain in Drosophila. Int J Mol Med *21*, 469-476.

Bannister, A. J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641-643.

Bashaw, G. J., and Baker, B. S. (1995). The msl-2 dosage compensation gene of Drosophila encodes a putative DNA-binding protein whose expression is sex specifically regulated by Sex-lethal. Development *121*, 3245-3258.

Bassett, A., Cooper, S., Wu, C., and Travers, A. (2009). The folding and unfolding of eukaryotic chromatin. Curr Opin Genet Dev *19*, 159-165.

Bell, O., Conrad, T., Kind, J., Wirbelauer, C., Akhtar, A., and Schubeler, 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.

Belote, J. M., and Lucchesi, J. C. (1980). Male-specific lethal mutations of Drosophila melanogaster. Genetics *96*, 165-186.

Bertram, M. J., and Pereira-Smith, O. M. (2001). Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif. Gene *266*, 111-121.

Bowman, B. R., Moure, C. M., Kirtane, B. M., Welschhans, R. L., Tominaga, K., Pereira-Smith, O. M., and Quiocho, F. A. (2006). Multipurpose MRG domain involved in cell senescence and proliferation exhibits structural homology to a DNA-interacting domain. Structure *14*, 151-158.

Broadhurst, A. V. (2001). Hydroxylapatite chromatography. Curr Protoc Protein Sci *Chapter 8*, Unit8 6.

Brymora, A., Valova, V. A., and Robinson, P. J. (2004). Protein-protein interactions identified by pull-down experiments and mass spectrometry. Curr Protoc Cell Biol *Chapter 17*, Unit 17 15.

Buscaino, A., Kocher, T., Kind, J. H., Holz, H., Taipale, M., Wagner, K., Wilm, M., and Akhtar, A. (2003). MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. Mol Cell *11*, 1265-1277.

Buscaino, A., Legube, G., and Akhtar, A. (2006). X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. EMBO Rep *7*, 531-538.

Cai, Y., Jin, J., Swanson, S. K., Cole, M. D., Choi, S. H., Florens, L., Washburn, M. P., Conaway, J. W., and Conaway, R. C. (2009). Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. J Biol Chem.

Chayen, N. E. (1998). Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. Acta Crystallogr D Biol Crystallogr *54*, 8-15.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834-840.

Clapier, C. R., Chakravarthy, S., Petosa, C., Fernandez-Tornero, C., Luger, K., and Muller, C. W. (2008). Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. Proteins *71*, 1-7.

Cline, T. W. (1979). A male-specific lethal mutation in Drosophila melanogaster that transforms sex. Dev Biol *72*, 266-275.

Copps, K., Richman, R., Lyman, L. M., Chang, K. A., Rampersad-Ammons, J., and Kuroda, M. I. (1998). Complex formation by the Drosophila MSL proteins: role of the MSL2 RING finger in protein complex assembly. Embo J *17*, 5409-5417.

Corona, D. F., Clapier, C. R., Becker, P. B., and Tamkun, J. W. (2002). Modulation of ISWI function by site-specific histone acetylation. EMBO Rep *3*, 242-247.

Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol *319*, 1097-1113.

Dekker, F. J., and Haisma, H. J. (2009). Histone acetyl transferases as emerging drug targets. Drug Discov Today *14*, 942-948.

Demakova, O. V., Kotlikova, I. V., Gordadze, P. R., Alekseyenko, A. A., Kuroda, M. I., and Zhimulev, I. F. (2003). The MSL complex levels are critical for its correct targeting to the chromosomes in Drosophila melanogaster. Chromosoma *112*, 103-115.

Dou, Y., Milne, T. A., Tackett, A. J., Smith, E. R., Fukuda, A., Wysocka, J., Allis, C. D., Chait, B. T., Hess, J. L., and Roeder, R. G. (2005). Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. Cell *121*, 873-885.

Dyson, H. J., and Wright, P. E. (1998). Equilibrium NMR studies of unfolded and partially folded proteins. Nat Struct Biol *5 Suppl*, 499-503.

Ebert, A., Lein, S., Schotta, G., and Reuter, G. (2006). Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res *14*, 377-392.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev *18*, 2973-2983.

