APO-DYSTROPHIN: LOCALIZATION AND EXPRESSION OF Dp116, Dp70, and Dp40 DURING EMBRYOGENESIS

by

CHAVDAR IVANOV

A thesis submitted to
The University of Birmingham for
the degree of
MASTER'S in Philosophy

School of Biosciences
College of Life and Environmental Sciences
The University of Birmingham
September 2017

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

Abstract

Dystrophin and its apo-forms are structural proteins which connect intracellular actin scaffolding in a cell with intercellular space. They are the main factor in guarding and keeping the sarcolemma intact after significant stress. Actin connects to dystrophin and forms a multistage complex, the dystrophin-associated protein complex (DAPC), with laminin 2 in the extracellular space. Muscular dystrophy is a hereditary disorder caused by progressive skeletal weakness based on defects in the sarcolemma, leading to cell death and inability to recover after simple physical activity. Despite recent advances in understanding the causes of muscular dystrophy, investigators still do not have a clear understanding of how DAPC contributes to disease development. Using immunohistochemistry and western blotting, this study attempts to increase understanding on how apo-dystrophin is expressed at various stages of embryogenesis. The results find programmed expressions of different types of apo-dystrophin at different stages of embryonic development of skeletal and cardiac muscle in a variety of mutants. Further research is needed to discover how different levels of apo-dystrophin are expressed and the effect their positions have on the future development of the various forms of muscular dystrophy.

Acknowledgements

I wish to thank Professor Janet Smith for her invaluable assistance in guiding my research. I am grateful for the support of Trushar Patel, who supported me and encouraged me to continue with my project.

I would like to express my gratitude to Professor Christopher Bunce for his patience and understanding.

List of tables

List of Figures

CHAPTER 1

Introduction

1.1 The dystrophin gene

Dystrophin is a protein produced by the dystrophin gene. It is the human genome's largest gene, at 2.5 Mbp long and with 79 coding exons (Koening et al. 1987, Coffey et al. 1992). Due to its large size, this gene has a high mutational rate (Culligan et al. 2001). It is located at Xp21.1 on the human X-chromosome. Investigators have identified eight different promotors (Nishio 1994). These promotors generate various protein isoforms expressed in a variety of tissues. These eight promotors express four proteins, which have a molecular mass of 427 kDa. In addition to these four proteins, the dystrophin gene makes shorter forms of dystrophin which have smaller molecular masses: 260, 140, 116 and 71 kDa, respectively. The four full forms of dystrophin are known as brain (B), muscle (M), Purkinje (P) and Lymphoblastoid (L) (Neuman 2005) (Fig.1). They all have a common second exon and different first exons of Dp427. Muscle (M) is expressed in skeletal, smooth, and cardiac muscle (Schofield et al. 1993). This M promotor is activated after differentiation of myogenic cells (Neuman 2005). Sequence analysis of this promotor shows that it consists of several conserved motifs, including E box, CarG boxes, and MCAT boxes (Neuman 2005) (Figure 1.

Dp427 Brain (B) is expressed in the brain, primarily in neurons located in the hippocampus and cortex (Gorecki et al. 1992). Figure 1 illustrates the location of Dp427 (B) between L and M promotors, the start of which is found approximately 500 bp upstream from the common exon 2 in the myogenetic cell (Neuman U. 2005). All four promotors have

different 5' Untranslated Regions (5'UTR), which are accompanied by different exonic splicing enhancers or silencers. All of the promotors are responsible for synthesizing different types of proteins based on different alternative splicing. As Neuman points out, the 5' Untranslated Region (5'UTR) of brain dystrophin mRNA is highly conserved in humans, mice, and chickens (Neuman 2005). B promotor has nine triplets upstream from the initiation codon AUG which contain AUG, and which are followed by the UAG codon which allows for different lengths of protein to be produced (Neuman 2005).

Purkinje (P) protein is another form of 427 kDa protein expressed in the Purkinje cells located in the cerebellum. Localization of the P promotor is between M and the second common exon in the dystrophin gene (Fig.1) (Neuman 2005). mRNA regulated by this promoter is found in Purkinjie neurons of the brain. This promotor regulates the expression of alternatively spliced mRNA, which has an AUG codon which can be found in close proximity to a 27 bp UAG codon (Neuman 2005) (Figure 1). The AUG codon's proximity to the 27bp UAG codon plays a significant role in alternative splicing. According to analysis by Neuman, another promotor, the lymphoblastoid cell dystrophin promotor, is positioned upstream from the B promotor. As mentioned previously, the first exon of the L promotor is spliced to exon 3, which differs from the rest of the promotors – M, B, and P. Their first unique exon is spliced to exon 2. As a consequence, the translation of L dystrophin mRNA starts at AUG at position 124. (Neuman 2005) (Figure 1).

The dystrophin gene has four internal promotors. They generate expression of the retinal (R) form Dp260, the brain (B3) form Dp140, the Schwan cell (S) form Dp116, and the general (G) form, also known as Dp71, as well as proteins transcribed from those four internal promotors. The common feature of all short forms of dystrophin is possession of different spectrin-like repeats with the largest number in Dp260 and the least in Dp71. The

proteins are as follows: the Dp260 retinal (R) form, located in the retina and consisting of 15 spectrin-like repeats, a cysteine-rich domain, and a C- terminal domain. According to Neuman, "It is likely that the first unique exon and the promotor are located at intron 29 of the DMD gene" (Neuman 2005). The Dp260 is founded in the outer plexiform layer but is also expressed in low quantities in brain and cardiac muscle (Figure 1). The Dp140 brain (B) form is expressed in brain glial cells and developing kidneys (Lidov et al. 1995). It has at least five spectrin-like repeats, a cysteine- rich domain, and a C-terminal domain. Translation of Dp140 starts at exon 51 where AUG is the first exon in frame, but the actual first exon of Dp140 has been sequenced, and it is in exon 44 where AUG has been found, but not in frame. This explains why exons 44-50 give a long 5'UTR mRNA for Dp140, which may be implicated in the alternative splicing (Neuman 2005) (Figure 1). The Dp116 Schwan cell (S) form is present in Schwan cells (Byers et al. 1993). The Dp116 form has at least two spectrin-like repeats, a cysteine-rich domain, and a C-terminal domain. The promotor is localized at intron 55, but the first in frame AUG triplet is in the exon between introns 55-56 (Neuman 2005) (Figure 1).

Shorter dystrophin isoforms are located in the peripheral and central nervous systems. The shortest form of all is Dp71, which is distinct from the full form of Dp427 and has cysteine- rich and C-terminal domains. This paper identifies two novel proteins that are shorter than Dp71: Dp40 and Dp30. Intron 62 has been identified as a promotor for Dp71 and acts as a housekeeping gene. In addition, it is rich in GC and has a TATA-less promotor, which leads to BRE being found upstream from its promotor (Neuman 2005).

Dp71 expression in early embryogenesis is down-regulated, with two transcription factors, Sp1 and Sp3, playing an essential role in the development of the early embryo and

preventing interference with synthesis of the full form of dystrophin Dp427 in muscle tissue in later development (Bermudez de Leon et al. 2004). Dp71 combined with Dp427 has been proven to play a central role in normal cardiac muscle development. Dp427 is found in the cardiac sarcolemma and T-tubules, whereas Dp71 is found predominantly at T-tubules.

Cardiomyopathy develops in cases where Dp427 is not expressed. The Mdx mouse and DMD-null mouse both develop cardiomyopathy because of lack of Dp427 (Masubuchi et al. 2013). Dp71 is present in the kidneys, lungs, the liver, and mainly in the blood-brain barrier. Proteins Dp427, Dp140 and Dp71 are all present in the brain, but the most abundant is Dp71. The number of expressed dystrophin proteins is increased by alternative splicing at the 3' end of the dystrophin gene (Feener et al. 1989)
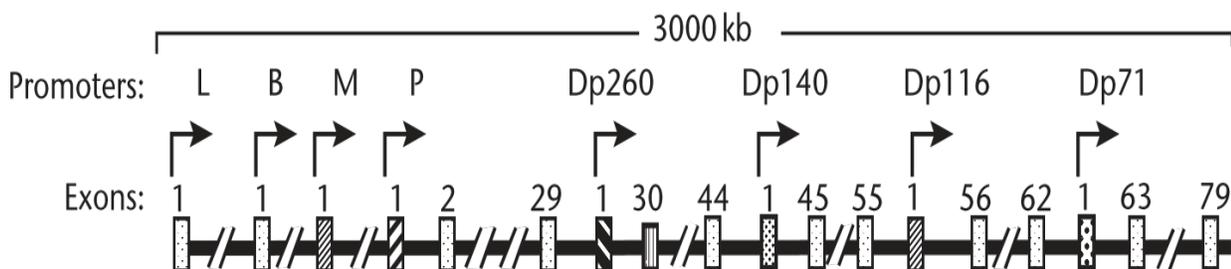
Figure 1. Scheme of dystrophin gene with promotors and exons. Introns where transcription starts for apo-dystrophin forms are shown. The transcription start site of each product is shown as a bent arrow. L, B, M and P show the position of lymphoblastoid cell, brain, muscle and Purkinje cell promotors, respectively. Each product has a unique first exon (Adapted from Neuman 2005).

Despite the large amount of information gathered in the last decade about dystrophin and its apo-forms, the role of dystrophin and its apo-forms in embryogenesis is not clear. For example, the function of the dystrophin-associated protein complex (DAPC) is still not clearly understood. It is necessary to uncover facts regarding dystrophin, its apo-forms in

early embryonic development, and how they interact and combine with the DAPC complex to understand more clearly the genesis of muscular dystrophy, which might eventually contribute to developing a novel treatment for the disease.

1.2   Dystrophin protein-structure and function

Full form Dp427 has four functional domains: the actin-binding domain which is at the N- terminal, the central rod domain, the cysteine rich, and the C-terminal domain. The actin binding N-terminal domain consists of 220 amino acids in the full form. The rod-like domain is located in the central zone and has four linking regions called hinges. It consists of 24 α- helical spectrin like repeats, each with 110 amino acids. It is believed that this part stabilizes the molecule and makes it more flexible during muscle contraction. Experiments have determined that one more actin-binding site exists at the beginning of this domain which includes units 11 to 17 (Rybakova et al. 1996, Amann et al. 1998, 1999, Koening, Kunkel et al. 1990). The next short subdomain consists of 40 amino acids and is called the WW domain, which is responsible for β-sheet protein- binding (Bork, Sudol 1994). The WW and part of the cysteine-rich domain provide space for interaction between dystrophin and β-dystroglycan (β- DG), which is a part of DAPC that spans the sarcolemma. The Cysteine-rich domain (CYS) is capable of binding Ca2+ cations.

The last C-terminal domain consists of a dystrobrevin (DB)-binding site and provides space for interaction between dystrophin and syntrophin (Roberts 2001). Dp260, Dp140 and Dp116 have different N-terminal domains; therefore, Dp260 and its rod domains are shorter than the rod domain in Dp427. Dp71 is different from Dp427, and its N- terminal domain has seven new amino acids. Additionally, the transcription of Dp71 mRNA is spliced on exons 71- 74 and 78, which increases the number of expressed proteins. Even if exons 71-74 are lost but exon 78 is saved, then the C-terminal domain is the same as other dystrophin isoforms. If exon 78 is spliced out, 31 new amino acids emerge in the C-terminal domain, 13 of which are expressed, resulting in a change in the reading frame (Lederfain 1992). Dp71 has different actin-binding and β-dystroglycan-binding sites, and the lack of a WW domain changes interaction with other proteins (Howard 1998). Figure 2 depicts schematic characteristics of Dp427, Dp260, Dp140, Dp116, and Dp70.
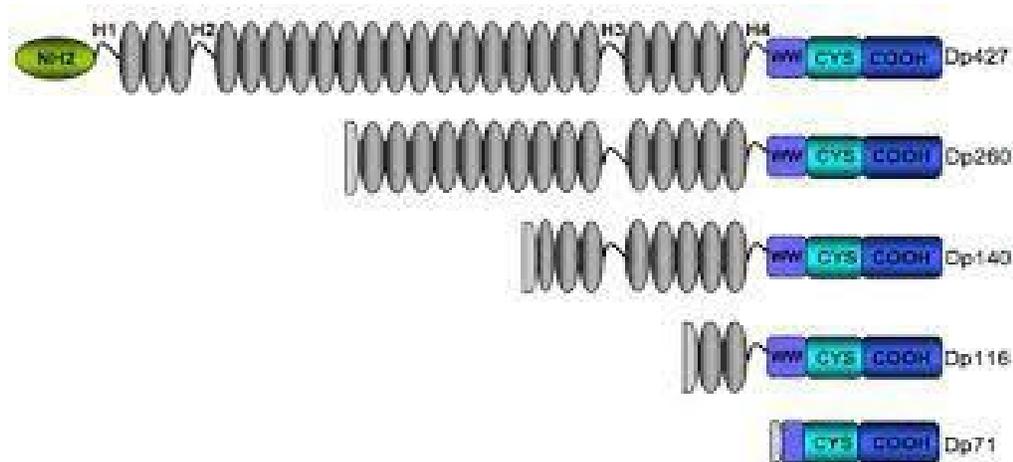
Figure 2. Structure of the full form of dystrophin 427 kDa and its apo-forms. These include the NH2 region, the H1, H2, H3 and H4 (regions which are flexible hinge regions and are rich in proline); WW (located in the cysteine-rich domain); cysteine-rich and C-terminal domains (based on Blake et al. 2002).

1.3 Development of dystrophin protein in embryogenesis

Skeletal muscle development in embryogenesis should first be highlighted in order better to understand the role of dystrophin during this process. The somite in early embryogenesis is influenced by the Hox genes. These genes are activated during somite formation, and it appears that the embryo perceives the somite and expression of the Hox genes as one complex. The appropriate Hox gene is activated in the somite where this gene is present. Hox genes are responsible for specialization of somite's in the anterior-posterior axis. Once established in each somite, the Hox gene is retained in this somite even if the somite has been moved to another area of the embryo. "Somite's give rise to the cartilage of the vertebrae and ribs, in addition to the muscles of the rib cage, limbs, the back of the abdominal wall, and the dermis of the dorsal skin" (Ordahl and Le Douarin 1992).

The flexibility and the pluripotency of the somite can be seen in the fact that when it is separated from the pre-somatic mesoderm, its cells can become any of the somite-derived structures. A portion of the cells located in the ventral-medial region of the somite can undergo mitosis, which results in loss of the ability to convert to epithelia. As a result, they are transformed into mesenchymal cells. The sclerotome is a structure emerges from the somite cells which have evolved from mesenchymal cells. Consequently, these cells becoming chondrocytes. Fate mapping with chimeric cells has helped to determine the fate of the rest of the epithelial part of the somite and is divided into three regions--two lateral, and one medial (Ordahl and Le Douarin 1992). The myotome is the layer produced out of the lower part of the previously formed structure called the dermatomyotome, which is comprised of double layers of cells from the somite. The dermatomyotome develops from the two lateral regions of the somite and is the precursor to myoblasts, which in turn develop into muscle cells. Two types of muscle formation exist--epaxial muscle groups (the deep muscle of the back), which develop from the myoblasts located close to the neural tube, and hypaxial muscle groups, which develop from the myoblasts farthest from the neural tube and which result in formation of muscles of the body, limbs, and tongue (Venters et al. 1999). As mention above, one of the double layers of the dermatomyotome--more precisely, the dorsal and central part--is named the dermatome and gives rise to the skin of the back, or dermis. Understanding the formation of the myotome is crucially important for understanding the stages of development of the future myoblast.

Studies involving transplantation and knockout mice have resulted in identification of several genes which produce proteins that influence the development of the myotome. These proteins are Wnt1, Wnt3, Sonic hedgehog, Bone morphogenetic protein 4 (BMP4) as well Pax1, Pax3 and Pax7. As mentioned above, the medial portion of the somite, which

results in formation of the epaxial muscle cells, is under the influence of the Wnt1 and Wnt3a, which the dorsal region of the neural tube releases. The epaxial muscle cells are influenced by Sonic hedgehog, as well, which the floor plate and notochord produces (Munsterberg et al.1995; Stern et al.1995; Ikeya and Takada 1998).

BMP4 facilitates the migration of the cells as well (Cossu et al. 1996; Pourquie et al. 1996; Dietrich et al. 1998). All of the factors which influence the development of the dermatomyotome are responsible for activating the specific muscle-gene expression through production of transcription factors that target these genes and, in some cases, are the reason for the movement of the myoblast from the dorsal region to a more appropriate ventral position , which slows down its own differentiation.

The transcription factor BMP4 is influenced negatively by Sonic hedgehog and Noggin. As mention above, BMP4 is produced from the later plate mesoderm (Cossuet al. 1996; Pourquie et al 1996; Dietrich et al. 1998). Negative signals from Sonic hedgehog and Noggin decrease the influence of BMP4. For example, Sonic hedgehog inhibits the influence of BMP4 only to the lateral sclerotome which hampers the propagation of this transcription factor to the medial and ventral part of the sclerotome. This prevents the transformation of the sclerotome cells into muscle cells from the last two regions. The interwoven influence of transcription factors mentioned above can be seen in Fig.3. The neural tube divides into a dorsal and a ventral section. The notochord develops from the floor plate of the dorsal section. The dermatomyotome and sclerotome are formed laterally from the neural tube. Transcription factors including Sonic hedgehog and Noggin, made from the notochord and the floor plate, produce contradictory influences. These transcription factors influence the ventral region of the somite to become the sclerotome, and the sclerotome in turn tarts to produce the Pax1 transcription factor. Production of Pax1 leads to differentiation of the

cells to become vertebrae and ribs. The dorsal part of the neural tube produces Wnt, BMP4, as well as a small amount of Sonic hedgehog transcription factor sufficient to block propagation of BMP4 to the rest of the neural tube. Surrounding cells from the dorsal part of the neural tube produce Pax3 transcription factor which marks the dermomyotome clearly for later development. The medial part of the dermomyotome is under the influence of Wnt and Pax3, while the ventral part of the dermomyotome is under the influence of Pax1 and Sonic hedgehog. This results in expression of the Myf5 transcription factor. All of these influences converge in the formation of the epaxial muscle groups (the deep muscles of the back) (Lassar et al.1989). The lateral part of the dermomyotome is under the influence of Wnt, which is secreted from the epidermis covering the dermomyotome, as well as Pax3 and BMP4 which is secreted from the lateral plate. All of these transcription factors converge and stimulate secretion of the MyoD transcription factor. This in turn leads to differentiation into hypaxial muscle groups (body wall and limb muscles) (Thayer et al.1989). MyoD and Myf5 are part of the group of transcription factors structured in a helix-loop-helix format called myogenic bHLH (Maroto et al. 1997; Tajbakhsh et al. 1997). These transcription factors are expressed in muscle precursors and muscle cells.

The next stage of muscle cell development is muscle cell fusion, which follows the exit of the myoblast from the cell cycle. The cell fusion process has three steps. First, several myoblasts fuse to form a single multinuclear myoblast (Mints and Baker 1967). This multinuclear myoblast divides until halted by the appearance of growth factors such as Fibroblast growth factor (FGF) (Menko and Boettiger 1987; Boettiger et al. 1995). Second, the myoblast initiates production of fibronectin into its extracellular matrix. And finally, $\alpha 5\beta 1$ integrin receptor binds to fibronectin. This fibrin-integrin receptor complex ensures the structural alterations that will guarantee irreversibility of the further differentiation of the myoblast into muscle cell (Menke and Boettiger 1987).
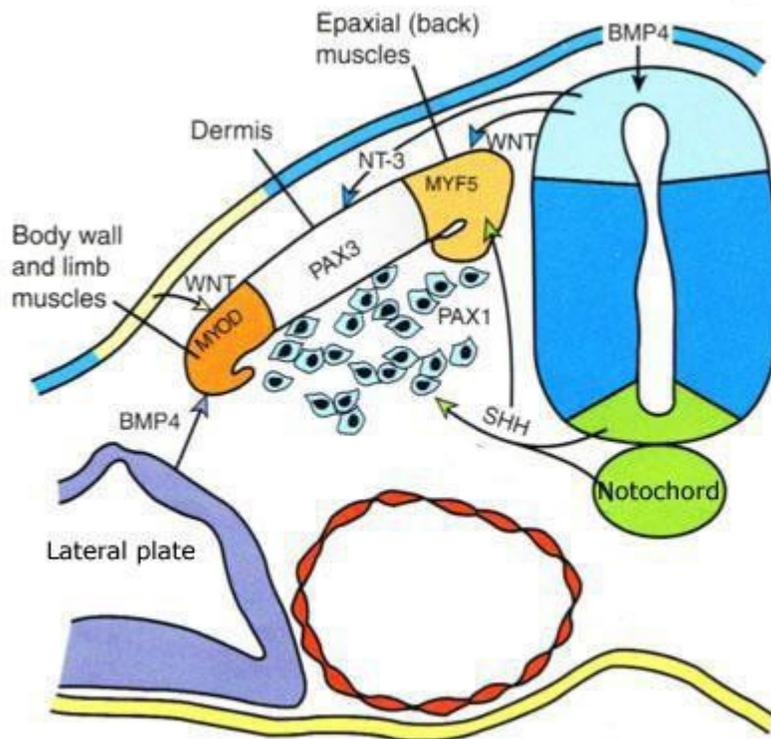
Figure 3. Schematic representation of transcription factors and their influence upon further development of embryonic skeletal and cardiac muscle. The dorsal part of the neural tube secretes BMP4, Wnt, Pax3 and NT-3. The ventral part of the neural tube, with the floor plate and notochord, secrete Sonic hedgehog and Pax1, which then influence the production of Myf5 from the medial part of the dermomyotome. This in turn leads to differentiation of the epaxial muscle groups. The lateral part of the dermomyotome, under the influence of Wnt, forms the dermis. The lateral part also produces BPM4, Pax1, and Pax3, which leads to differentiation of the hypaxial muscle groups. (reproduction; http://Healthappointments.com/chapter-11-muscular-system-essays/).