Fauth, T., Muller-Planitz, F., Konig, C., Straub, T., and 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.

Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev *17*, 1870-1881.

Fukunaga, A., Tanaka, A., and Oishi, K. (1975). Maleless, a recessive autosomal mutant of Drosophila melanogaster that specifically kills male zygotes. Genetics *81*, 135-141.

Futaki, S., Tatsuto, K., Shiraishi, Y., and Sugiura, Y. (2004). Total synthesis of artificial zinc-finger proteins: Problems and perspectives. Biopolymers *76*, 98-109.

Gelbart, M. E., and Kuroda, M. I. (2009). Drosophila dosage compensation: a complex voyage to the X chromosome. Development *136*, 1399-1410.

Gelbart, M. E., Larschan, E., Peng, S., Park, P. J., and 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.

Gilfillan, G. D., Straub, T., de Wit, E., Greil, F., Lamm, R., van Steensel, B., and Becker, P. B. (2006). Chromosome-wide gene-specific targeting of the Drosophila dosage compensation complex. Genes Dev *20*, 858-870.

Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem *182*, 319-326.

Girton, J. R., and Johansen, K. M. (2008). Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. Adv Genet *61*, 1-43.

Gorman, M., Franke, A., and Baker, B. S. (1995). Molecular characterization of the male-specific lethal-3 gene and investigations of the regulation of dosage compensation in Drosophila. Development *121*, 463-475.

Gorman, M., Kuroda, M. I., and Baker, B. S. (1993). Regulation of the sex-specific binding of the maleless dosage compensation protein to the male X chromosome in Drosophila. Cell *72*, 39-49.

Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Gu, W., Szauter, P., and Lucchesi, J. C. (1998). Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of Drosophila melanogaster. Dev Genet *22*, 56-64.

Gu, W., Wei, X., Pannuti, A., and Lucchesi, J. C. (2000). Targeting the chromatin-remodeling MSL complex of Drosophila to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. Embo J *19*, 5202-5211.

Gupta, A., Sharma, G. G., Young, C. S., Agarwal, M., Smith, E. R., Paull, T. T., Lucchesi, J. C., Khanna, K. K., Ludwig, T., and Pandita, T. K. (2005). Involvement of human MOF in ATM function. Mol Cell Biol *25*, 5292-5305.

Gupta, V., Parisi, M., Sturgill, D., Nuttall, R., Doctolero, M., Dudko, O. K., Malley, J. D., Eastman, P. S., and Oliver, B. (2006). Global analysis of X-chromosome dosage compensation. J Biol *5*, 3.

Hallacli, E., and Akhtar, A. (2009). X chromosomal regulation in flies: when less is more. Chromosome Res *17*, 603-619.

Hibbert, R. G., Mattiroli, F., and Sixma, T. K. (2009). Structural aspects of multi-domain RING/Ubox E3 ligases in DNA repair. DNA Repair (Amst) *8*, 525-535.

Hilfiker, A., Hilfiker-Kleiner, D., Pannuti, A., and 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.

Ivaldi, M. S., Karam, C. S., and Corces, V. G. (2007). Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in Drosophila. Genes Dev *21*, 2818-2831.

Izzo, A., Regnard, C., Morales, V., Kremmer, E., and Becker, P. B. (2008). Structure-function analysis of the RNA helicase maleless. Nucleic Acids Res *36*, 950-962.

Jacobs, S. A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Jin, Y., Wang, Y., Johansen, J., and Johansen, K. M. (2000). JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. J Cell Biol *149*, 1005-1010.

Kageyama, Y., Mengus, G., Gilfillan, G., Kennedy, H. G., Stuckenholz, C., Kelley, R. L., Becker, P. B., and Kuroda, M. I. (2001). Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site. Embo J *20*, 2236-2245.

Kalkhoven, E. (2004). CBP and p300: HATs for different occasions. Biochem Pharmacol *68*, 1145-1155.

Kelley, R. L., Lee, O. K., and Shim, Y. K. (2008). Transcription rate of noncoding roX1 RNA controls local spreading of the Drosophila MSL chromatin remodeling complex. Mech Dev *125*, 1009-1019.

Kelley, R. L., Meller, V. H., Gordadze, P. R., Roman, G., Davis, R. L., and Kuroda, M. I. (1999). Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell *98*, 513-522.

Kelley, R. L., Solovyeva, I., Lyman, L. M., Richman, R., Solovyev, V., and 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.

Kelley, R. L., Wang, J., Bell, L., and Kuroda, M. I. (1997). Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. Nature *387*, 195-199.

Kind, J., Vaquerizas, J. M., Gebhardt, P., Gentzel, M., Luscombe, N. M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in Drosophila. Cell *133*, 813-828.

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet *41*, 376-381.

Kugler, J. M., and Lasko, P. (2009). Localization, anchoring and translational control of oskar, gurken, bicoid and nanos mRNA during Drosophila oogenesis. Fly (Austin) *3*, 15-28.

Kumaran, R. I., Thakar, R., and Spector, D. L. (2008). Chromatin dynamics and gene positioning. Cell *132*, 929-934.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lane, M. C., O'Toole, P. W., and Moore, S. A. (2006). Molecular basis of the interaction between the flagellar export proteins FliI and FliH from Helicobacter pylori. J Biol Chem *281*, 508-517.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Larschan, E., Alekseyenko, A. A., Gortchakov, A. A., Peng, S., Li, B., Yang, P., Workman, J. L., Park, P. J., and Kuroda, M. I. (2007). MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell *28*, 121-133.

Lee, C. G., Chang, K. A., Kuroda, M. I., and Hurwitz, J. (1997). The NTPase/helicase activities of Drosophila maleless, an essential factor in dosage compensation. Embo J *16*, 2671-2681.

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

Legube, G., McWeeney, S. K., Lercher, M. J., and Akhtar, A. (2006). X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in Drosophila. Genes Dev *20*, 871-883.

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

Li, X., Wu, L., Corsa, C. A., Kunkel, S., and Dou, Y. (2009). Two mammalian MOF complexes regulate transcription activation by distinct mechanisms. Mol Cell *36*, 290-301.

Lucchesi, J. C., Kelly, W. G., and Panning, B. (2005). Chromatin remodeling in dosage compensation. Annu Rev Genet *39*, 615-651.

Lucchesi, J. C., and Manning, J. E. (1987). Gene dosage compensation in Drosophila melanogaster. Adv Genet *24*, 371-429.

Lucchesi, J. C., and Skripsky, T. (1981). The link between dosage compensation and sex differentiation in Drosophila melanogaster. Chromosoma *82*, 217-227.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger, K., and Richmond, T. J. (1998). The histone tails of the nucleosome. Curr Opin Genet Dev *8*, 140-146.

Lyman, L. M., Copps, K., Rastelli, L., Kelley, R. L., and Kuroda, M. I. (1997). Drosophila male-specific lethal-2 protein: structure/function analysis and dependence on MSL-1 for chromosome association. Genetics *147*, 1743-1753.

Mandrioli, M., and Borsatti, F. (2006). DNA methylation of fly genes and transposons. Cell Mol Life Sci *63*, 1933-1936.

Marin, I. (2003). Evolution of chromatin-remodeling complexes: comparative genomics reveals the ancient origin of "novel" compensasome genes. J Mol Evol *56*, 527-539.

Marmorstein, R., and Roth, S. Y. (2001). Histone acetyltransferases: function, structure, and catalysis. Curr Opin Genet Dev *11*, 155-161.

Marmorstein, R., and Trievel, R. C. (2009). Histone modifying enzymes: structures, mechanisms, and specificities. Biochim Biophys Acta *1789*, 58-68.

Megee, P. C., Morgan, B. A., and Smith, M. M. (1995). Histone H4 and the maintenance of genome integrity. Genes Dev *9*, 1716-1727.

Meller, V. H., and Rattner, B. P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. Embo J *21*, 1084-1091.

Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I., and 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.

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., *et al.* (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. Mol Cell *21*, 811-823.

Meyer, B. J. (2000). Sex in the wormcounting and compensating X-chromosome dose. Trends Genet *16*, 247-253.

Morales, V., Regnard, C., Izzo, A., Vetter, I., and Becker, P. B. (2005). The MRG domain mediates the functional integration of MSL3 into the dosage compensation complex. Mol Cell Biol *25*, 5947-5954.