The second step is the alignment of the myoblast. For the alignment to succeed, identical cells must be close to each other. The alignment is implemented by the cell membrane glycoproteins (Knudsen 1985; Knudsen et al.1990). The next and final step is the fusion itself. This fusion is initiated by a group of metalloproteinases named meltrins along with calcium ionophores such as A23187 (Shainberg et al. 1969).

Dystrophin appears for the first time in the embryonic somite on day 9.5 (Merrick et al. 2009; Huang et al. 2000). Deficiency of dystrophin in skeletal muscle leads to the major clinical forms of Duchenne muscular dystrophy (DMD), characterized by the progressive dystrophy of muscle tissue found in patients and in several animal models (Brown and Lucy 1999). Ten percent of patients die from cardiac complications in which dystrophin is not only present in the cardiac muscle but also is absent from cardiac Purkinje fibers (Bies et al. 1992). DMD also contributes to mental retardation, which is present in 30% of DMD cases.

Researchers have documented that Dp71 is the first dystrophin isoform to appear in early embryogenesis. This appearance is supported by an increased concentration of transcription factors Sp1 and Sp3. A family of transcription factors exists, consisting of the following: Sp1, Sp2, Sp3, Sp4, BTEB1, and BTEB2. The Sp1 transcription factor is found very widely and binds to so-called GC-boxes (Kadonaga et al., 1987). It contains zinc fingers as well as glutamine and serine/threonine-rich amino regions. The rest of the transcription factors—Sp2, Sp3 and Sp4—are homologous to Sp1 and highly conserved. Two more related factors, BTEB1 and BTEB2, are included in this family. They bind to the GC-box (Imataka et al. 1992; Sogawa et al. 1993). It has also been shown that Sp1 serves as an activator, while the Sp3 transcription factor may serve both as an activator and a depressor. In this case, Sp3 is a co-activator for the Dp71 promoter. Later in embryogenesis, Dp71 expression, well defined in myoblasts, declines, with the differentiation of myoblasts into myotubes.

It has also been shown that the 224 bp 5' flanking region of the Dp71 promoter, containing several Sp-binding sites (SP-A to SP-D), is responsible for basal activity of the Dp71 promoter (Bermudez de Leon et al. 2004). These transcription factors are downregulated with muscle cell differentiation. They down-regulate synthesis of the Dp71

dystrophin isoform, which helps avoid interference with synthesis of Dp427 later in the process of differentiation of muscle cells.

The role of Dp71 in the development of cardiac muscle has been demonstrated, but what remains unclear is its role in cardiac muscle development during embryogenesis. More research is necessary to determine the role of Dp71 and Dp427 and to establish whether Dp71 appears first during embryogenesis and whether this protein disappears later in embryogenesis like it does in myotubes of early skeletal muscle development. A group of researchers has concluded that normal cardiac muscle would need both forms of dystrophin in order to develop and function normally, and that a lack of Dp427 could lead to cardiomyopathy in mice (Masubuchi et al. 2013).

1.4 Dystrophin-associated protein complex (DAPC)

The structure and localization of dystrophin protein and the associated protein complex are known primarily in muscle tissue (Campbell et al. 1989). This complex links extracellular space through the sarcolemma with the cytoskeletal structure of the cell. Three subcomplexes of DAPC are confirmed to exist: dystroglycans–α, β (α, β-DG). α-dystroglycan is an extracellular protein, which connect β-dystroglycan with the laminin 2 protein, and this subcomplex binds dystrophin and utrophin. The second subcomplex is the sarcoglycan complex, consisting of α, β, ϒ and δ subunits (SG). It is a membrane-spanning subcomplex and connects sarcospan proteins with β-dystroglycan. The third subcomplex is cytoplasmic and connects dystrophin and utrophin with dystrobrevin (DB) and the syntrophin molecule (SYN). Syntrophin is an adapter molecule through which the whole complex can interact with signal transduction proteins and ion channels (Hazai et al. 2014). This is a promising

area for further experiments (See Fig. 4).
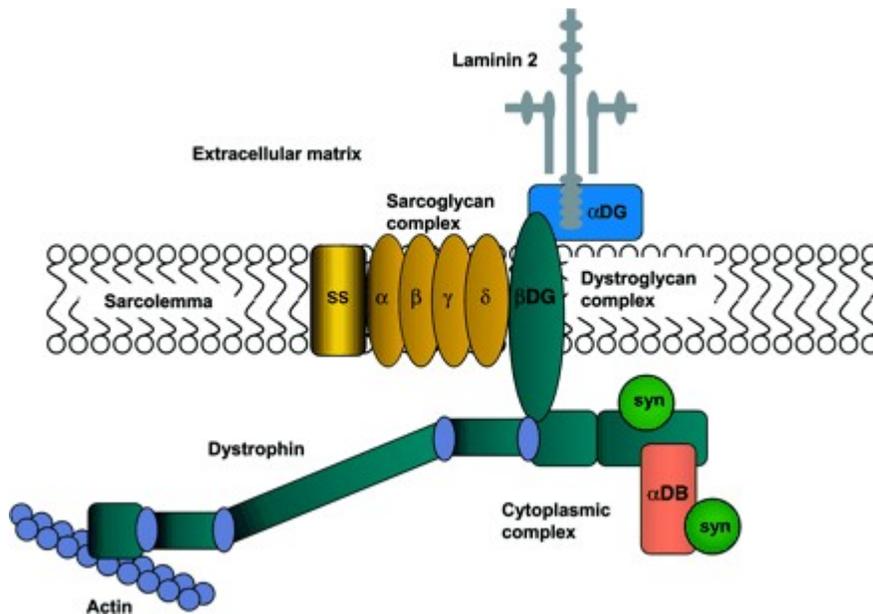
Figure 4. Structure of dystrophin-associated protein complex (DAPC). Dystroglycan sub complex-α, β- dystroglycan, laminin 2; Sarcoglycan sub complex α, β, ϒ and δ; and a third sub complex, which is intracellularly cytoplasmic and consists of dystrobrevin connected to dystrophin, utrophin and syntrophin. (Adapted from Blake and coworkers, 2002)

1.4.1 Dystrobrevin (DB)

Dystrobrevin directly binds to the C-terminal of dystrophin. There are two proteins-- α and β- dystrobrevin. The existence of two separate genes has been confirmed. The gene for α- dystrobrevein is located on chromosome 18 in humans and mice and has three promotors, which give rise to tissue-specific proteins. These three promotors can be further spliced, giving rise to more types of proteins. The three most important isoforms of α-DB are α-DB1, α-DB2, and α-DB3 (Holzfeind et al. 1999). These are present in cardiac and skeletal muscle

and have different C-termini (Blake et al. 1996), each of which is the location for several phosphorylation sites. There is evidence of the existence of two more isoforms with lower molecular mass, α-DB4 and α-DB5 which are present in skeletal muscle (Peters et al. 1998). α-BD1 is a cytoplasmic protein which binds dystrophin, utrophin, and syntrophin.

According to Yoshida et al. (2000), α-DBs can bind to sarcoglycan, but there is evidence that α-DB interacts directly with β-dystroglycan (Chang and Campanelli 1999). This isoform is most commonly found in cardiac and skeletal muscle. It can also be found in the neuromuscular junction and in pericapillary astrocytes and other glial cells (Blake et al. 1998, 1999), whereas α-DB2 is found in the sarcolemma (Nawrotzki et al. 1998). α-DB3 is found in skeletal and heart tissue, but the exact location is still unknown.

The gene of β-DB is located on chromosome 2 in humans and chromosome 12 in mice. As a consequence of splicing of the C-terminal, several isoforms of β-DB are expressed. There are no phosphorylation sites at the C-terminal of the β-DB molecule. These proteins are not expressed in the muscle but rather in several other tissues, including the brain. According to Blake (1996), β -DB is located in neurons of several regions of the cerebral cortex, hippocampus, and cerebellum (1999). β-DB interacts with syntrophins and dystrophin proteins (Dp71, Dp140, Dp427), and Blake assumes that it contributes to the formation of DAPC (see Figs. 4 and 6).

1.4.2 Dystroglycan (DG)

Dystroglycans are two proteins, α and β, that are strongly associated with one another. B- DG is a transmembrane protein that spans the sarcolemma, and α-DG is an extracellular protein. Both proteins are expressed by the dystroglycan gene, then post-transcriptionally glycosylated (Ibrahimova-Beskrovnaya et al. 1993). α-DG is associated with laminine-1 and

laminine-2 (Ervasti et al. 1993), and agrin (Campanelli et al. 1994). β-DG is a transmembrane protein which anchors α-DG and, with its C-terminal, interacts with the dystrophin protein (Jung et al. 1995). The C-terminal of β-DG interacts directly with the WW domain of dystrophin and binds with utrophin and α-DG (see Fig. 5).

1.4.3 Sarcoglycan (SG)

Sarcoglycans are found primarily in skeletal and cardiac muscle. There are α-, β-, ϒ- and δ-SG (Crosbie at al. 1997). α-SG is expressed in skeletal and cardiac muscle, whereas the rest are expressed in the smooth muscle (Barresi et al. 2000). An ε-SG has also been identified in several tissues, including brain tissue.

It is hypothesized that ε-SG can substitute for α-SG in the sarcoglycan-sarcospan complex formed in smooth muscle. Besides anchoring to the membrane and stabilizing DAPC, sarcolgycans act as an integral part in intracellular signal transduction (Yoshida et al. 1998). Mutations in its gene cause various type of muscular dystrophy (Busby 1999) (see Fig.5)

1.4.4 Sarcospan (SS)

Sarcospan (SS) is expressed primarily in skeletal and cardiac muscle. It has four transmembrane domains, an intercellular N-domain and a C-terminal domain. A sarcospan- deficient mouse does not exhibit muscular dystrophy (Lebakken et al. 2000) (see Fig.4).

1.4.5 Syntrophin (SYN)

The syntrophin protein family consists of five proteins: α-, β1-, β2-, ϒ1- and ϒ2-syntrophin (Adams et al.1993). Syntrophins directly bind to dystrophin, utrophin, α-DB1, α- DB2 and β- DG. It has been shown that DAPC contains four syntrophin binding sites (dystrophin and α-DB, each having two sites) (Newey et al. 2000). Syntophin also binds to signal transduction proteins and ion channels in DAPC in muscle and brain tissue, for example, voltage- gated Na+ channels, nNOS, and the aquaporin-4 ion channel (Inoue et al. 2002), as well as microtubule-associated serine/threonine kinase and mechanically activated protein kinase-3 (Hasegawa et al. 1999, Gee et al. 1998). Distribution of syntrophin is uneven, as α-SYN is present in the sarcolemma of muscles, β1-SYN is located in fast-contraction type 2 muscle fibers (Peters et al. 1997), and β2-SYN is located at the sub-neuronal site of the skeletal muscle's neuromuscular junction (Peters et al. 1994). In muscle, α-SYN is found in nNOS and aquaporin-4 channels in the sarcolemma. In CNS, α- and β-SYN are present in postsynaptic densities of synapses (Roberts 2001) where α-SYN, β-DB and Dp427 form a DAPC and anchor nNOS to the postsynaptic membrane (Blake et al. 1999). There is no indication of disease occurring when these proteins are absent (see Fig. 5).

1.4.6 Caveolin 3

Caveolin is the principal protein of the caveolae, is integrated into the plasma membrane, and has a molecular mass of 21-24 kDa. Several variants of caveolin exist, including caveolin 1 and caveolin 3. Caveolae are also an integral part of some signal transduction pathways. Research has proven that caveolin-3 is part of the signaling system coupled with G- protein (α- and β- subunits), protein kinase Cα, the Scr-family tyrosine

kinase, and Ras proteins. Caveolin 3 is expressed during the differentiation of the C2C12 cell culture (Song et al. 1996) and also binds β-DG to its WW region. Dystrophin and caveoline-3 have remarkably similar structure, and the C-terminus of β- DG is the binding site for these two proteins. In case of lack of the Dystrophin Dp427 in the organism, caveolin-3 can compensate for the lack of Dp427 to some extent. This overexpression is connected to a reduction of levels of dystrophin and β-dystroglycan (Song et al. 1996) (See Fig. 5). When there is no lack of Dp427 because of their structural similarity they compete for the binding place at the C-terminal of β-DG.
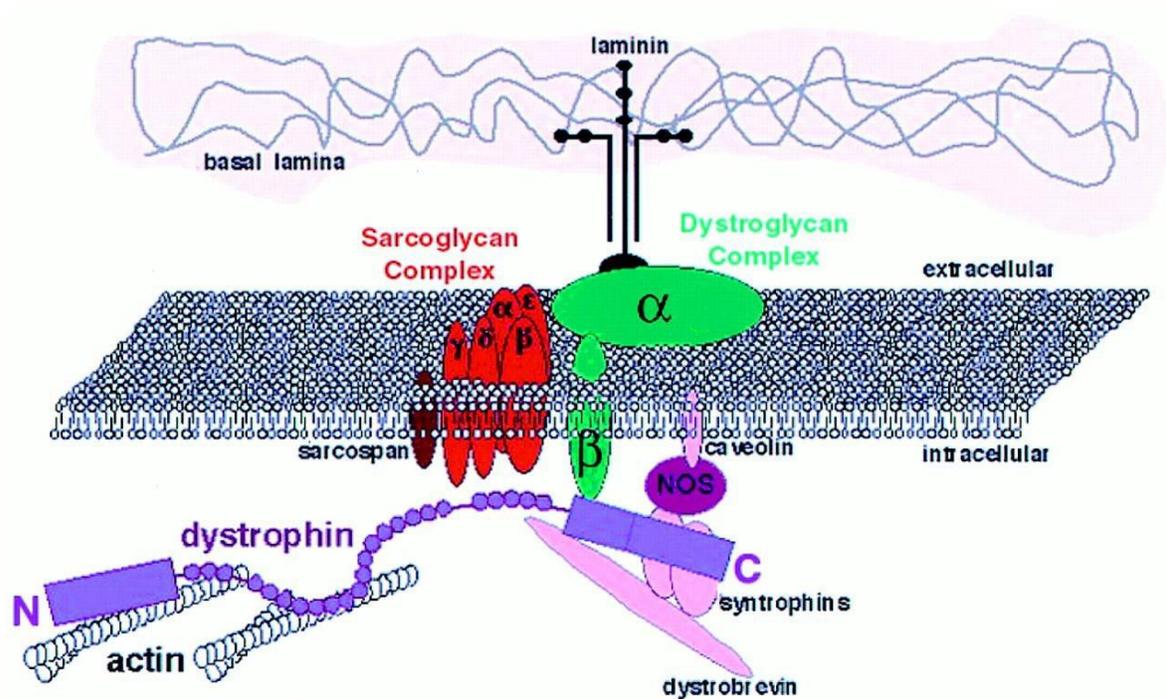
Figure 5. Schematic representation of the Caveolin-3 transmembrane protein. In the figure, Caveolin- 3 is shaded pink and is clearly represented as an integral protein and a part of the NOS membrane transduction process and, together with syntrophin, plays an important role as a bridge connecting intracellular space with extracellular space (adapted from Song et al.1996).

1.4.7 Utrophin

Utrophin is remarkably similar to Dystrophin protein. Some researchers theorize that the lack of Dp427 could be compensated for through genetic modifications to increase the production of Utrophin. The utrophin gene is found on chromosome 6 in humans and 10 in mice and has a molecular mass of 395 kDa. The primary structures of both dystrophin and utrophin are remarkably similar. Utrophin is produced from all autosomal chromosomes (Peng et al. 1998). Expression of utrophin can be expected to be mainly in skeletal, cardiac (Pons et al. 1996), and smooth muscle (Nguyen et al. 1991) but can also be found in vascular endothelia (Matsumura et al. 1992), retinal glial cells (Clauepierre et al. 1999), platelets (Earnest et al. 1995), peripheral nerves' Schwann cells (Matsumura et al. 1993), and a variety of types of kidney tissue (Lotz 1987).

Dystrophin is localized in the troughs of the neuromuscular junction, while utrophin is found in junctional fold crests. The distribution of utrophin is vastly different between adult muscle cell and in the differentiating muscle cell. The localization of utrophin in the adult muscle is on the neuromuscular junctions. In the differentiating muscle is found around the sarcolemma (Bewick et al 1992). There is a difference between dystrophin and utrophin actin- binding sites. Utrophin lacks the additional actin-binding sites associated with dystrophin's rod-like domain. Utrophin possesses a short extension of the N-terminus which dystrophin lacks. It is believed that this extension contributes to stronger affinity to actin, which binds cytoskeletal β-actin more strongly in vitro than does dystrophin (Moores et al. 2000). The primary structure of the C-terminus resembles that of dystrophin. It appears, therefore, that utrophin can bind β-DG (Matsumura et al. 1992), α-DB1 (Peterson et al. 1998) and syntrophin (Kramarcy 2000) and can also bind that part of the complex that includes sarcoglycans (Matsumura et al. 1992).
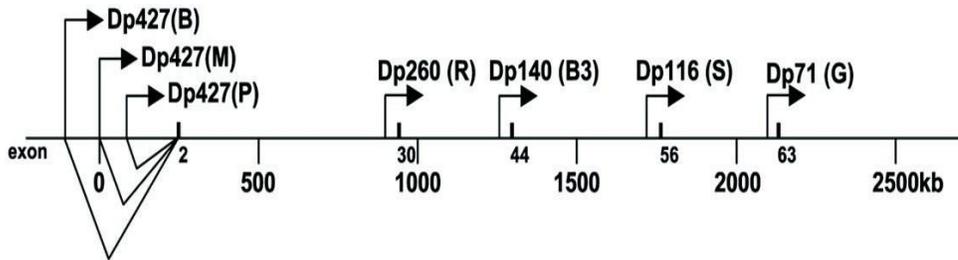
Figure 6. Schematic representation of the dystrophin gene, showing exons and introns, in addition to the molecules of dystrophin, utrophin, and all of the forms of dystrobrevin. Differences in the primary structure of all proteins are shown (Adapted from Matsumura et al.1992)

25

1.5 Aims

The main goal of this project is to localize apo-dystrophin in WT, Cav3$^{-/-}$ deficient, MDX mutant and DMHet mice, using immunohistochemistry in embryos E13.5, E15.5 and E17.5. The second goal is to determine how apo-dystrophin is distributed among different stages of development of the embryos and in which organs apo-dystrophin is most expressed during embryogenesis. The third goal is to determine the molecular mass and the concentration of expressed apo-dystrophin Dp116, Dp71 and Dp40 at different stages of development of Cav3$^{-/-}$ deficient, MDX mutant and DMHet mice and to compare the results to the expression of Dp116, Dp71 and Dp40 to find standard deviation of the mean.