Morales, V., Straub, T., Neumann, M. F., Mengus, G., Akhtar, A., and Becker, P. B. (2004). Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. Embo J *23*, 2258-2268.

Mukherjee, A. S., and Beermann, W. (1965). Synthesis of ribonucleic acid by the X-chromosomes of Drosophila melanogaster and the problem of dosage compensation. Nature *207*, 785-786.

Narlikar, G. J., Fan, H. Y., and Kingston, R. E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. Cell *108*, 475-487.

Nguyen, D. K., and Disteche, C. M. (2006). Dosage compensation of the active X chromosome in mammals. Nat Genet *38*, 47-53.

Nielsen, P. R., Nietlispach, D., Buscaino, A., Warner, R. J., Akhtar, A., Murzin, A. G., Murzina, N. V., and Laue, E. D. (2005). Structure of the chromo barrel domain from the MOF acetyltransferase. J Biol Chem *280*, 32326-32331.

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

Oh, H., Park, Y., and Kuroda, M. I. (2003). Local spreading of MSL complexes from roX genes on the Drosophila X chromosome. Genes Dev *17*, 1334-1339.

Pace, C. N. (1990). Measuring and increasing protein stability. Trends Biotechnol *8*, 93-98.

Palmer, M. J., Mergner, V. A., Richman, R., Manning, J. E., Kuroda, M. I., and Lucchesi, J. C. (1993). The male-specific lethal-one (msl-1) gene of Drosophila melanogaster encodes a novel protein that associates with the X chromosome in males. Genetics *134*, 545-557.

Park, S. W., Kang, Y., Sypula, J. G., Choi, J., Oh, H., and Park, Y. (2007). An evolutionarily conserved domain of roX2 RNA is sufficient for induction of H4-Lys16 acetylation on the Drosophila X chromosome. Genetics *177*, 1429-1437.

Park, S. W., Kuroda, M. I., and Park, Y. (2008). Regulation of histone H4 Lys16 acetylation by predicted alternative secondary structures in roX noncoding RNAs. Mol Cell Biol *28*, 4952-4962.

Park, Y., Kelley, R. L., Oh, H., Kuroda, M. I., and Meller, V. H. (2002). Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins. Science *298*, 1620-1623.

Park, Y., Mengus, G., Bai, X., Kageyama, Y., Meller, V. H., Becker, P. B., and Kuroda, M. I. (2003). Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex. Mol Cell *11*, 977-986.

Paro, R., and Hogness, D. S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc Natl Acad Sci U S A *88*, 263-267.

Prestel, M., Feller, C., Straub, T., Mitlohner, H., and Becker, P. B. (2010). The activation potential of MOF is constrained for dosage compensation. Mol Cell *38*, 815-826.

Rattner, B. P., and Meller, V. H. (2004). Drosophila male-specific lethal 2 protein controls sex-specific expression of the roX genes. Genetics *166*, 1825-1832.

Richter, L., Bone, J. R., and Kuroda, M. I. (1996). RNA-dependent association of the Drosophila maleless protein with the male X chromosome. Genes Cells *1*, 325-336.

Schagger, H. (2006). Tricine-SDS-PAGE. Nat Protoc *1*, 16-22.

Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem *166*, 368-379.

Schroder, E., Jonsson, T., and Poole, L. (2003). Hydroxyapatite chromatography: altering the phosphate-dependent elution profile of protein as a function of pH. Anal Biochem *313*, 176-178.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990-995.

Scott, M. J., Pan, L. L., Cleland, S. B., Knox, A. L., and Heinrich, J. (2000). MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in Drosophila. Embo J *19*, 144-155.

Semeshin, V. F., Demakov, S. A., Shloma, V. V., Vatolina, T. Y., Gorchakov, A. A., and Zhimulev, I. F. (2008). Interbands behave as decompacted autonomous units in Drosophila melanogaster polytene chromosomes. Genetica *132*, 267-279.

Smith, B. C., and Denu, J. M. (2009). Chemical mechanisms of histone lysine and arginine modifications. Biochim Biophys Acta *1789*, 45-57.

Smith, E. R., Cayrou, C., Huang, R., Lane, W. S., Cote, J., and Lucchesi, J. C. (2005). A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. Mol Cell Biol *25*, 9175-9188.

Smith, E. R., Pannuti, A., Gu, W., Steurnagel, A., Cook, R. G., Allis, C. D., and 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.

Sousa, R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. Acta Crystallogr D Biol Crystallogr *51*, 271-277.

Spierer, A., Seum, C., Delattre, M., and Spierer, P. (2005). Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. J Cell Sci *118*, 5047-5057.

Straub, T., and Becker, P. B. (2008). DNA sequence and the organization of chromosomal domains. Curr Opin Genet Dev *18*, 175-180.

Straub, T., Grimaud, C., Gilfillan, G. D., Mitterweger, A., and Becker, P. B. (2008). The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. PLoS Genet *4*, e1000302.

Stuckenholz, C., Meller, V. H., and Kuroda, M. I. (2003). Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster. Genetics *164*, 1003-1014.

Suganuma, T., and Workman, J. L. (2008). Crosstalk among Histone Modifications. Cell *135*, 604-607.

Sun, Y., Jiang, X., Xu, Y., Ayrapetov, M. K., Moreau, L. A., Whetstine, J. R., and Price, B. D. (2009). Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. Nat Cell Biol *11*, 1376-1382.

Sural, T. H., Peng, S., Li, B., Workman, J. L., Park, P. J., and 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.

Taipale, M., and Akhtar, A. (2005). Chromatin mechanisms in Drosophila dosage compensation. Prog Mol Subcell Biol *38*, 123-149.

Taipale, M., Rea, S., Richter, K., Vilar, A., Lichter, P., Imhof, A., and Akhtar, A. (2005). hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. Mol Cell Biol *25*, 6798-6810.

Thomas, T., and Voss, A. K. (2007). The diverse biological roles of MYST histone acetyltransferase family proteins. Cell Cycle *6*, 696-704.

Tsunaka, Y., Kajimura, N., Tate, S., and Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. Nucleic Acids Res *33*, 3424-3434.

Vermaak, D., and Malik, H. S. (2009). Multiple Roles for Heterochromatin Protein 1 Genes in Drosophila. Annu Rev Genet *43*, 467-492.

Voss, A. K., and Thomas, T. (2009). MYST family histone acetyltransferases take center stage in stem cells and development. Bioessays *31*, 1050-1061.

Waddell, W. J. (1956). A simple ultraviolet spectrophotometric method for the determination of protein. J Lab Clin Med *48*, 311-314.

Wang, Y., Zhang, W., Jin, Y., Johansen, J., and Johansen, K. M. (2001). The JIL-1 tandem kinase mediates histone H3 phosphorylation and is required for maintenance of chromatin structure in Drosophila. Cell *105*, 433-443.

Wutz, A., and Gribnau, J. (2007). X inactivation Xplained. Curr Opin Genet Dev *17*, 387-393.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Yan, Y., Harper, S., Speicher, D. W., and Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. Nat Struct Biol *9*, 862-869.

Zhang, P., Du, J., Sun, B., Dong, X., Xu, G., Zhou, J., Huang, Q., Liu, Q., Hao, Q., and Ding, J. (2006a). Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. Nucleic Acids Res *34*, 6621-6628.

Zhang, W., Deng, H., Bao, X., Lerach, S., Girton, J., Johansen, J., and Johansen, K. M. (2006b). The JIL-1 histone H3S10 kinase regulates dimethyl H3K9 modifications and heterochromatic spreading in Drosophila. Development *133*, 229-235.

Zhimulev, I. F., Belyaeva, E. S., Semeshin, V. F., Koryakov, D. E., Demakov, S. A., Demakova, O. V., Pokholkova, G. V., and Andreyeva, E. N. (2004). Polytene chromosomes: 70 years of genetic research. Int Rev Cytol *241*, 203-275.

Zhou, S., Yang, Y., Scott, M. J., Pannuti, A., Fehr, K. C., Eisen, A., Koonin, E. V., Fouts, D. L., Wrightsman, R., Manning, J. E., and *et al*. (1995). Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. Embo J *14*, 2884-2895.