CHAPTER 2

Materials and Methods

2 Immunohistochemistry (IHC)

2.1 Embedded embryos in Paraffin.

Embryos were dissected and placed in 4 % PFA solution on a Gyro-Rocker shaker at 4°C for 3-24 hours' time. The time of exposure was optimized by embryonic developmental stage as shown in the table below:

Step	Solution	E11.5	E12.5	E13.5	E14.5	E15.5	E16.5	E17.5
1	4% PFA (4°C)	3 hrs.	5 hrs.	20 hrs.	20 hrs.	20 hrs.	24 hrs.	24 hrs.**
2	40% C_2H_5OH	30 min	30 min	60 min	60 min	90 min	90 min	24 hrs.
3	70% C_2H_5OH	30 min	30 min	60 min	60 min	90 min	90 min	24 hrs.
4	95% C_2H_5OH	30 min	30 min	60 min	60 min	90 min	90 min	24 hrs.
5	100% C_2H_5OH	3x30 min*	3x30 min*	3x60 min*	3x60 min*	3x90 min*	3x90 min*	48 hours
6	Xylene	3x30 min	3x30 min	3x60 min	3x60 min	3x90 min	3x90 min	72 hours

Table 1. Fixation, dehydration and clearing incubation times for E11.5-E17.5 stage embryos in 4% paraformaldehyde, ethanol (C_2H_5OH) and xylene. *=final wash was performed overnight at 4°C; **=decalcification for E17.5 embryos (Merrick, 2006)

E11.5-E17.5 embryos were washed twice with sterile PBS for ten minutes at room temperature. The embryos were dehydrated in serial washes with ethanol (40%, 70%, 95% and 100%) for the time presented in the table above and corresponding to the stage of development. Embryos were placed into glass bottles, and the ethanol was substituted with

xylene (Fisher Scientific, UK). Embryo E17.5 underwent an additional step for decalcification prior to dehydration. Decalcification was performed with 0.5 M of EDTA solution (pH 7.5) for 72 hours at 4°C, and the solution was changed every day. Next, the embryo was washed out in dH$_2$O overnight. Embryos were then washed with xylene. Molten wax (60° C paramat; Fisher Scientific UK) was poured into glass bottle with embryos inside. The bottles were placed in an oven (Stuart Scientific, UK; Hybridization oven/shaker S120H). For E11.5-E13.5 embryos, wax was changed twice wax over a 24-hour period, while the wax for E14.5-E17.5 was changed three times during a 24-48-hour time period. Heated forceps were used to remove embryos from the wax in order to prevent the wax from solidifying too quickly, which otherwise would result in damaging the tissue. Orientation of embryos was obtained as required in 22 mm x 22 mm squire peel-away disposable embedding molds (Fisher Scientific, UK), and then the molds were labeled and immersed in molten wax. Blocks were left to cool for six to eight hours and then placed in a cold room at 4°C until ready for sectioning using a microtome.

2.2 Microtome sectioning of paraffin blocks.

A water bath was set at 30°C and an incubator at 37°C prior to starting the sectioning. Paraffin blocks were first removed from the cold room. The paraffin blocks were then clamped into a Reichert-Jung 1150/Auto cut microtome (Leica Microsystems Ltd, UK) and carefully aligned with the blade. Sections were cut at 5 μm thick and placed on top of the water in the 30°C bath. Sections were then collected onto super frost slides (VWR International, UK). Excess water between sections and slides was carefully dried. Slides were placed in a 37°C oven overnight to ensure the attachment between the section and the slide.

2.3 Visualization

I experimented with different dilutions of the monoclonal antibody MANEX 7374J 8FA. MANEX 7374J 8FA is a monoclonal antibody which binds to exon 73 and 74, both of which are close to the C-terminal domain. The second primary antibody was MANDRA1 77 7A10 (Table 3). The table below depicts the concentration changes of the antibody, which allowed a good visualization.

Table 2. Dilutions of monoclonal antibody to visualize IHC

Concentration of antibody	1:1000	1:750	1:500	1:250	1:100	1:75	1:50
Name of primary antibody	MANEX 7374J 8F10	MANEX 7374J 8F10	MANEX 7374J 8F10	MANEX 7374J 8F10	MANEX 7374J 8F10	MANEX 7374J 8F10	MANEX 7374J 8F10
Results	No visualization	No visualization	Little visualization	Little visualization	No clear visualization	Visualization	Clear visualization

Deparaffinization using Xylol was carried out twice, first for ten minutes, and the second time for five minutes. Rehydration was then carried out with different alcohol concentrations, starting with absolute alcohol 100% twice, for two minutes each time, followed by 95%, 90%, 70%, 50% and 30%, respectively, each for two minutes.

In order to block endogenous peroxidase, slides were then immersed in a solution consisting of 5 ml of 3% hydrogen peroxide mixed into 45 ml of water, for five minutes. The slides then were immersed for an hour, after which they were washed in PBST three times, ten minutes each time. Blocking non-specific sites with TNB for 30 minutes was performed

prior to applying the primary antibody on tissue specimens. In this case, two types of antibody, MANEX 7374J 8F10 and MANDRA1 7A10, were used and the slides left overnight at 4°C. The following day, PBST was used to wash the slides three times, ten minutes each time. After this step, biotinylated Anti-Mouse IgG whole antibodies were applied as secondary antibodies for an hour at room temperature at a concentration of 1:500, dissolved in TNB.

The slides were then washed again with PBST three times, ten minutes each time. The slides were incubated with Streptavidin-horseradish peroxidase diluted with TSA-Kit (SA-HRP; TSA-Kit) in a ratio of 1:500. Following this, the slides were moved into a DAB solution (3, 3'- diaminobenzidine; Dako, UK), which stained the protein of interest a brown color. The DAB solution consisted of 1ml of DAB (1 tablet of DAKO®DAB chromogen dissolved in 10 ml of PBS) added to 100 ml of H2O. Slides were incubated for 30 minutes, followed by washing with PBST again three times, ten minutes each time.

Next, a signal amplification was produced by Avidin- biotin complex (ABC) using an amplification reagent: TSA-Kit with a dilution of 1:50. Different slides were immersed in this solution for different lengths of time: nine, 12, and 15 minutes, followed by washing with PBST three times, ten minutes each.

The slides were then incubated with SA-HRP in TNB at a dilution of 1:500. A greater dilution, such as 1:250 or even 1:100, is also possible for 30 minutes, at room temperature. The slides were again washed with PBST three times, ten minutes each time.

The next step was immersing the slides in DAB for six minutes and washing them twice with tap water. The next step was to place the slides in Hematoxylin for one and one-half minutes. Slides were then washed in Scott's water for five minutes. Slides were then dipped in acid alcohol. The slides were then placed for a second time in a second container with Scott's water and washed for five minutes under running tap water.

The sections were then dehydrated with increasing concentrations of alcohol, starting with 30%, 50%, 70%, 90%, 95%, and 100%, twice at each concentration for two minutes. Finally, the sections were mounted with DEPEX (VWP International, UK) a mounting medium, and covered with a coverslip. Table 3 depicts the antibodies used.

Table 3. Antibodies used

Primary Antibody (Titer)	Secondary Antibody (Titer)
MANEX 7374J 8F10- Provided by The Wolfson Center for Inherited Neuromuscular Disease, UK (1:1000; 1:500; 1:250; 1:100; 1:75; 1:50) Primary Antibody (Titer).	Biotinylated anti-mouse IgG antibody; Amershman Pharmacia Biotech, UK (1:1000; 1:500; 1:100)
MANDRA1 77 7A10- Provided by The Wolfson Center for Inherited Neuromuscular Disease (1:1000)	Biotinylated anti-mouse IgG antibody. Amershman Pharmacia Biotech, UK (1:1000)

To obtain results from IHC, I used the following embedded embryos in paraffin: WT embryo E15.5 and E17.5; $Cav3^{-/-}$ deficient mice embryo E15.5 and E17.5; MDX E15.5 and E17.5; and β-dystroglycan E15.5. IHC was performed six times with each phenotype with 10 slides, using two primary antibodies: MANEX 7374J and MANDRA1 77 7A10.

2.4 Western Blotting

Western blotting was performed with embryos with the following stage of development and genotypes:

Wild type (C57BL/10; E13.5, E15.5 and E17.5)

E13.5 (C57BL/10, *mdx, cav3⁻/⁻* and *mdxcav3⁻/⁻* double mutant het--DMhet).

E15.5 (C57BL/10, *mdx, cav3⁻/⁻* and *mdxcav3⁻/⁻*double mutant het--DMhet).

E17.5 (C57BL/10, *mdx, cav3⁻/⁻* and *mdxcav3⁻/⁻*double mutant het--DMhet) (Larner, 2012)

Pregnant mice were culled, and the uterus removed. Embryos were dissected and ground using a mortar and pestle. RIPA buffer was added to immerse embryos fully (3 ml) and then homogenized. The samples were aliquoted into 5 ml bijoux tubes.

The molecular mass of Dp116 is 116kDA. Dp70 has a molecular mass of approximately 70-80 kDa, while Dp40 has a molecular mass of 40kDa. Below is a table representing the percentile of polyacrylamide and the range of protein separation.

Percentile (%) polyacrylamide	Range of protein separation (kDa)
5	57-212
7	36-94
10	16-68
12	13-50
15	12-43

Table 4. The percentile of polyacrylamide used in SDS-PAGE gels to separate different sized proteins (Ameshmant Pharmacia Biotech UK Ltd Hybond™ ECL™ Nitrocellulose membrane reference manual)

To determine the protein yield, samples were diluted and subsequently assayed against a set of standards for BSA. BSA standards were diluted with sterile ddH2O in concentrations of 0, 1, 2, 5, 10, 15 and 25 µg/ml. 100 µl of each standard was plated into a 96-well plate (Nunc plastic ware, UK) in duplicate. The samples were diluted in sterile water in 1:100 and 1:400 ratios, and 100 µl of each sample plated into a 96-well plate alongside the standards. An equal volume of Coomassie® Protein assay reagent (Pierce, UK) was added into each well and mixed well with the protein sample. The plate was allowed to incubate for 20 minutes before reading at λ=595nm on an Exam precision microplate reader (Molecular Devices, UK) which revealed the standard curve for estimating sample concentration. Table 4 indicates the protein concentrations used:

Table 5. Protein concentrations used

Protein concentration	BSA	ddH2O	Coomassie protein assay reagent	Total
0	0	100	100	200
1	0.5	99.5	100	200
2	1.0	99.0	100	200
5	2.5	97.5	100	200
10	5	95.0	100	200
15	7.5	92.5	100	200
25	12.5	87.5	100	200

2.4.1 Electrophoresis

In order to determine the concentration of SDS-PAGE gel which would produce the best visualization of the protein, a number of different concentrations, ranging from 3% to 12.5%, were tried. The best results for Dp116, Dp71, Dp40 and Dp30 were obtained using 7% gel running on low voltage of 60- 110 V.

The compounds that were used to make 7% resolving gel were 8ml of total volume of 4.0 ddH2O, 1.87 ml of 30% Acryl amide, 2.0 ml of Tris HCL (pH=8.8), 80 μl of 10% SDS, 80 μl of 10% APS, and 8.0 μl of TEMED. The mixture was poured three-fourths of the way up into the glass platters. A small quantity of bipropanol was poured over the mixture to level the gel surface. The gel was left 30 minutes to solidify. Next, a 6% stacking gel was prepared using the following materials: 2.6 ml of ddH20, 1.0 ml of 30% Acyl acrylamide, 1.25 ml of 0.5 Tris pH=6.8, 50 μl of 10% SDS, 50 μl of 10% APS, and 5 μl of TEMED. APS and TEMED were used to solidify the gel. This mixture was then poured on top of the resolving gel and was left for 30 minutes to solidify.

The next step was preparing the samples. In this case, 15 μl was used for the embryonic samples. Four embryonic samples were used: WT-wild type; MDX; Cav3$^{-/-}$ deficient mutant; and DM-double mutant, which is a homozygote for mdx and heterozygote for Cav3 protein. The mixture was prepared based on the protein concentration determined previously: for 15 μl of the sample, 7.5 μl of dye was used, and the rest was ddH2O. The mixture was boiled at 100°C for 5 minutes, after which the mixture was centrifuged at G 13000 for one minute, then kept on ice. Combs were removed after the gel had solidified completely. The gel holder was unclipped from the gel casting tray and was placed into the electrophoresis tank (BioRat Laboratories Ltd, UK). Inner and outer chambers were then filled with SDS running buffer (3 g of TRIS, 14.42 g of glycine, 1 g of

SDS and dH$_2$O to 1 L).   A quantity of 4 µl of protein marker (New England Biolabs LTD, UK) was loaded at the beginning of each set. The electroporator was connected to the power supply. The gel ran at 70 V, then changed to 110 V once the dye was stacked and the resolving gel added. Transfer blotting was conducted by using a turbo machine to transfer the protein out of the gel and onto the paper.

The next step was incubating the membrane in the blocking buffer for an hour on the shaker. The blocking buffer was prepared prior to taking the membrane from the turbo machine. The blocking buffer was made from 50 ml of TBS, 2.5 grams of nonfat dry milk, and 50 µl of Tween20.

For antibody detection, a fresh concentration of antibody was prepared and used at ratios of 1:1000, 1:100 or 1:50, and at a dilution of 1:20000 for β-actin, and the membrane was then incubated at 4℃ overnight.

The following day, the membrane was washed in a washing buffer three times for ten minutes each time, followed by incubation in biotinylated anti-mouse IgG at a concentration of 1:1000 for an hour whilst shaking, followed by washing the membrane in TBS three times for ten minutes each time. The membrane was then incubated in strepto-conjugated light sensitive solution at a concentration of 1:20000. The membrane was then washed with TBST three times for ten minutes each time. After that, the membrane was read using an imaging system (LI- comBiotech.uk).

I used the following formula to complete the statistical part of the experiment. The standard deviation was derived as follows: $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$; variance= σ^2 ; standard error= $(\sigma_{\bar{x}}) = \frac{\sigma}{\sqrt{n}}$ where $\bar{x}$ is sample's mean; n is the sample size and compared with statistical analysis of variance ANOVA generated by Microsoft excel.   Each western

blotting was repeated six times with different stages of development of the embryos (n is indicated in figures 16, 19 and 22) and was compared to WT E13.5, E15.5 and E17.5 to deduce the standard deviation of the mean. The standard deviation was obtained using analysis of variance (ANOVA) using Microsoft excel.

CHAPTER 3

Results

3.1 WT (wild type); MDX mouse; Cav3-/- deficient mouse; DMHet mouse

Apo-dystrophin forms are the product of the same dystrophin gene as the full form of dystrophin. Dp116, Dp71 and Dp40 play a crucial role in protecting the sarcolemma from distractive eccentric movements in the sarcolemma, in the absence of the full form of dystrophin Dp427. In the developing embryo, as the result of a genetic disorder, Dp427 is not synthesized, and the short forms of dystrophin Dp116, Dp71 and Dp40 are overexpressed. The overexpression of all apo-forms is an adaptive mechanism developed during evolution to protect the sarcolemma from the harmful forces of especially eccentric movements. Concentric movements do not lead to massive necrosis after birth.

Two methods were used to determine the localization and expression of apo-dystrophin and its distribution among different developing organs and systems in embryogenesis: immunohistochemistry, which visualizes dystrophin localization in different organs of developing embryo, and Western blotting, which uses of whole embryo already prepared to determine the molecular mass of the protein. It would have been better to have prepared samples from different organs in the embryo. This would have aligned very well with the already visualized dystrophin in the embryo using immunohistochemistry. However, I used already-prepared samples of whole embryos and as a result did not have the option of choosing to sample specific organs separately.

Immunohistochemistry was used to determine the location of apo-dystrophin in skeletal muscle, cardiac muscle, the brain, the retina, and the lungs of mice embryos E15.5 and E17.5. Different magnifications of electron microscopy allowed visualization of dystrophin and its apo- forms. In small magnifications, visualization was achieved in different organ systems of the embryo, while `larger` magnification allowed visualization in detail in different organs and systems of the embryo. The WT, the MDX, and the $Cav3^{-/-}$ deficient mice embryos were used to determine and visualize short forms of dystrophin.

Western blotting was then used to establish the length of the short forms of dystrophin. While it helped to determine the precise molecular mass of the short form of dystrophin, it was not particularly useful in determining the different apo-forms of dystrophin in the various organ localizations. In other words, Western blotting gave an overall picture of the dystrophin distribution at different stages of development of the embryo, but it was not able to determine the type of apo-dystrophin found in each organ. The samples used in Western blotting were made from whole embryos, which impaired the ability to determine exact types of apo-dystrophin that were found and their most expressed forms in different organs.

3.1.1 WT E15.5 and E17.5
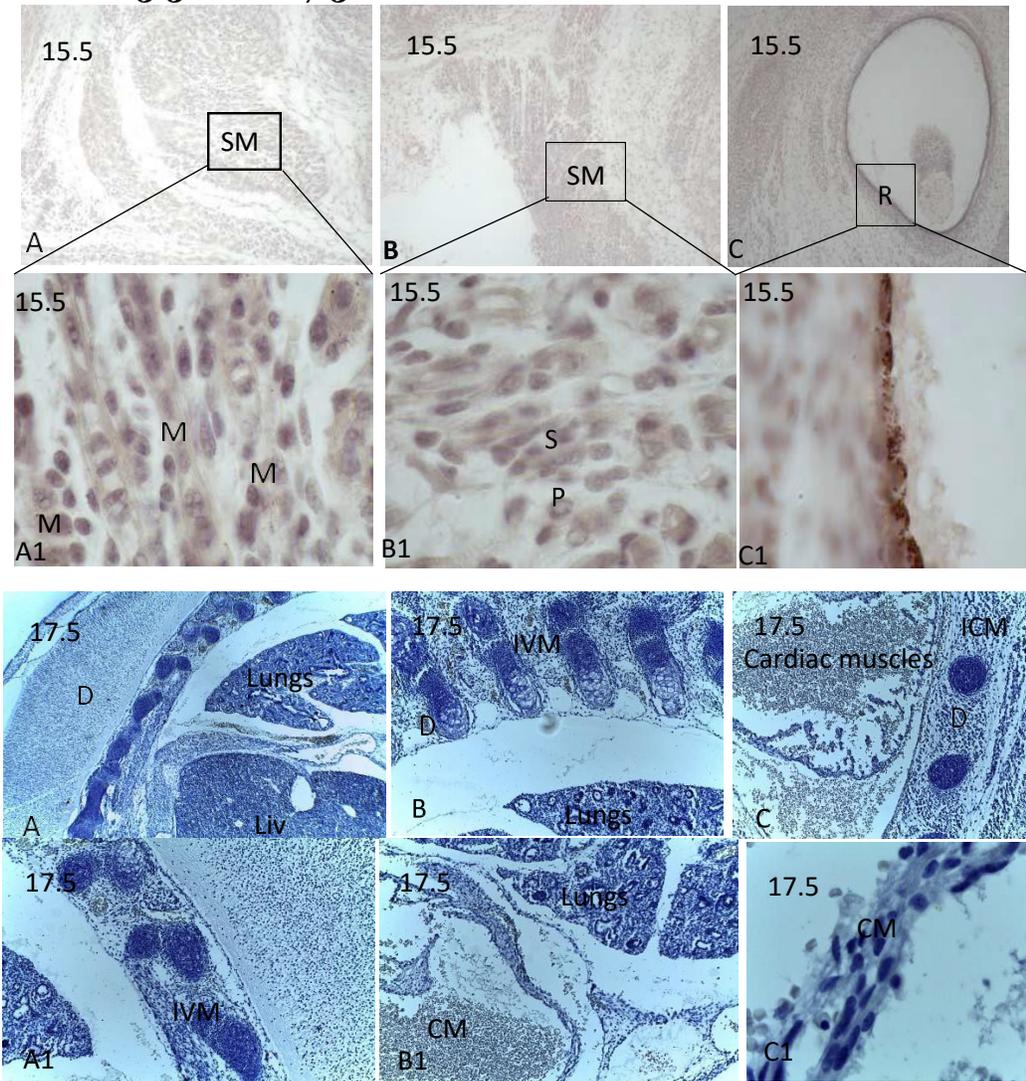
Figure 7. WT embryos 15.5 and 17.5. The first two samples represent developing striated muscles of the torso, and the third sample represents developing eye tissue in a 15.5 embryo. In embryo 17.5 the developing spinal cord, lungs, liver and heart can be seen clearly in the second and third row. In A and B, the skeletal muscle (SK)/A1 and B1 are the skeletal muscles; M represents a primary myotube, and S represents a secondary myotube. C represents the retina; dystrophin is widespread for WT 15. Dystrophin is abundant in the intercostal (ICM) and the inter- vertebrate muscles (IVM). It is also plentifully represented in the cardiac muscle (CM) and the lungs. All of this can be observed in the pictures of the WT 17.5 embryo. Results of Western blotting show that the Dp116 apo-form of Dystrophin in the mutant type of mice is overexpressed in embryogenesis due to lack of production of the full form of dystrophin.

Pathology is prominent in the Cav3$^{-/-}$ deficient mice, especially in the development of the heart in the mice embryo. Cardiomyopathy is very noticeable in the developing mouse heart. Clinically dilated cardiomyopathy is a disease that involves the myocardium where there is a visual enlargement of the left and right ventricles, which impairs systolic function and later leads to congestive heart failure and perivascular fibrosis. Myocyte necrosis and cellular infiltration could be present, but they are not prominent (Wynne and Braunwald 2000).

3.1.2 Cav3$^{-/-}$ E15.5 and E17.5
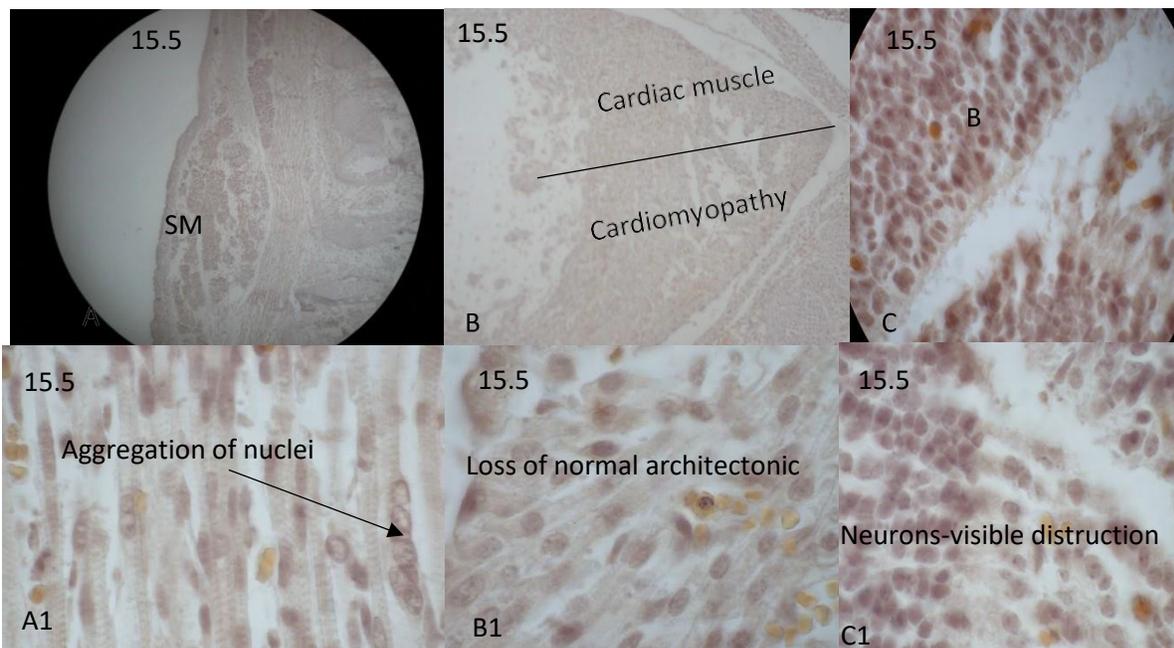
Figure 8. Cav3-/- deficient mice. The first sample shows the developing muscle of the back in an embryo of 15.5 days. The second sample is cardiac muscle, and the third sample is a developing brain in an embryo of 15.5 days This mutant typically develops very characteristic phenotypically expressed pathology. In small magnification, the destruction of skeletal muscle is visible (A), but under larger magnification (A1) the aggregation of nuclei is prominent; in cardiac muscle in small magnification cardiomyopathy is apparent (B), under larger magnification the normal architecture of the cardiac muscle is lost. Interruption of normal dendrite and glia configuration is observable in brain tissue under low magnification (C), distraction is visible under larger magnification (C1).
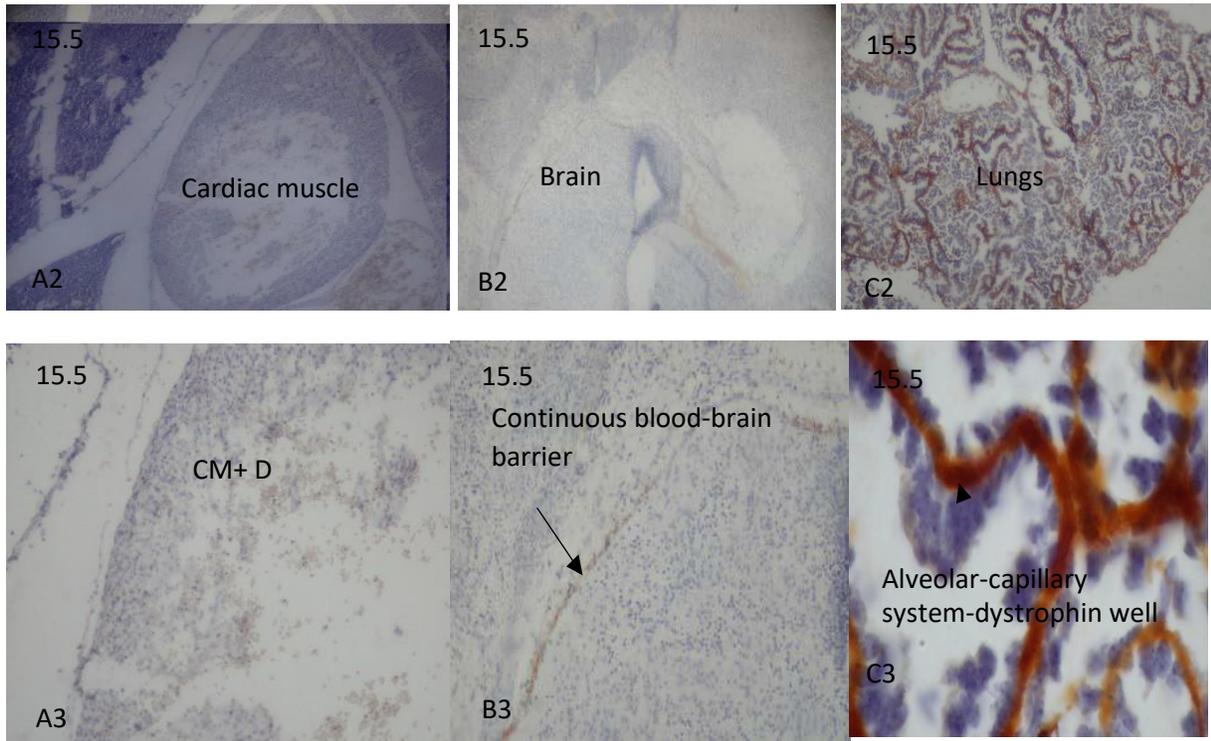
Figure 9. Cav3-/- deficient mutant. The first sample is cardiac muscle, while the second is brain tissue, and the third is the lungs tissue of the Cav3$^{-/-}$ deficient mutant. As expected, the cardiac muscle is hypertrophic at low magnification (A2); at higher magnification, the destruction is obvious, and dystrophin presence in the zone of destruction is visible (A3). The blood-brain barrier is a place where the dystrophin is located in the endo-epithelium of the capillary system (B2 and B3). The location of the dystrophin in the alveoli of the capillaries is the novel finding in this project (C2 and C3), but further research is needed to determine which apo-dystrophin form is present or the probability of a combination of the full form of dystrophin Dp427 and one or more forms of the apo-dystrophin Dp116, Dp71 and Dp40. Dp30 was detected as well, but its expression was vague. Further investigation is required to determine its location and its expression.
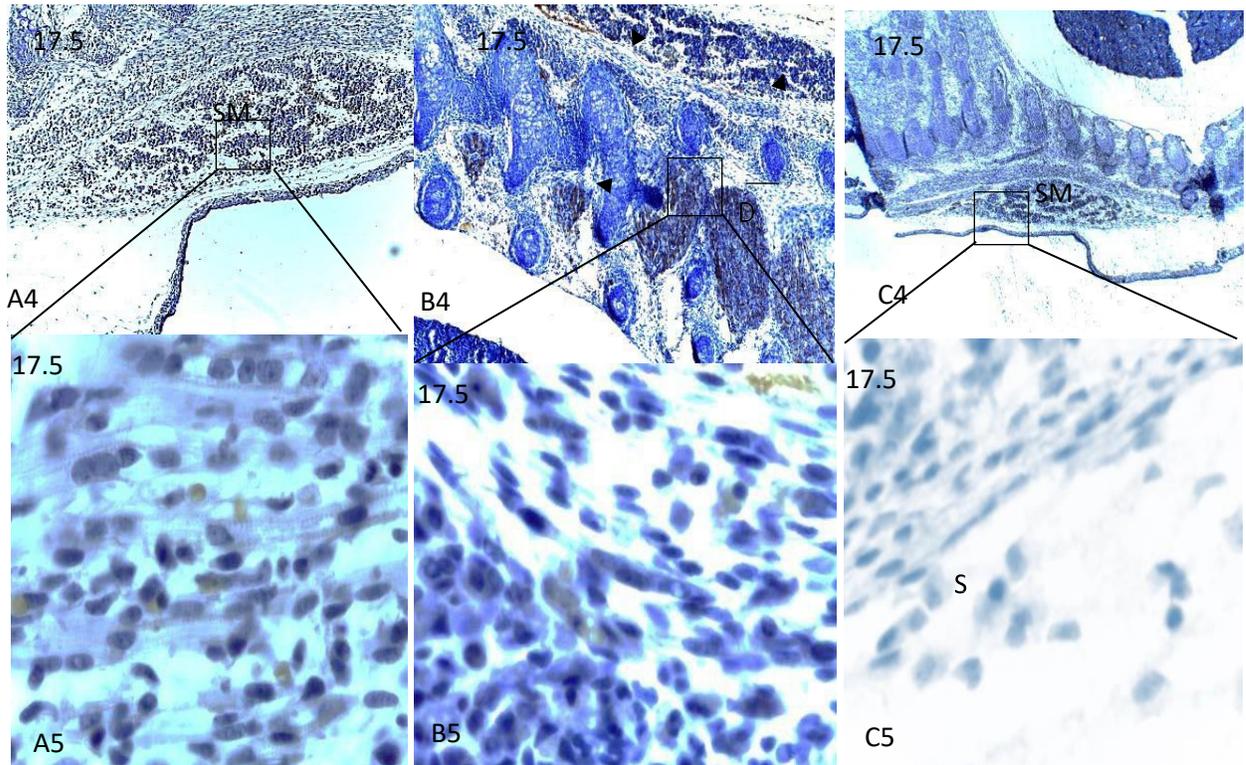
Figure 10. The first, second and third samples show developing back muscle tissue of the Cav3-/- deficient mutant in a 17.5 days old embryo. As other papers suggest, in this stage of development overabundant expression of dystrophin is seen in skeletal muscles. This may be the result of adaptation in the mutant, which lacks the full form of dystrophin Dp427.

In Figures 8, 9, and 10, the IHC shows areas where dystrophin was located mainly in the skeletal muscle (SM). Cardiomyopathy is very prominent in the Cav3-/- deficient mice. The reason for this finding could be that only the short forms of dystrophin are present, and Western blotting shows a spike of the presence of Dp71 in these mutants. Dp427 is not present in all kinds of mutants, and normal cardiac development can be achieved only in co-operation of the Dp427 and Dp71.

In the case of Cav3-/- there is hyper-production of the D p 71. The r e s u lt is cardiomyopathy (B and B1).

The brain in this type of mutant should be untouched by pathology, but at high magnification the neurons (C1) have disappeared, which suggests that the lack of the full form of dystrophin could be responsible for the alteration in the development of the Cav3-/- mice brains. Dystrophin was visualized in the cardiac muscle in embryo 15.5 (A2), but an abundance of dystrophin was seen in the inter-vertebrate muscle in embryo 17.5 (A4). Also, good visualization was achieved in the skeletal muscle sarcolemma in embryo 17.5 (A5). Some of the characteristic pathology in the Cav3-/- mice was very well visualized at greater magnification of the skeletal muscle. Where the nuclei were aggregated in one myoblast (A1), there was loss of normal architectonic structure in the miocardiocyte (B1).

The most interesting discovery was visualizing the apo-dystrophin in the continuous blood-brain barrier (B3) and the alveolar- capillary barrier in the lungs of embryo 15.5(C3). In Fig. 8, C5, the satellite cell was visualized. This is the source of the regeneration of injured muscle cells. The satellite cell in case of injury turns itself into a pluripotent cell, which becomes the source of division, growth, and recovery of the injured muscle. The mechanism is impaired in muscular dystrophy: instead of real healing, dead muscle cells are substituted with connective tissues such as collagen or adipose tissue.

Muscle tissue in muscular dystrophy is subjected to a heavy necrotic process triggered by simple movement followed by regeneration, especially in mice muscle; in human muscle tissue, regeneration does not occur, and the muscle tissue is replaced by adipose or connective tissue. The process of necrosis and regeneration establishes waves of development of muscular dystrophy, particularly in the skeletal muscle tissue. The waves of necrosis followed by regeneration, whether in muscle tissue in mice or connective or adipose tissue in humans, are characteristic of muscular dystrophy.

The principal reason for necrosis is apoptosis. Clinically defined, apoptosis is a genetically regulated self-destruction based on outside or inside stimuli. In cases of the DMD, many muscle cells die during the apoptotic period triggered by external stimuli, such as eccentric movement. This usually provokes and activates internal stimuli in the sarcolemma resulting from the lack of dystrophin and some of the proteins associated with dystrophin, which consequently trigger instability of the actin-dystrophin-basal lamina complex, leading to a massive apoptotic episode. These stimuli trigger a mechanism of activation of procaspase enzymes. The cell membrane has killer receptors on its surface which bind specific ligands that precipitate a cascade of reactions resulting in self-destruction (Elmore 2007). The process starts when a ligand occupies the killer receptor; for example, the Fas ligand produced by lymphocytes, which then binds to the Fac receptor on the cell surface. This induced clustering of Fas proteins induces production of an intracellular adaptor molecule that binds and aggregates the pro-caspace-8 molecule, which cleave and activate one another (Elmore 2007). The activated caspace-8 molecule then activates downstream pro-caspases to induce cell death. Damaged or stressed cells usually trigger the internal mechanism. In this case, the mitochondria of these cells produce and release protein cytochrome c into the cytosol, which binds and activates a protein called Apaf-1. The result is an activated cascade in which cell DNA is damaged.

For this process to be activated, another protein called p53 must be available with the help of proteins that belong to the Bcl-2 family. P53 participates in the gene transcription which encodes and releases cytochrome c from mitochondria. P53, in turn, regulates the activation of procaspases. Some of the these are Bcl-2, and Bcl-Xl, which inhibit apoptosis. Other members of the same family which promote apoptosis are Bax and Bad, which can be activated by another member such as Bid. Their action includes activation of cytochrome c from the mitochondria, while the Bad protein inhibits the death-inhibitor members of the

family. Clinical characteristics of apoptosis include cell shrinkage and condensation.

In the final stages, the cytoskeleton collapses, the nuclear envelope dissolves, and the DNA breaks into fragments. There is interruption of the plasma membrane. The holes in it aid leakage of its internal contents (Elmore 2007).

Mutant MDX in mice is in many aspects analogous to DMD in humans, but the difference between MDX and DMD is significant. As mentioned above, there are waves of necrosis and regeneration in mice, while in humans, full recovery is never achieved. The MDX mutant is still a good model because of the similarity in clinical characteristics of muscular dystrophy between mice and humans.

3.1.3 MDX E15.5 and E17.5
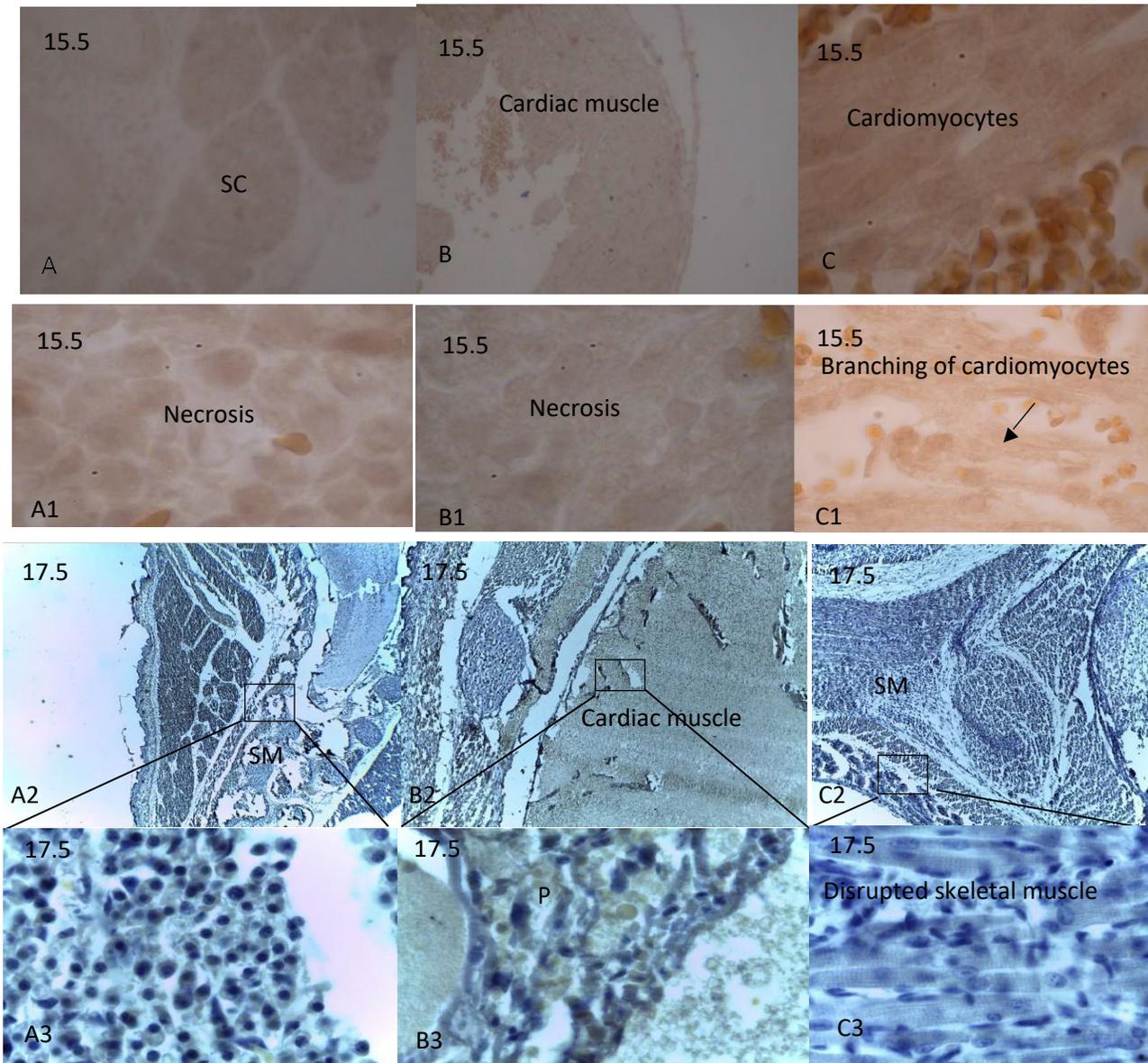
Figure 11. MDX in mice embryos 15.5 and 17.5. The first sample shows skeletal muscle of the torso in an MDX mutant at 15.5 days. The second sample is the cardiac muscle in the same MDX mutant, and the third sample in the top figure is cardiac muscle in greater magnification of the MDX mutant. The second-row images are of skeletal muscles in the developing torso of an MDX mice embryo. The second sample is cardiac muscle, and the third sample is skeletal muscle in the MDX mutant. In the MDX mice, deficient myotubes are hypotrophic (A and A1). The myotube alignment is disrupted and displaced (A1 and C3). Disrupted architecture of the muscle with missing muscle myotubes is apparent (C3). The nuclei are centrally located (A3). There is a morphological finding in the heart tissue of MDX mice that the cardiomyocytes are missing, and there is a branching of the cardiomyocytes (B1 andC1).

Hypo-staining in the heart tissue of MDX mice is the result of missing cardiomyocytes. The brain and the retina cells were seemingly normally developed. The reasons for this could be overexpression of utrophin despite the total lack of the full and short form of dystrophin. Utrophin compensates and probably helps mice brains and retinas to achieve full and normal development and may explain the full recovery in the mouse's skeletal muscles. This represents one of the principal reasons for the full recovery after a massive necrotic episode in mice. Utrophin compensates for the scarcity of dystrophin and its apo-forms in mice, while in humans, production of utrophin is depressed in parallel to the lack of the dystrophin.

There are no IHC slides for the representation of apo-dystrophin in MDX embryos, due to lack of MDX embryos, but it is possible that there would be same path of expression for utrophin in mice embryos as a consequence of finding new ways to make up for the lack of the full form of dystrophin Dp427. Further experiments should be conducted to determine the representation of apo-dystrophin in MDX embryos and how apo-dystrophin would change the development of the mouse embryo.

This raises another interesting question. In the case of an IHC of the MDX mouse embryo, how would apo-dystrophin mimic the real damage from the lack of the Dp427? The spike in concentration of the Dp40 in the Cav3-/- is dramatically manifested, which once again shows that overexpression of the Dp40 compensates for the lack of the full form of dystrophin in this type of mutant. In fully presented results, the most interesting mutant is the Cav3-/- because it shows a dramatic change in the concentration of production of the short forms of dystrophin.

3.1.4 β-dystroglycan E15.5

Finally, there are some results from using a primary monoclonal antibody against β-dystroglycan. β-dystroglycan is an integral and important part of the DAPC. In the WT mice, the results of staining are the same as in the case of using primary monoclonal antibodies against dystrophin. This demonstrates that the β-dystroglycan is part of the DAPC and is integrated closely with dystrophin.

The MDX-deficient mice will show the pathology in which the full form of dystrophin is missing, but in such a case, the β-dystroglycan would bind to the existing short form of dystrophin or utrophin, and as a result, it is difficult to determine whether there is any pathology. Utrophin, in many cases, is a very promising protein that may substitute for the lack of dystrophin in those patients presenting with different forms of muscular dystrophy.
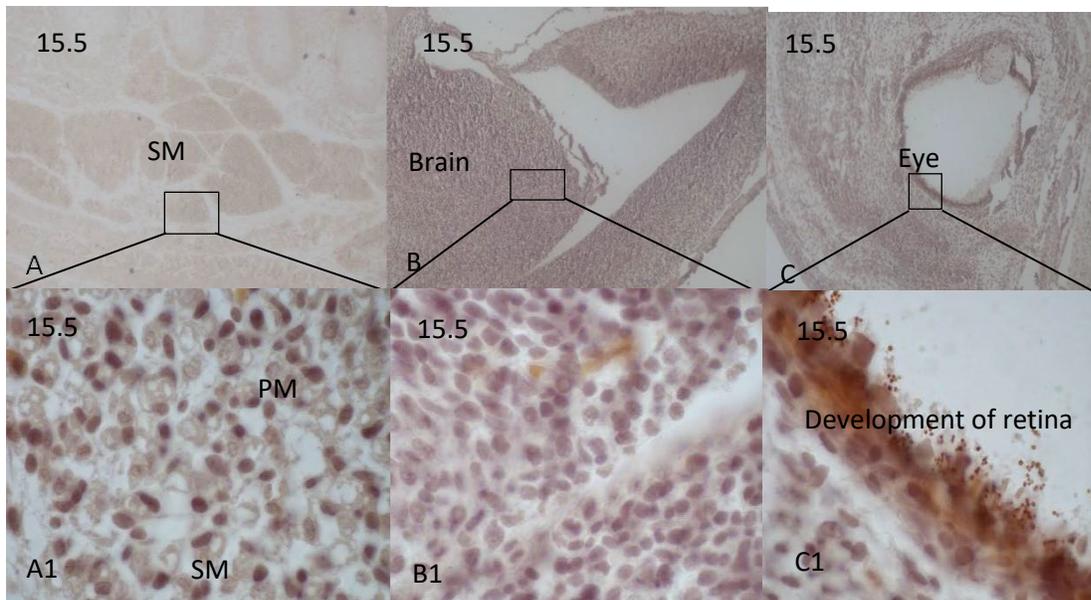
Figure 12. β-dystroglycan. The first sample is skeletal muscle located around the developing spine of the embryo. The second sample is developing brain tissue, and the third sample is developing eye tissue in a 15.5-day-old embryo. β-dystroglycan is an integral part of the DAPC complex and plays a particularly important role in creating stability in the DAPC. The primary and the secondary myotubes are well visualized in A1 of embryo 15.5. The architectonic structure of the brain is very well developed with no alteration overall, but the striking expression of β-dystroglycan can be seen in the developing retina. The short form, the Dp260 of apo-dystrophin, overcomes the other short forms of apo-dystrophin (Neuman 2005).

Directly connected to dystrophin, β-dystroglycan plays a stabilizing role but also actively contributes to formation of a bridge between the cell muscle's cytoplasm and laminin 2 through the α-dystroglycan. In muscular dystrophy, a missing but particularly important link in the chain is the full form of dystrophin. This creates instability, and the sarcolemma is prone to injuries even with slight exertion. It is important to note that β-dystroglycan, even if it exists in the chain, will not rescue the sarcolemma from injury.

3.2 Western Blotting

All results of Western blotting were extrapolated from samples made from whole embryos. Different stages in the development of specially prepared embryos were used to determine the concentration of the protein in the WT and MDX, Cav3$^{-/-}$ and DMHet (MDXCav3$^{-/-}$ double mutant). Unknown samples were diluted and compared with a set of bovine serum albumin standards. ANOVA test was use and result was analyzed. ANOVA statistical analysis produced the following results:

Dp116; ANOVA P=0.91 (Including here the average results from E13.5, E15.5 and E17.7)

Dp71; ANOVA P=0.77

Dp40; ANOVA P=0.78

E13.5

Dystrophin
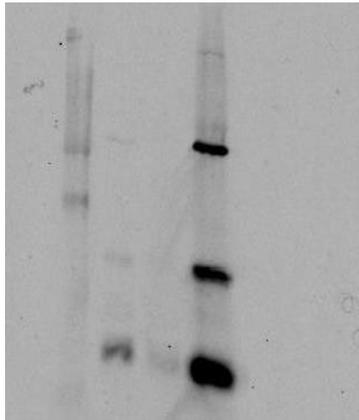
Dp116

Dp71

Dp40

Dp30

β-actin

M   WT Cav3⁻ᐟ⁻ DMHet

E15.5

Dystrophin

Dp116

Dp71

Dp40

Dp30

β-actin

52kDa

M WT Cav3⁻ᐟ⁻DMHet

M WTCav3⁻ᐟ⁻DMHet

M WTCav3⁻ᐟ⁻DMHet

E17.5

Dystrophin
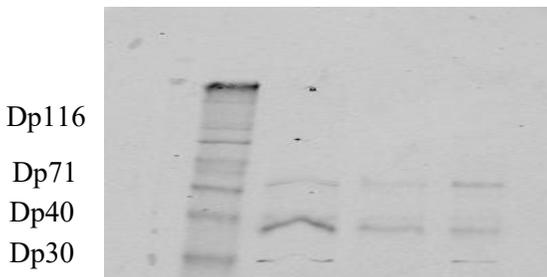
Dp116
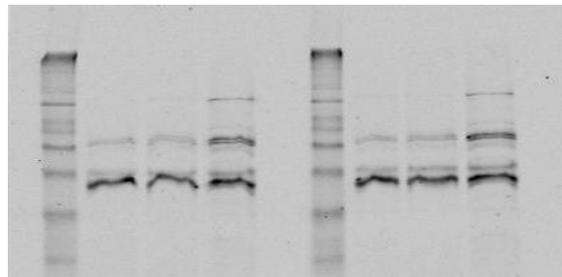
Dp71

Dp40

Dp30

β-actin   β-actin

M   WT   Cav3⁻ᐟ⁻ DMHet   M WT   Cav3⁻ᐟ⁻ DMHet   M WT   Cav3⁻ᐟ⁻ DMHet

Figure 13. Western blotting of E13.5, E15.5, and E17.5. Dp30 was detected, but its expression is not particularly prominent. It is prominently expressed in E13.5 in DMHet, but its expression is diminished in E15.5 and E17.5 stage.

Figure 14 presents the expression of Dp116 in embryonic development.
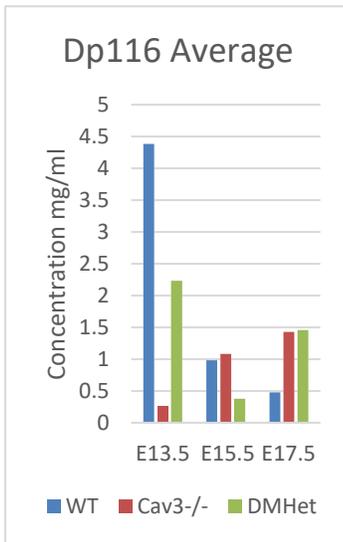
Figure 14. Dp116 average. Dp116 average shows clearly that the expression of Dp116 is very explosive in the WT in E13.5 while the same protein is under-expressed in Cav3$^{-/-}$, and its expression in DMHet is almost half of the WT expression. In late stage development of E15.5, the production is changing. While Dp116 is dramatically suppressed in WT, Cav3$^{-/-}$ is triple the amount compared to E13.5, and the same protein is again suppressed in DMHet in E15.5. Dp116 is quadruple the amount in Cav3$^{-/-}$ in E17.5 compared to E13.5, and in WT the production of Dp116 even dipped, while in DMHet expression is reduced compared to E13.5 but increased compared to E15.5.
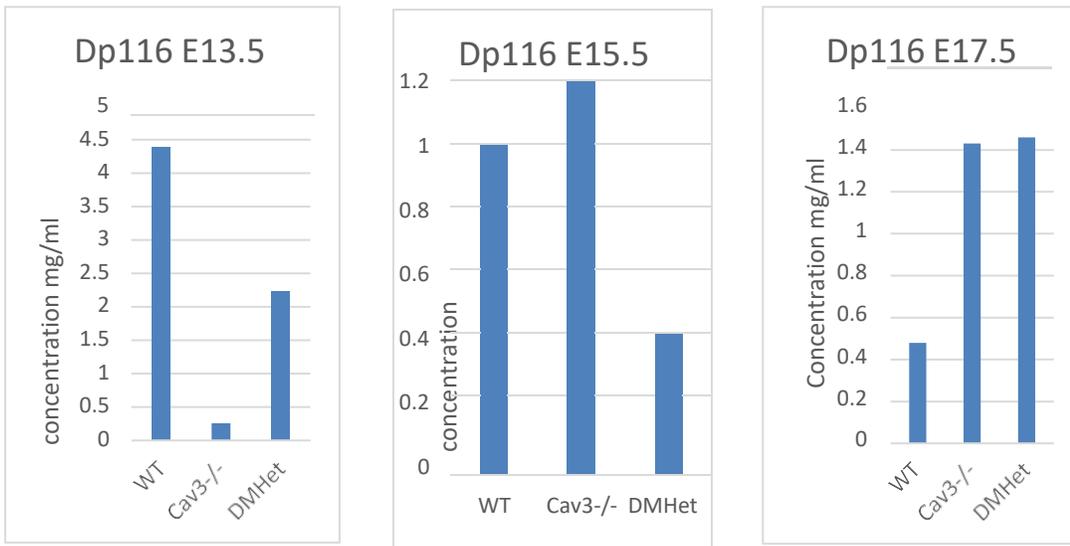
Figure 15. Dynamic expression of Dp116 at different stages of development.
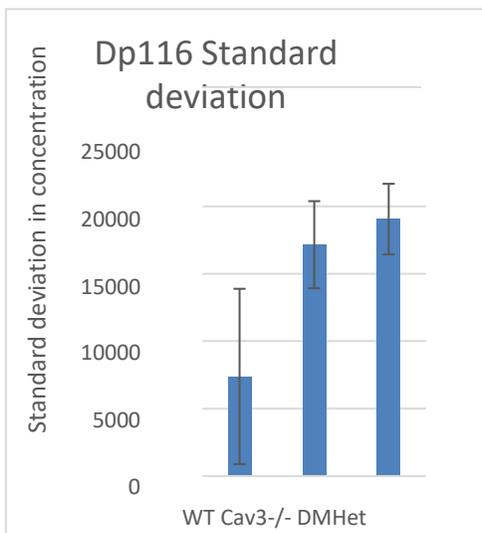
Figure 16. Standard deviation of Dp116 at different stages of embryonic development. Error bars show the standard deviation of the mean, calculated using the formula presented on p. 35 (ANOVA; P>0.05 Dp116 E13.5, E15.5 and E17.5; controls n=6)

Dp71 has been connected to the development of cardiomyopathy in the absence of the full form Dp427. It could be speculated that Cav3$^{-/-}$ mutants developed cardiomyopathy as a result of over-expression on Dp71 which, combined with genetic mutation, leads to complete absence of the production of Dp427.
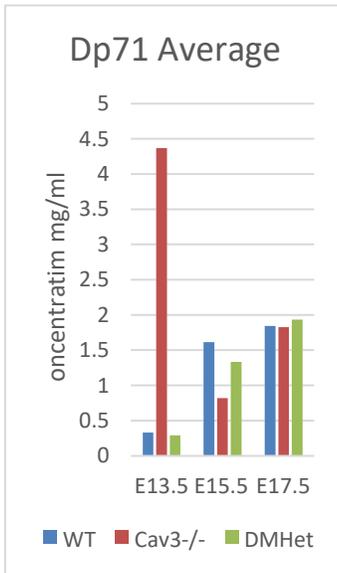
Figure 17. Dp71. As predicted, over-expression of this form is observed in E13.5. At this point in embryonic development, Dp71 is overexpressed four-fold, which may be the result of the fact that at this stage, initial development is ending. The expression slows down in the next stage as a result of compensatory forces resulting from a lack of Dp427. Dp71 levels are initially very low, expressed in WT in stage E13.5, which could be explained by the fact that Dp427 is expressed normally in WT, and that Dp71, combined with the full form of dystrophin, completes the requirements for normal muscular development. The appearance of Dp71 in DMHet is low in stage E13.5. This could be because this is a heterozygous mutant. In both WT and DMHet, Dp71 normalizes its production during late stages.
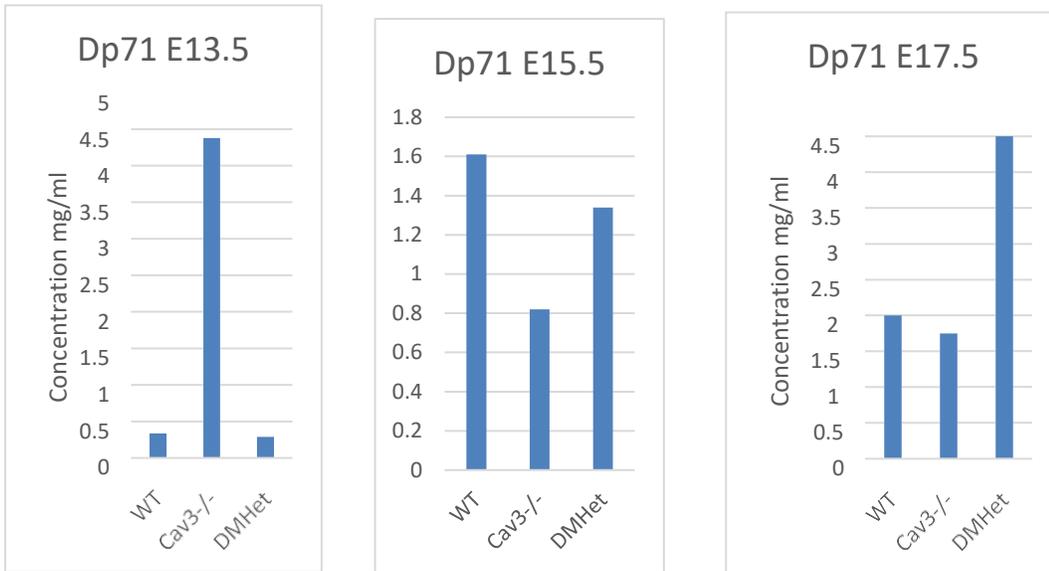
Figure 18. Expression of Dp71. Dp71 is over-expressed in DHMet in E17.5. This could be the result of adaptation occurring at this stage, in the absence of the full form of dystrophin Dp427.
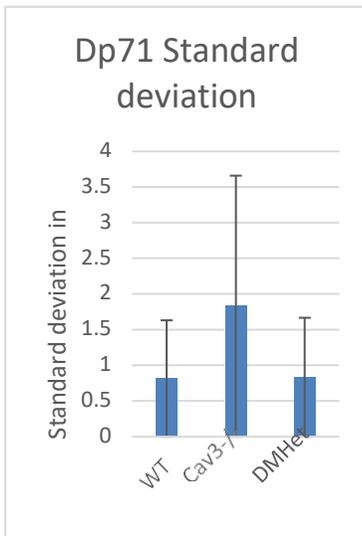
Figure 19. Standard deviation in the Dp71 form at different stages of development of the embryo. Error bars show the standard deviation of the mean, calculated using the formula presented on p. 35 (ANOVA; P>0.05 Dp71 E13.3, E15.5 and E17.5; controls n=6)

The Dp40 apo-form shows intensive expression in Cav3-/- in all stages of development of the embryo. This could be explained with the adaptation mechanism that this type of mutant developed during embryogenesis. It could be speculated that the combination of Dp40 and Dp71 is the reason for developing of cardiomyopathy, but further research is needed to confirm or reject this finding.
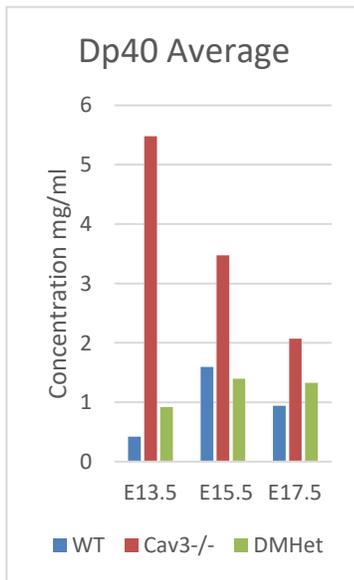
Figure 20. Average expression of Dp40. Over-expression of Dp40 is present in E13.5. It could be speculated that this form of apo-dystrophin is affiliated with Dp71 and that both are participating in developing cardiomyopathy, as IHC shows in the Cav3[-/-] mutant. Dp40 expression is 4.5 times more in Cav3[-/-] in early stages than in the other mutants, and the expression of Dp40 slows down in late stages of development. Dp40 in WT is depressed at early stage development (E13.5) compared to the level of expression during a subsequent wave of production in later embryogenesis.
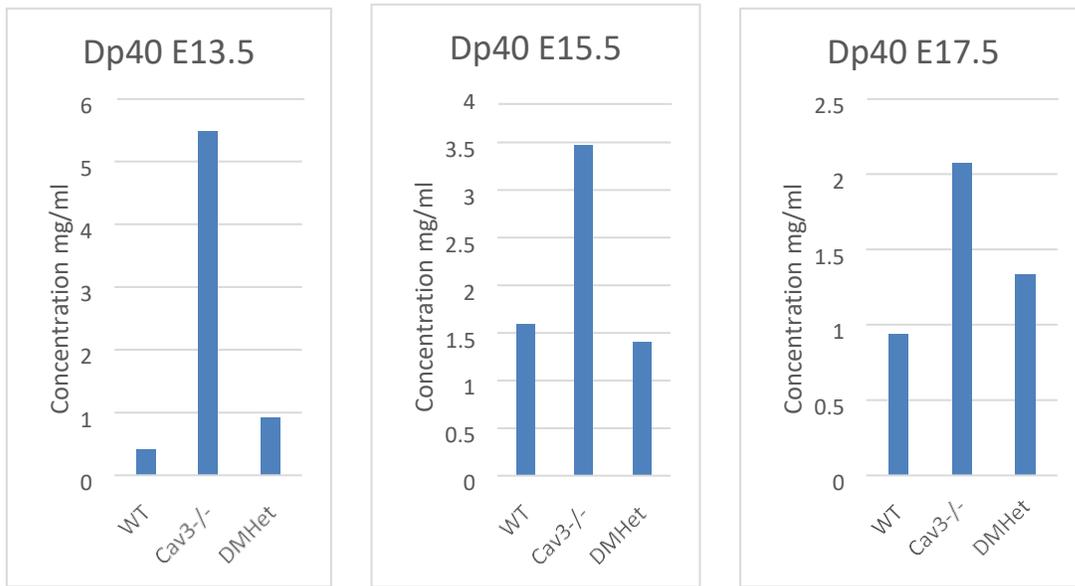
Figure 21. Spikes in expression of Dp40 in all stages of development in Cav3-/- represent an adaptation that arises due to lack in synthesis of the full form of dystrophin Dp427 based on genetic mutation in the dystrophin gene.
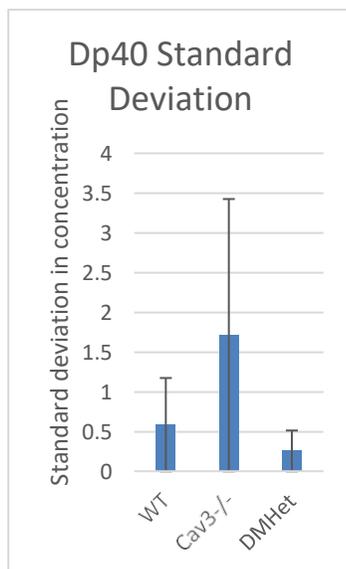
Figure 22. Standard deviation of Dp40 in all stages of development. Error bars show the standard deviation of the mean, calculated using the formula presented on p. 35 (ANOVA; P>0.05 Dp40 E13.5, E15.5 and E17.5 controls n=6)

Chapter 4

Discussion

The results from IHC in the Cav3$^{-/-}$ mutant show increased number and size of myotubes and hypertrophy in trunk muscles, along with particularly acute hypertrophy of the heart. In the same mutant, the thickness of the intercostal muscle is not changed, but there is an increased fiber density. Neuron destruction is visible in this type of mutant (Fig.8 C1 sample), which can lead to cognitive disfunction. Branching and fiber splitting is not characteristic in the Cav3$^{-/-}$ mutant. However, an increased number of myonuclei with displacement with some aggregation (Fig.8 A1 sample) is a prominent trait in the Cav3$^{-/-}$ mutant.

Cardiomyopathy is the dominant trait for the Cav3$^{-/-}$ mutant (Fig.8 B sample). Some cardiac atrium-trabecular defects (Fig.9 A2 sample) are observed, but results are inconclusive. For this reason, additional experiments with Cav3$^{-/-}$ mutants should be conducted. Apo-dystrophin is aligned very well in the alveolar-capillary system (Fig.9 C2 sample) in the Cav3$^{-/-}$ mutants. In western blotting, expression of Dp116 apo-dystrophin is depressed in early stage of development (E13.5) in the Cav3$^{-/-}$ mutant, while in later development, expression is doubled in E15.5 and tripled in E17.5, which can be interpreted as a compensatory mechanism for lack of expression of Dp427. I observed a massive spike in the expression of Dp71 in the Cav3$^{-/-}$ mutant at E13.5. In contrast, the expression of Dp71 should normally be more prominent in WT, but in my experiments, the expression is depressed in WT. A possible explanation for the spike of Dp71 in Cav3$^{-/-}$ is compensation for the lack of Dp427 in this mutant later in development. Dp40 also demonstrates a spike in expression in the E13.5 Cav3$^{-/-}$ mutant embryo, which is consistent with the result from

IHC for increased fiber density.

IHC of the MDX mutant shows reduction in the number and size of myotubes with decrease of fiber density (Fig.11 A sample), along with disruption in myotube alignment (Fig.11 C3 sample). Increased branching and fiber splitting (Fig.11 C1 sample) is also observed. There is no visible increase and displacement of myonuclei in the MDX mutant. Cardiac thickening is visible (Fig.11 B2 sample), but cardiac abnormality in the atrium and trabecular are not found. Theoretically, the MDX mutant should have this type of abnormality. It is necessary to repeat this IHC staining to explore these results further.

IHC of WT in E15.5 and E17.5 shows normal size myotubes (Fig. 7A samples for both E15.5 and E17.5), along with normal fiber density (Fig.7 A and B samples E15.5, as well A, B and C samples E17.5). There is no branching and fiber splitting in WT individuals. Myonuclei are located normally but are not displaced, and the cardiac wall is normal in WT (Fig. 7 B1 sample). Western blotting results demonstrate an explosion in the expression of Dp116 in WT later in development, while expression of Dp116 is normalized. In contrast, the expression of Dp71 is depressed in early stages of embryogenesis in WT and later is normalized again. The same pattern can be seen with regard to Dp40 in WT.

Western blotting for the concentration of DMHet shows expression in concentration of Dp116 two times greater than in than Cav3^{-}/$^{-}$, which is probably connected with adaptation mechanisms in individuals with mutations in the dystrophin gene. The concentration of expression of Dp116 in DMHet is extremely low in later development, while the concentration of Dp71 is depressed in early stages of development but is normalized later in embryogenesis. The concentration of expression of Dp40 in DMHet remains steady for the duration of embryogenesis. This is likely a compensation for lack of the Dp427 protein.

List of References

[1] M. D. Adams, M. H. Butler, T. M. Dwyer, M. F. Peters, A. A. Murnane, and S. C. Froehner. Two forms of mouse syntrophin, a 58 kDa dystrophin-associated protein, differ in primary structure and tissue distribution. *Neuron*, 11(3):531–540, 1993.

[2] M. E. Adams, N. Kramarcy, S. P. Krall, S.G. Rossi, R. L. Rotundo, R. Sealock, and S. C. Froehner. Absence of alpha-syntrophin leads to structurally aberrant neuromuscular synapses deficient in utrophin. *The Journal of Cell Biology* 150(6):1385-1398,2000.

[3] K. J. Amann, B. A. Renley, and J. M. Evrasti. A cluster of basic repeats in the dystrophin domain binds F-actin through an electrostatic interaction. *Journal of Biology and Chemistry*, 273(43):28419–28423,1998.

[4] G. Artz and J. Wynne. Restrictive cardiomyopathy. *Current Treatment Options in Cardiovascular Medicine* 2(5):431-438, 2000.

[5] R. Barresi, S.A. Moore, C. A. Stolle, and K. P. Campbell. Expression of gamma-sarcoglycan in smooth muscle and its interaction with smooth muscle sarcoglycan-sarcospan complex. *Journal of Biology and Chemistry*, 275(49):38554–38560,2000.

[6] D. J. Blake, R. Nawrotzki, M. F. Peters, S. C. Froehner, and K. E. Davies. Isoform diversity of dystrobrevin, the murine 87-kDa postsynaptic protein. *Journal of Biology and Chemistry*, 271(13):7802-7810, 1996.

[7] P. Bork and M.Sudol.The WW domain: a signalling site in dystrophin? *Trends in Biochemical Science.* 19(12):531–533, 1994.

[8] M. Bermudez de Leon, C. M. Pablo Gomez, S. L. Moralez-Lazaro, V. Tapia-Ramirez, V. Valadez-Graham, Recillas-Tagra, D. Yaffie, U. Nudel, and B. Cisneros. Dystrophin Dp71 is downregulated during myogenesis: Role of Sp1 and Sp3 on the Dp71 promoter activity. *Journal of Biology and Chemistry,* 280:5290–5299,2005.

[9] G. S. Bewick, L. V. Nicolson, C. Young, E. O'Donnell, and C. R. Slater. Different distributions of dystrophin and related proteins at nerve-muscle junctions. *Neuroreport,* 3:857–860, 1992.

[10] R. Bies, D. Freedman, R. Roberts, B. M. Perryman, and T. C. Casky. Expression and localization of dystrophin in human cardiac Purkinje fibers. *American Heart Association,* 86:147–153, 1992.

[11] D. J. Blake, R. Nawrotzki, N. Y. Loh, D. C. Gorecki, and K. E. Davies. Beta-dystrobrevin of dystrobrevin, the murine 87-kDa postsynaptic protein family. *Journal of Biology and Chemistry,* 271(13):7802–7810, 1996.

[12] C. S. Brown and A. J. Lucy. (Editors). *Dystrophin: Gene, Protein, and Cell Biology.* Cambridge: Cambridge University Press, 1997.

[13] K. M. Busby. The limb-girdle muscular dystrophies-multiple genes, multiple mechanisms. *Human Molecular Genetics,* 8(10):1875–1882, 1999.

[14] T. J. Byers, H. G. Lidov, and L. M. Kunkel. An alternative dystrophin transcript specific to peripheral nerve. *National Genetics,* 4(1):1875–1882,1993.

[15] J. T. Campanelli, L. S. Roberds, K. P. Campbell, and Scheller, R. H. A role for dystrophin-associated glycoproteins and utrophin in agrin-induced AChR clustering. *Cell,* 77(5):663–674, 1994.

[16] K. P. Campbell and S. D. Kahl. Association of dystrophin and an integral membrane glycoprotein. *Nature*, 338(6212):259–262., 1989.

[17] T. Claudepierre, F. Rodius, M. Frasson, V. Fontaine, S. Picaud, H. Dreyfus, D. Mornet, and A. Rendon. Differential distribution of dystrophins in rat retina. *Investigative Ophthalmology and Visual Sciences,* 40:1520–1529,1999.

[18] G. Cossu, R. Kelly, S. Tajbakhsh, S. Di Donna, E. Vivarelli, and M. Buckingham. Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm. *Development,* 122:429–437,1999.

[19] W. Chung and J. T. Campanelli. WW and EF hand domains of dystrophin-family proteins mediate dystroglycan binding. *Molecular Cell Biology Research Communications,* 2(3):162–171, 1999.

[20] A. J. Coffey, R. G. Roberts, E. D. Green, C. G. Cole, R. Butler, R. Anand, F. Giannelli, and D. R. Bentley. Construction of a 2.6-Mb in yeast artificial chromosome spanning the human dystrophin gene using an STS-based approach. *Genomics,* 12:474–484,1992.

[21] R. H. Crosbie, J. Heighway, D. P. Venzke, J. C. Lee, and K. P. Campbell. Sarcospan, the 25kDa transmembrane component of the dystrophin-glycoprotein complex. *Journal of Biology and Chemistry,* 272(50):31221–31224, 1997.

[22] K. Culligan, L. Glover, P. Dowling, and K. Ohlendieck. Brain dystrophin-glycoprotein complex: Persistent expression of β-dystroglycan, impair oligomerization of Dp71 and up-regulation of utrophins in animal models of muscular dystrophy MBC. *Cell Biology,* 2:2, 2001.

[23]   S. Dietrich, F. R. Schubert, C. Healy, P. T. Sharpe, and A. Lumsden. Specification of the hypaxial musculature. *Development,* 125:2235–2249, 1998.

[24]   V. N. D'Souza, N. Man, and G. E. Morris. A novel dystrophin isoform is required for normal retinal electrophysiology. *Human Molecular Genetics* 4, 837–842,1995.

[25]J. P. Earnest, G. F. Santos, S. Zuerbig, and J. E. Fox. Dystrophin-related protein in the platelet membrane skeleton. Integrin-induced change in detergent-insolubility and cleavage by calpain in aggregating platelets. *Journal of Biochemical Chemistry*, 270:27259–27265, 1995.

[26]   S. Elmore. Apoptosis: a review of programmed cell death. *Toxicol Pathology* 35:4: 495-516, 2007.

[27]J. M. Evrast and K. P. Campbell. A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *The Journal of Cell Biology*, 122:809–823, 1993.

[28]   C. A. Feener, M. Koening, and L. M. Kunkel. Alternative splicing of human dystrophin mRNA generates isoform at the carboxyl terminal. *Nature,* 338(6215):509–511,1989.

[29]   S. H. Gee, R. Madhavan, S. R. Levinson, J. H. Caldwell, R. Sealock, and S. C. Froehner. Interaction of muscle and brain sodium channels with multiple members of syntrophin family of dystrophin-associated protein. *Journal of Neuroscience*, 18(1):128–137, 1998.

[30]   D. C. Gorecki and E. A. Barnard. Specific expression of G-dystrophin (Dp71) in the brain. *Neuroreport,* 6:893–896, 1995.

[31] M. Hasegawa, A. Cuenda, M. G. Spillantini, G. M. Thomas, V. Buee-Scherrer, P. Cohen, and M. Goedert. Stress-activated protein kinase 3 interacts with PDZ domain of alpha1-syntrophin. A mechanism for specific substrate recognition. *Journal Biology and Chemistry,* 274(18):12626–12631, 1999.

[32]   D. Hazai, K. Halasy, D. Mornet, F. Hajos, and V. Jancsik. Dystrophin splice variants are distinctly localized in the hippocampus. *Acta Biologica Hungarica,* 57(2): 141-146,2006.

[33]   P. J. Holzfeind, H. J. Ambrose, S. E. Newey, R. A. Nawrotzki, D. J. Blake, and K. E. Davies. Tissue-selective expression of alpha-dystrobrevin is determined by multiple promoters. *Biochemistry and Chemistry*, 274(10): 6250:6258,1999.

[34]   P. L. Howard, H. J. Klamut, and P. N. Ray. Identification of a novel actin binding site within Dp71 dystrophin isoform. *FEBS Letter,* 441:337–341,1998.

[35]X. Huang, F. Poy, R. Zhang, A. Joachimiak, M. Sudol, and M. J. Eck. Structure of a WW domain containing fragments of dystrophin in complex with beta-dystroglycan. *Nature*

Structural and Molecular Biology, 7(8):634-638, 2000.

[36]   O. Ibrahimova-Beskrovnaya, J. M. Ervasti, C. J. Leveille, C. A. Slaughter, S. W. Sernett and K. P. Campbell. Primary structure of dystrophin–associated glycoproteins linking dystrophin to the extracellular matrix. *Nature*, 335(6362):696–702,1992.

[37] M. Ikeya and S. Takada. Wnt singling from dorsal neural tube is required for the formation of the medial dermamyotome. *Development,* 125:4696–4697,1998.

[38]   M. Inoue, Y. Wakayama, J. W. Lui, M. Murashi, S. Shibuya, and H. Onikiof. Ultrastructural localization of aquaporin 4 and alpha1-syntrophin in vascular feet of brain astrocytes. *Tohoku Journal of Experimental Medicine*, 197(2):87–93,2002.

[39]   D. Jung, B. Yang, J. Meyer, J. S. Chamberlain, and K. P. Campbell. Identification and characterization of the dystrophin anchoring site on beta-Dystroglycan. *Journal of Biology and Chemistry,* 270(45):27305–27310, 1995.

[40]   T. J. Kadonaga, K. R. Karner, F. R. Maziars, and R. Tijan. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell,* 51(6):1079–1090, 1987.

[41] N. H. Keep, S. J. Winder, C. A. Moores, S. Walke, F. L. Norwood, and J. Kendrick-Jones. The 2.0 A structure of the second calponin homology domain from actin-binding region of the dystrophin homology utrophin. *Journal of Molecular Biology,* 285:1257–1264, 1999.

[42]   N. H. Keep, S. J. Winder, C. A. Moores, S. Walke, F. L. Norwood, and J. Kendrick-Jones. Crystal structure of actin-binding region of utrophin reveals a head-to-tail dimer. *Structural Folding and Design,* 7:1539–1546, 1999.

[43]   M. Koening, E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener, and L. M. Kunkel. Complete cloning of Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of DMD gene in normal and affected individuals. *Cell,*50:509–517, 1987.

[44]   K. A. Knusden. The calcium-dependent myoblast adhesion that proceeds cell fusion is mediated by glycoproteins. *Journal of Cell Biology,* 101:891–897,1985.

[45]K. A. Knusden, S. A. McElwee, and L. Myers. A role for the neural cell adhesion molecule, N-CAM, in myoblast interaction during myogenesis. *Developmental Biology,* 138:159– 168, 1990.

[46]   L. M. Kunkel, A. P. Monaco, E. Hoffman, M. Koening, C. Feener, and C. Bertelson. Molecular studies of progressive muscular dystrophy (Duchenne). *Enzyme*. 38(1–4):72–75, 1987.

[47]   C. S. Lebakken, D. P. Venke, R. F. Hrstka, C. M. Consolino, J. A. Faulkner, R. A.

Williamson, and K. P. Campbell. Sarcospan-deficient mice maintain normal muscle function. *Molecular Cell Biology*, 20(5):1669–1677, 2000.

[48]   D. Lederfein, Z. Levy, N. Augier, D. Morent, G. Morris, O. Fuchs, D. Yaffe, and U. Nudel. A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other non-muscle tissue. *Proceedings of the National Academy of Sciences USA,* 89(12):5346–5350, 1992.

[49]   H. G. Lidov, S. Selig, and L. M. Kunkel. Dp140: a novel 140 kDa CNS transcript from the dystrophin locus. *Human Molecular Genetics,* 4(3):329–335,1995.

[50]   S. Lin, F. Gaschen, and J. M. Burgunder. Utrophin is a regeneration-associated protein transiently present at the sarcolemma of regenerating skeletal muscle fibers in dystrophin-deficient hypertrophic feline muscular dystrophy. *Journal of Neuropathology and Experimental Neurology,* 57:780–790, 1998.

[51] B. P. Lotz and A. G. Engel. Are hyper-contracted muscle fibres artifacts and do they cause rupture in the kidney? *Journal of Cell Science*, 37:1466–1475, 1987.

[52] C. Marcelle, M. R. Stark, and M. Bronner-Fraser. Coordinate actions of BMPs, Wnts, Shh, and Noggin mediate patterning of the dorsal somite. *Development*, 124:3955–3963, 1997.

[53] M. Marotto, R. Reshef, A. E. Munsterberg, S. Koester, M. Goulding, and A. B. Lasser Ectopic Pax3 activates MyoD and Myf5 expression in embryonic mesoderm and neural tissue. *Cell,* 89:139–148, 1997.

[54] N. Masubuchi, Y. Shidon, S. Kondo, J. Takato, and K. Hanoaka. Subcellular localization of dystrophin isoforms in cardiomyopathy and phenotypic analysis of dystrophin isoforms in cardiomyocytes and phenotypic analysis of dystrophin-deficient mice reveal cardiac myopathy is predominantly caused by a deficiency in full-length of dystrophin. *Experimental Animals,* 62(3), 211-217, 2013.

[55] K. Matsumura, H. Yamada, T. Shimuzu, and K. P. Campbell. Differential expression of dystrophin, utrophin and dystrophin-associated proteins in peripheral nerves. *FEBS Letter*, 334:281–285, 1993.

[56] D. Merrick, L. K. Stadler D. Larner, and J. Smith. Muscular dystrophy begins early in embryonic development deriving from stem cell loss and disrupted skeletal muscle formation. *Disease Model Mechanisms* 2(7-8): 374-388, 2009.

[57] C. A. Moores and J. Kendrick-Jones. Biochemical characterization of the actin-binding properties of utrophin. *Cell Motility Cytoskeleton*, 46:116–128, 2000.

[58]   A. E. Munsterberg, J. Kitajewski, D.A. Bumcroft, A. P. McMahon, and A. B. Lassar.

Combinatorial signalling by sonic hedgehog and Wnt family members induce myogenic bHLH gene expression in the somite. *Genes Development,* 9:2911–2922, 1990.

[59]R. Nawrotzki, N. Y. Loh, M. A. Ruegg, K. E. Davies, and D. J. Blake. Characterization of alpha-dystrobrevin in muscle. *Journal of Cell Sciences*, 111(17):2595–2605, 1998.

[60]   S. E. Newey, M. A. Benson, C. P. Ponting, K. E. Davies, and D. J. Blake. Alternative splicing of dystrobrevin regulates the stoichiometry of syntrophin binding to the dystrophin protein complex. *Current Biology*, 10(20):1295–1298, 2000.

[61] S. Neuman, M. Kovalio, D. Yaffe, and U. Nudel. The drosophila homologue of the dystrophin gene – introns containing promoters are the major contributors to the large size of the gene. FEBS Letter 579(24): 5365-5371, 2005.

[62]   T. M. Nguyen, J. M. Ellis, D. R. Love, K. E. Davies, K. C. Gatter, G. Dickson, and G. E. Morris. Localization of DMDL gene-encoded dystrophin-related protein using a panel of nineteen monoclonal antibodies: presence at neuromuscular junctions, in the sarcolemma of dystrophic skeletal muscle, in vascular and other smooth muscles, in the proliferating brain cell lines. *Journal of Cell Biology*, 115:1695–1700, 1991.

[63]   H. Nishio, Y. Takeshima, N. Narita, H. Yanagawa, Y. Suzuki, Y. Ishikawa, R. Minami, H. Nakamura, and M. Matsuo. Identification of a novel first exonin human dystrophin gene and of a new promoter located more than 500 kb upstream of the nearest known promoter. *Journal of Clinical Investigation, 94*(3):1037–1042, 1994.

[64]   C. P. Ordahl and N. Le Douarin. Two myogenic lineages within the developing somite. *Development*, 114:339–353, 1992.

[65]H. B. Peng, N. R. Kramarcy, D. F. Daggett, H. Rauvala, J. R. Hassell, and N. R. Smalheist. The relationship between perlecan and dystroglycan and its implication in the formation of the neuromuscular junction. *Cell Adhesion Communication Journal*, 5(6):475–489, 1998.

[66]   M. F. Peters, M. E. Adams, and S. C. Froehner. Differential association of syntrophin pairs with dystrophin complex. *Journal of Cell Biology*, 138(1):81–93, 1997.

[67]M. F. Peters, N. R. Kramarcy, and S. C. Froehner. Beta 2-syntrophin: localization at the neuromuscular junction in skeletal muscle. *Neuroreport*, 5(13):1577–1580, 1994.

[68]   M. F. Peters, K. F. O'Brien, H. M. Sadoulet-Puccio, L. M. Kunkel, M. F. Adams, and S. C. Froehner. Beta-dystrobrevin, a new member of dystrophin family. Identification, cloning and protein association. *Journal of Biology and Chemistry*, 272(50):31561–31569, 1997.

[69]   M. F. Peters, H. M. Sadoulet-Puccio, M. R. Grady, N. R. Kramarcy, L. M. Kunkel, R. Sealock, and S. C. Froehner. Differential membrane localization and intermolecular

associations of alpha-dystrobrevin isoforms in skeletal muscle. *Journal of Cell Biology,* 142(5):1269–1278, 1998.

[70]   F. Pons, A. Robert, H. Fabbrizio, G. Hugon, J. C. Califano, J. A. Fehrentz, J. Martinez, and D. Mornet. Utrophin localization in normal and dystrophin-deficient heart. *Circulation,* 90:369–374, 1994.

[71] O. Pourquie et al. Lateral and axial signals involved in somite patterning: A role for MBP4.*Cell,* 84:461–471, 1996.

[72]M. E. Pownall, M. K. Gustafsson, and C. P. Emerson. Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annual Review of Cell Development and Biology*, 18:747–783, 2002.

[73] K. Sogawa, H. Imataka, Y. Yamasaki, H. Kusume, H. Abe, and Y. Fuji-Kuriama. cDNA cloning and transcription properties of a novel GC box-binding protein, BTEB2. *Nucleic Acid Research*, 21(7):1527–1532, 1993.

[74] K. Song, P. E. Sherrer, L. Tang Zao, T. Okamoto, M. Cafel, C. Chu, S. D. Kohtz, and P. M. Lisanti. Expression of Caveolin-3 in skeletal, cardiac and smooth Muscle cells. *Journal of Biochemical Chemistry*, 275(25):15161–15165, 1996.

[75] J. Schofield, D. Houzelstein, K. Davies, M. Buckingham, and Y. H. Edward. Expression of the dystrophin related (utrophin) gene during mouse embryogenesis. *Developmental Dynamics,* 198(4):254–264, 1993.

[76]Y. Seko, T. Yamazaki, Y. Shinkai, H. Yagita, K. Okumura, S. Naito, K. Imataka, J. Fujii, and Y. Yazaki. Cellular and molecular bases for the immunopathology of the myocardial cell damage involved in acute viral myocarditis with special reference to dilated cardiomyopathy. *Japanese Circulation Journal,* 56(10):1062-1072,1992.

[77]A. Shainberg, G. Yagil, and D. Yaffe. Control of myogenesis in vitro by Ca2+concentration in nutritional medium. *Experimental Cell Research*, 58:163–167, 1969.

[78]J. Smith and J. Schofield. Stable integration of an mdx skeletal muscle cell line into dystrophic (MDX) skeletal muscle: Evidence for cell stem status. *Cell Growth and Differentiation,* 8:927–934, 1997.

[79]H. M. Sturn, M. C. Brown, and S. D. Hauschka. Myogenesis in paraxial mesoderm: Preferential induction by dorsal neural tube and by cell expressing Wnt-1. *Development,* 121:3675–3686, 1995.

[80]   R. G. Roberts. Dystrophins and dystrobrevins. *Genome Biology*, 2(4):3006.1–3006.7, 2001.

[81] I. N. Rybakova, K. J. Amann, and J. M. Evrasti. A new model for the interaction of dystrophin with F-actin. *Journal of Cell Biology*, 135(3):661–672, 1996.

[82] S. Tajbakhsh, D. Rocauncourt, G. Gossu, and M. Buckingham. Redefining the genetic hierarchies controlling skeletal myogenesis: Pax3 and Myf5 act upstream of MyoD. *Cell*, 89:127–138, 1997.

[83] J. M. Tinsley, A. C. Potter, S. R. Phelps, R. Fisher, J. I. Trickett, and K. E. Davies. Amelioration of the dystrophic phenotype of mdx mice using a truncated utrophin transgene. *Nature,* 384:349–353, 1996.

[84] S. J. Venters, S. Thorsteindottir, and M. J. Duxon. Early development of the myotome in the mouse. *Develelopmental Dynamics* 216:219–231, 1999.

[85] M. Yoshida, H. Hama, M. Ishikawa-Sakurai, M. Imamura, Y. Mizuno, K. Araishi, E. Wakabayashi-Takai, S. Noguchi, T. Sasaoka, and E. Ozawa. Biochemical evidence for association of dystrobrevin with the sarcoglycan-sarcospan complex as a basis for understanding sarcoglycanopathy. *Human Molecular Genetics,* 9(7):1033-1040, 2000.

[86] T. Yoshida, Y. Pan, H. Hanada, Y. Iwata, and M. Shigekawa. Bidirectional signalling between sarcoglycans and integrin adhesion system in cultured L6 myocytes. *Journal of Biology and Chemistry*, 273(3):1583–1590, 1998.

POST-TRANSCRIPTIONAL MODIFICATIONS OF MRNA

by

CHAVDAR IVANOV

A thesis submitted to
The University of Birmingham for
the degree of
MASTER in Philosophy

School of Biosciences
College of Life and Environmental Sciences
The University of Birmingham
August 2019

Abstract

Post-transcriptional modification and regulation of mRNA plays a crucial role in the cellular cycle of an emerging organism's cellular division and growth. Alternative splicing and post-transcriptional modification of mRNA are responsible for adding a new layer to existing control mechanisms in cell metabolism. Post-transcriptional modification of mRNA includes A-to-I and C-to-U editing, as well as m^5C, m^6A, Cap1 and Cap2 modifications. mRNA editing causes a number of pathologies which emerge after modifications to the mRNA, but post- transcriptional remodeling increases stability of the mature mRNA in the cellular environment. The aim of this paper is to investigate and observe the correlations that arise from the intimate interaction of NSUN6 protein with other proteins connected to sexual differentiation in *Drosophila melanogaster* as well as the tight integration of its mRNA with other proteins. NSUN6 methyltransferase is of particular interest, as it is highly conserved among eukaryotes and belongs to the family of Nol1/Nol2/SUN domains (NSUN proteins). Applied UV crosslinking as well as chemical crosslinking with 5-azacytidine have revealed the function of this enzyme. NSUN6 exists in the cytoplasm and participates in methylation of the cytosine of mRNA. Understanding the mechanism and effect of this methylation can help understand its role in gene expression and has implications for a variety of pathological processes that may develop as a result.

Chapter 1

NSUN6 and Its Biological Role in the Cell Cycle

1.1 Introduction

RNA contributes to cell metabolism and regulates a variety of functions in the cell. The diversity of RNA functions in the cell requires a variety of RNA molecules. Transcription is a process that creates mRNA from DNA, which in turn is responsible for production of proteins from protein-coding genes. mRNA is also referred to as coding RNA. mRNA is responsible for translation together with the ribosome and translational complex, which possesses trans- and cis elements. The existence of another major group of RNAs—so-called non-coding RNA—is conditioned by the need to regulate other cellular processes, such as participating in generation of ribosomal rbRNA and transfer RNA (tRNA) as an integral part of translation. Other types of RNA, including miRNA and siRNA, form part of the machinery which suppresses gene expression, and which participates in gene silencing. In addition to these types of RNA, small nucleolar RNA (snRNA) is responsible for modification and processing of other non-coding RNAs as well as alternative splicing of coding pre-mRNA.

snRNAs that are U-rich also form part of the spliceosome machinery, and most spliceosomes contain the U1, U2, U4, U5 and U6 varieties of snRNAs. Long non-coding RNAs participate in dosage compensation in sexually developed individuals, and finally RNA can be part of riboswitches and RNA sensors.

Two major forces are responsible for driving the evolution and maintaining the variety of proteins needed for correct cellular function. These are post-transcriptional

modification of mRNA and alternative splicing. They preserve and keep in check the stability of expression of genes, guarantee smooth transmission of genetic information to the subsequent generation, and sustain the variety of proteins. To date, more than 140 post-transcriptional mRNA modifications have been discovered, all of which contribute to the process of correct transmission of genetic information.

RNA modifications can be divided into internal modifications and RNA editing. RNA internal modifications are m^6A-methylation of the adenosine base at the sixth position in the mRNA; m^5C methylation of cytidine at the fifth position; cap1 and cap2 methylation of the cap at its 5' end with a 7-methylguanosine nucleotide. RNA editing consists of A-to-I (adenosine-to-inosine) and C-to-U (cytosine-to-uracil) editing and the conversion of uridine to pseudo-uridine (Ψ).

1.2 RNA Splicing

RNA splicing is a process that leads to emergence of mature RNA. Eukaryotic cells contain group III introns which are present in many protein-coding mRNAs. These introns are not self-splicing and require the assembly and function of a large protein-RNA machine called a spliceosome (Meister 2011). Group I and Group II introns have also been documented. In all three groups of introns, there is a common mechanism that ensures the continuity of the evolutionary process. The common mechanism is the transesterification reaction with the nucleophilic attack mainly carried out by OH^- group on the phosphate backbone of the RNA. This process is energy-independent (Meister

2011).

Ubiquitous sequence elements are necessary for adequate pre-mRNA splicing. The 5' splice site of the eukaryotic cell contains a conserved GU element followed by a purine and AGU element. The 3' splice site is constructed of an AG element. There is a branch point sequence which contains the branch point adenosine (A) and polypyrimidine tract, located between the branch point in addition to the 3' splice site. These elements are highly conserved and are called *cis* elements.

Additional elements are required for correct processing of pre-mRNA splicing. Such elements are found in introns as well as exons, and they can influence splicing positively or negatively. They are called intronic splicing enhancers (ISE), intronic splicing silencers (ISS), exonic splicing enhancers (ESE), and exonic splicing silencers (ESS) (Meister 2011). Figure 1. represents schematically conserved intronic *cis* elements participating as guides for assembling of spliceosomes.

exon   5' splice site   branch point   polypyrimidine   3' splice site exon
                                                              tract

GURAGU —————————— YNYURAY——————— Y (10-12)———— AG
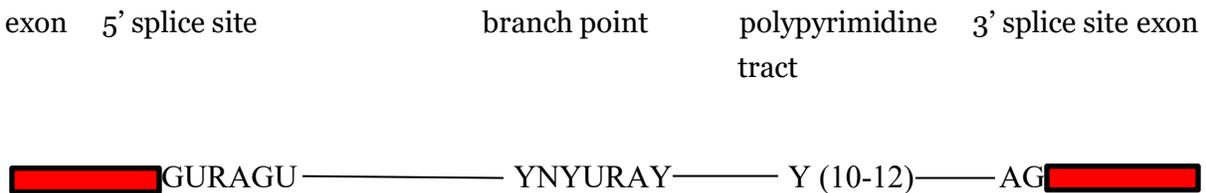
Figure 1. *Conserved intronic elements responsible for pre-mRNA splicing.* Eukaryotic cells have both a 5' splice site consisting of GURAGU, where R is a purine, and a 3' splice site consisting of AG. The branch point element contains an A base that is crucial for the first catalytic reaction of the splicing. The branch point harbors YNYURAY in its structure, where Y is a pyrimidine and R is a purine, and N can be any nucleotides. The polypyrimidine tract is located between the branch point and the 3' splice site (Meister 2011).

1.2.1 The Spliceosome

Another element that is essential for correct splicing is the spliceosome. The spliceosome is highly effective and is capable of changing its structure during pre-mRNA splicing. The spliceosome contains RNA-protein complexes that were formed prior to starting the splicing. The RNA-protein particles are composed of a non-coding U-rich small nuclear RNA that binds tightly to specific proteins to form so-called small ribonucleoprotein particles (U snRNPs) (RNA Biology, Meister G, 2011). The U snRNPs can be grouped in accordance with the non-coding RNA part of the spliceosome as follows: U1 snRNP, U2 snRNP, U4 snRNP, U5 snRNP and U6 snRNP, which are all necessary for splicing the pre-mRNA and are involved in forming of the major spliceosome. A U12-dependent minor spliceosome also exists and is responsible for the splicing of mRNA with U12 type introns.

U12 type introns contain a 5' splice site composed of A (or G) UAUCCUUU and the branch point element, which consists of the UCCUUA*A*CU sequence and which lacks a polypyrimidine tract. The U12-dependent minor spliceosome consists of U11, U12, U4atack and U6atack snRNPs (Meister 2011). All these components are referred to as *trans*-elements and are responsible for providing the correct platform for precise pre- mRNA splicing.

How do spliceosomes work? A concise explanation is needed before proceeding to the next step. Transcription is a process where pre-mRNA is synthesized from protein- coding genes with the help of RNA Polymerase II. This process can be divided into transcriptional initiation, elongation, and termination. In each specific step, different proteins are required, which combine with the main enzyme RNA Polymerase II. This process leads to production of pre-mRNA. Pre-mRNA consists of exons and introns

which require further processing to become mature mRNA.

The mechanisms which are involved in transforming the pre-mRNA into mature mRNA are as follows: capping of the pre-mRNA 5' end, as well as 3' end processing which includes the addition of a poly (A) tail. Capping of the pre-mRNA will be presented in additional detail later as a part of post-transcriptional modifications of mRNA.

The spliceosome is assembled through recruitment of the U1 snRNP to the 5' splice site, which precedes the formation of the E complex. U2 snRNP binds to the branch point shortly after the U1 snRNP binds to the 5' spice site. This leads to formation of pre-spliceosome complex A. The pre-assembled U4/U5 and U6 tri-snRNP complex is attached to complex A to form the precatalytic spliceosome or complex B. U1 and U4 snRNPs are displaced, and the active spliceosome is formed. This complex is called B^*.
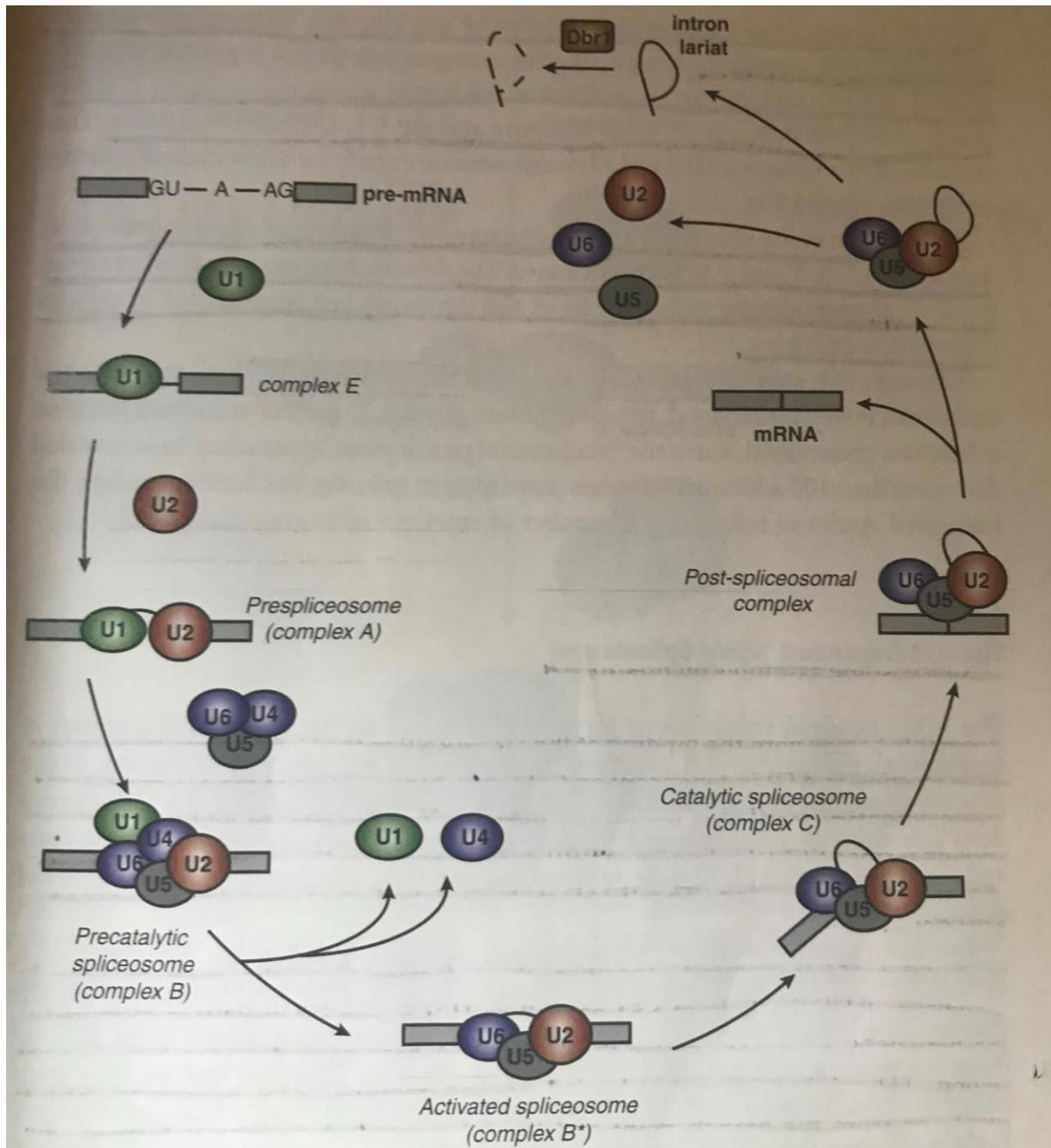
F*igure 2.* Schematic showing the process of splicing of pre-mRNA. In the first step (1), U1 snRNP converges to the 5' splice site where it binds (2). The reaction requires a full base pair of U1 snRNP with a *complementary* sequence within the 5' splice site. Simultaneously, another protein called SF1, also known as BBP, *branch point binding protein* (BBP), attaches to the branch point region. The U2AF (U2 auxiliary factor) complex consists of U2AF65 and U2AF35. U2AF65 binds to the SF1 protein, while U2AF35 binds to the AG in the 3' splice site. All of these factors then combine to form *spliceosome E complex* (2). The SF1 and U2AF help to recruit U2 snRNP to the branch point. This reaction is energy-dependent and requires ATP hydrolysis. During this joining process of the U2 snRNP at the branch point, the SF1 protein is pushed out from the branch point and helps create the *spliceosome A complex* (3). The next step is formation of the tri-snRNP complex, which is composed of U4, U6 and U5 snRNPs. In this complex, U4 and U6 are base paired, while U5 is attached by protein-protein interaction. This complex is important because it forms part of the new *precatalytic spliceosome (complex B)* (4). To be activated, the spliceosome must be subjected to rearrangement and displacement of the existing components of the *precatalytic spliceosome (complex B),* which

74

includes release of U1 and U4 snRNPs, which in turn leads to creation of the active *spliceosome B*complex* (5). The active spliceosome then binds to the 5' splice site, after which occurs the release of the 5' exon and formation of a covalent bond at the 5' end of the intron to the branch point. This complex is called *spliceosome complex C* (6). Further rearrangements form the *post-spliceosome complex,* which contains U6, U5, and U2 snRNPs (7). (adopted from Meister 2011)

The next step is disassembling of the *post-spliceosome complex* into the U6, U5, and U2 snRNPs. The U2 snRNP retains the intron lariat and splices mature mRNA (8). The last step is the complete dissolution of the remains of the *post-spliceosome complex* (9). The intron lariat is further processed by the debranching enzyme Dbr1, which is the prerequisite for the complete degradation of the intron lariat by 5' to 3' and 3' to 5' exonucleases (Meister 2011).

The process of splicing pre-mRNA is coupled with transcription and 5' capping. Although the mature mRNA is in the beginning stage of completion, the first few introns are spliced out co-transcriptionally during this synthesis. Construction of a correct cap structure is a necessary precursor to accurate splicing of the pre-mRNA. Formation of the cap-binding complex (CBP) enhances interaction of the U1 snRNP with the 5' splice site, which is necessary for later binding of the U6 snRNP and displacement of the U1 snRNP.

The C-terminal domain (CTD) of an RNA Pol II is a binding platform for factors that directly influence transcription and splicing. Phosphorylation of CTD is among the important regulators of the splicing, together with the so-called SR proteins which are rich in serine (S) and arginine (R) and which are used as a platform that promotes spliceosome assembly (Meister 2011). Newly synthesized mature mRNA consists of elements mentioned above—the exonic splicing enhancers and silencers (ESE, ESS) and the intronic splicing enhancers and silencers (ISE, ISS) that play a principal role in alternative splicing.

1.3 Alternative splicing

Alternative splicing is the process responsible for adding a new layer of complexity in increasing the variety of proteins that a cell can produce. Mechanisms of alternative splicing can be illustrated as follows:
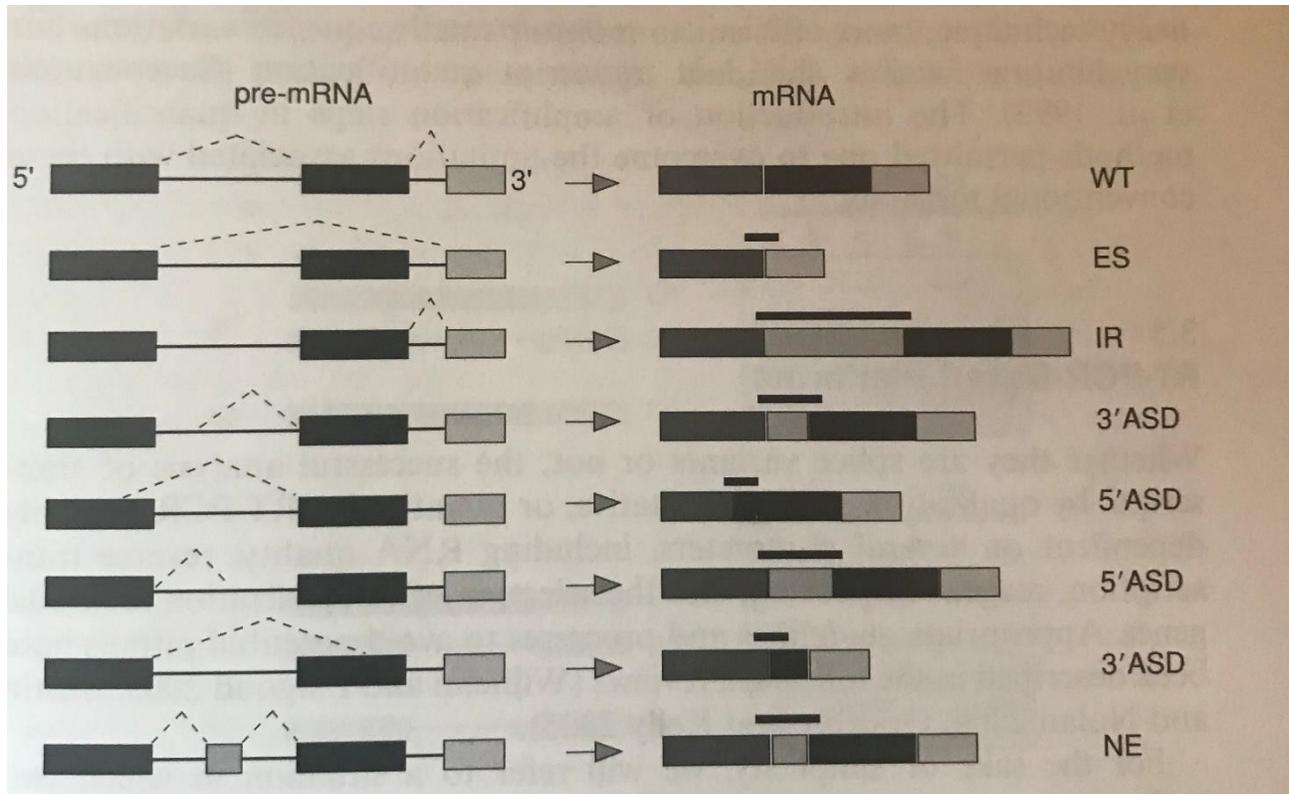
Figure 3. Types of alternative splicing and their influence on the processing of pre-mRNA. The first represents WT (wild type) processing of the pre-mRNA. Many of the exon splicing enhancers (ESE) are SR proteins. The major function of SR proteins is recruitment of U2AF65 to the weak polypyrimidine tract. If SR proteins are not attached to the site of the alternative splicing, the following exon will be excluded from the mature mRNA, as shown on the second line of the figure. SR proteins which bind to intron splicing enhancers (ISE) influence binding of U1 snRNP to the weak 5' splice site. This is the case of intron inclusion (line 3). Without a binding SR protein, the intron will not be recognized and is skipped, which is the normal process in splicing of pre-mRNA. 5'and 3' alternative splice donors (ASD) use cryptic 5' and 3' sites, which generate additional sequence elements which are not present in the WT and include a novel exon (NE) (line 8) (adopted from Bracco, 2006).

Alternative splicing leads to increased variability of the proteins in the cell including through utilizing the weak polypyrimidine tract as well as the 5' and 3' splice sites. Alternative splicing is aided by exon splicing enhancers and silencers as well as intron splicing enhancers and silencers, which either increase or decrease the affinity of transcription factors for the weak polypyrimidine tract.

1.3.1 SXL

SXL is the master gene in the hierarchy of an intricate and elaborate system developed in sex differentiation in *Drosophila Melanogaster*. Splicing silencers (ESS or ISS) are sequences that serve as binding platforms for different proteins such as PTB (polypyrimidine tract-binding protein), which belongs to the hnRNP family. SXL belongs to a family named hnRNP-A1. SXL is a main regulatory protein that controls and channels the sex alternative splicing in *Drosophila melanogaster*.

Female sex determination in *Drosophila melanogaster* is dependent on the presence of two X chromosomes. If in the early stage of development an XY genotype is established, SXL protein is not made. This leads to the production of a *Sxl* transcript which includes the third exon with a stop codon. As a result, non-coding RNA is made (Kelley and Kuroda 2000; Salz 2011). The result is production of truncated and inactive protein.

The SXL protein is expressed only in female flies. When the SXL protein is present, a female splicing pattern occurs, and the male splicing pattern is suppressed (D.B. 2003). In the early stages of embryonic development, a chain of autoregulation of *Sxl* is established based on presence of two X chromosomes. In the early stages of embryogenesis, more specifically in the syncytial blastoderm, in response to the activities of the two X chromosomes, the *SxlPe* promotor is activated. This promotor has a short

term of activity and prepares the environment for further activation of *SxpPm* which will later be responsible for initiating exon skipping (Salz 2011).
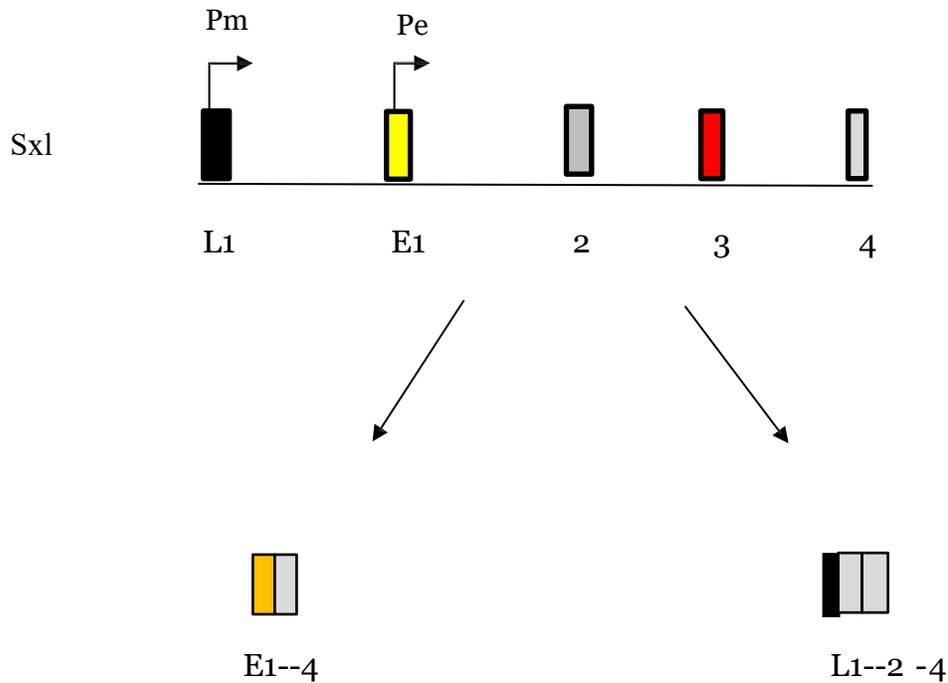
Figure 4. Early embryo-specific exon is represented in yellow. It is still unclear how E1 joins to exon 4, but in the early syncytial blastoderm, *SxlPe* is activated in embryos with two X chromosomes. This early promotor produces the SXL protein that further influences the late *SxlPm* promotor, which then initiates skipping of exon 3 in female individuals (Salz 2011).

SXL has two RNA binding domains (RBD). The mechanism of binding between the targeted RNA and the SXL protein was described as a result of crystallography obtained in the laboratory of Nagai (Varani and Nagai 1998). These results reveal that the SXL protein binds to RNA through both RBD structures, which face one other and form a cleft, which then interacts with a nine-nucleotide U-rich element interrupted by guanosines (Crowder, Kanaar et al. 1999). The N-terminal domain of the RBD contains a glycine-rich sequence which assists in the assembly of a larger protein complex with multiple binding sites (Wang and Bell 1994). The *Sxl* pre-mRNA consists of several SXL binding sequences. The two most important are located approximately 200 nucleotides upstream

and downstream from the male exon (Salz 011).

Alternative splicing of the *sxl* gene is considered crucial for determination of sex in *Drosophila melanogaster*. Spliceosome assembly is dependent on the presence of the SXL protein. This process can occur in two different ways depending on whether SXL protein is present. In females, the SXL protein interferes with the assembly of the spliceosome, thereby interacting with U1 snRNP at the 5' splicing site and with the U2AF65 at the 3' splicing site. This interaction precipitates formation of the dead-end complex and does not allow the U2 snRNP to bind to the forming spliceosome. This is the reason for exon 3 exclusion in the RNA active transcript. In the absence of a second X chromosome, the SXL protein is not produced. This leads to assembly of the spliceosome through the binding of U1 snRNP to the 5' splicing site and simultaneously with the binding of U2AF65 to AG of the 3' splicing site, which leads later to the binding of U2 snRNP to the spliceosome. This allows the formation of the full spliceosome, which leads to exon 3 inclusion, presenting UAG in the center of exon 3, and results in development of the male phenotype.

Sxl splicing regulation in *Drosophila Melanogaster*
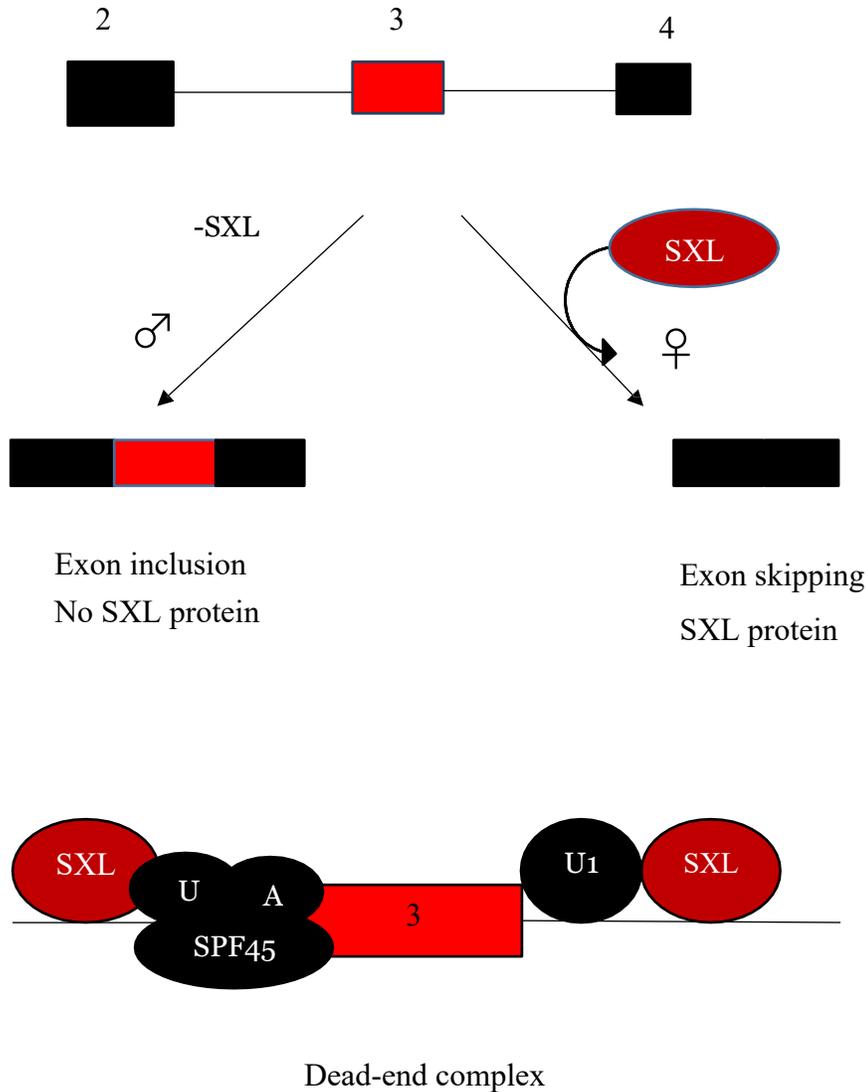
Figure 5. Schematic representation of the *sxl* gene that gives rise to the female phenotype. The spliceosome starts assembling around exon 3 with U1 snRNP binding to the 5' splicing site and U2AF65 binding to the 3' splicing site. SXL protein binds to the U1 snRNP and to U2AF65 and triggers exon 3 skipping, which in turn leads to development of the female phenotype. Without the SXL protein, the spliceosome assembles around exon 3, which then leads to the production of non-coding RNA, which leads to development of the male phenotype (D.B. 2003).

1.3.1.0 SXL autoregulation

SXL is a main player in the system that determines the sexual differentiation in *Drosophila melanogaster* and auto-regulates its own expression by promoting skipping of exon 3 from its pre-mRNA (Bell, Horabin et al. 1991). The regulation is elaborate and includes 5' and 3' UTR regions of exon 3, as well as 3' UTR of intron 2 and 5' UTR of intron 4. Multiple uridine-rich segments, referred to as upstream and downstream, have been identified in the introns that precede and follow exon3, respectively (Sakamoto, Inoue et al. 1992). The downstream uridine-rich sequence is preferential, forming stronger bonds with the SXL protein (Horabin and Schedl 1993). The upstream uridine- rich sequence and the SXL protein play an important role in determining whether exon 3 is skipped or retained. The 5' UTR of exon 3, which is very weak, is a particular target for SXL regulation (Horabin and Schedl 1993), but the 3' splice site is important as well, given its very strong affinity for and ability to bind strongly with SXL. Deleting the 3' splicing site would automatically lead to making 5' UTR the principal binding site. The opposite is also true-- mutation in the 5' UTR leads to SXL dysregulation.

How does the mechanism of skipping exon 3 work? In *Drosophila melanogaster* a homolog to U1 and U2 snRNP named SNF protein was identified. This protein interacts with SXL, leading to binding of SXL to U1 and U2 snRNP at the 5' and 3' splicing sites on exon 3. This turns off both binding sites, blocking the assembly of the spliceosome. (Deshpande, Samuels et al. 1996).
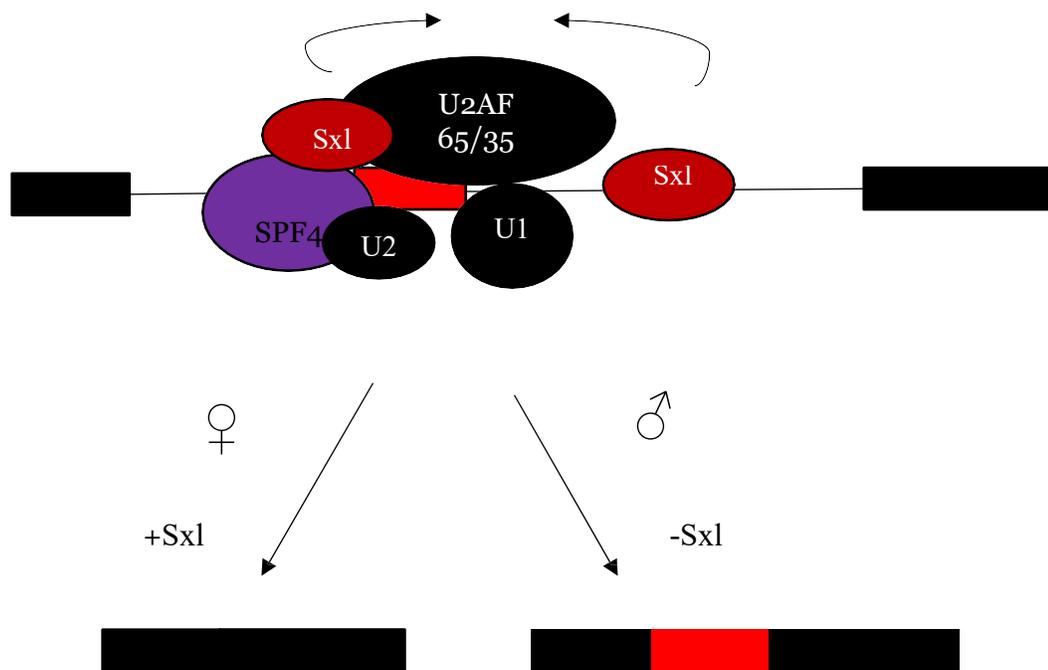
Figure 6. Autoregulation of the *Sex lethal (sxl)* gene *(D.B. 2003)*. The interaction that occurs between SXL, SPF45 and U2AF65 that leads to changes of autoregulation of the SXL protein is clearly visible.

The SXL protein regulates the transcription of the *Transformer* gene (Tra). SXL, as mention above, is produced only by females, which leads to splicing of *Transformer* to produce female-specific Tra-protein that controls the splicing of the *Doublesex (dsx)* and *Fruitless (fru)* genes. *Transformer* (tra) is transcribed in female and male flies but TRA protein uses two different alternative 3' splice sites that exist in intron 2, differentiating the splicing site used in female and male flies (Boggs, Gregor et al. 1987). "The male uses the upstream 3' splice site and contains a STOP codon in its sequence, which gives truncated and inactive TRA protein, while the female uses both upstream and more often downstream splice sites." (Sosnowski, Belote et al. 1989). The splicing sites can be divided as follows: an upstream site called "non-sex-specific" where SXL protein binds to the pyrimidine tract and blocks U2AF65 from making bonds with the pyrimidine tract, in order to allow emergence of the female phenotype using the female-specific splice site (Sosnowski, Belote et al. 1989). In the absence of SXL protein, two reasons can be highlighted to explain the unequivocal use of a non-specific sex site. The U2AF65 has a 100-fold higher affinity for the uridine-rich pyrimidine tract of the non-specific 3' splicing site than for the less uridine-rich pyrimidine tract associated with the female-rich specific site (Valcarcel, Singh et al. 1993). Additionally, the non-sex-specific site is closer to the 5' splice site, which explains the more frequent use of this site when the SXL protein is absent.

1.3.2 TRA

TRA is regulated by SXL and controls the splicing of the *Doublesex (dsx)* and *Fruitless (fru)* genes and consists of an extended RS domain. TRA does not contain an RNA binding domain, but it essential, as part of the assembly of the spliceosome, for further splicing of *dsx* mRNA and *fru* mRNA.
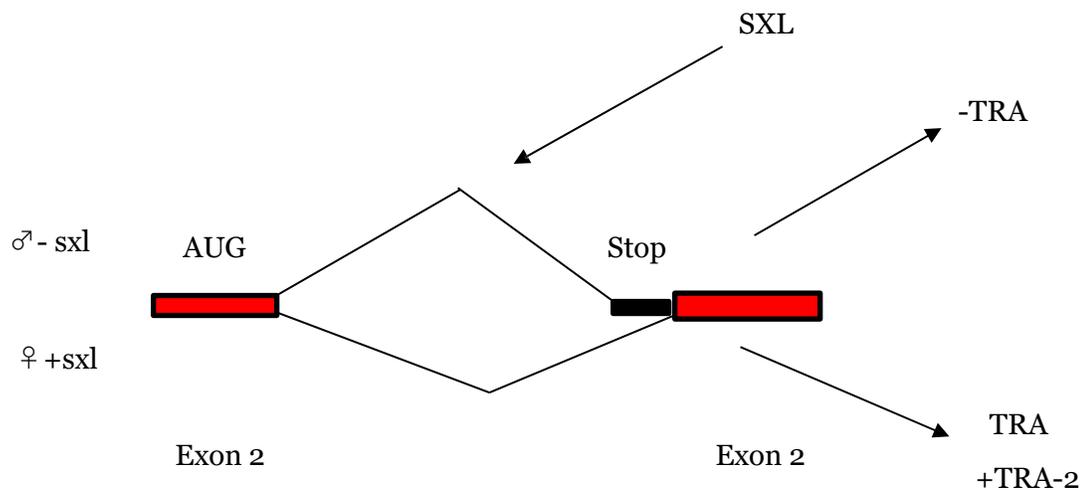
Figure 7. The alternative 3' splice site is used to make different RNA transcripts. The SXL protein inhibits use of the 3' splice site in *tra,* which leads to skipping the stop codon in the frame and which produces a TRA protein. TRA acts with TRA-2, expressed in both sexes, and induces use of a female-specific downstream 3' splice site. In the presence of SXL, the protein binds to its recognition motif, which is located at the downstream 3' splice site, blocking the recognition and consequent binding of the U2AF65 (Sosnowski, Belote et al. 1989). These changes lead to a shift in the downstream 3' splice site where the stop codon is deleted from the *Tra* mRNA, which further moves to active translation of the *Tra* mRNA and produces the active TRA protein (D.B. 2003).
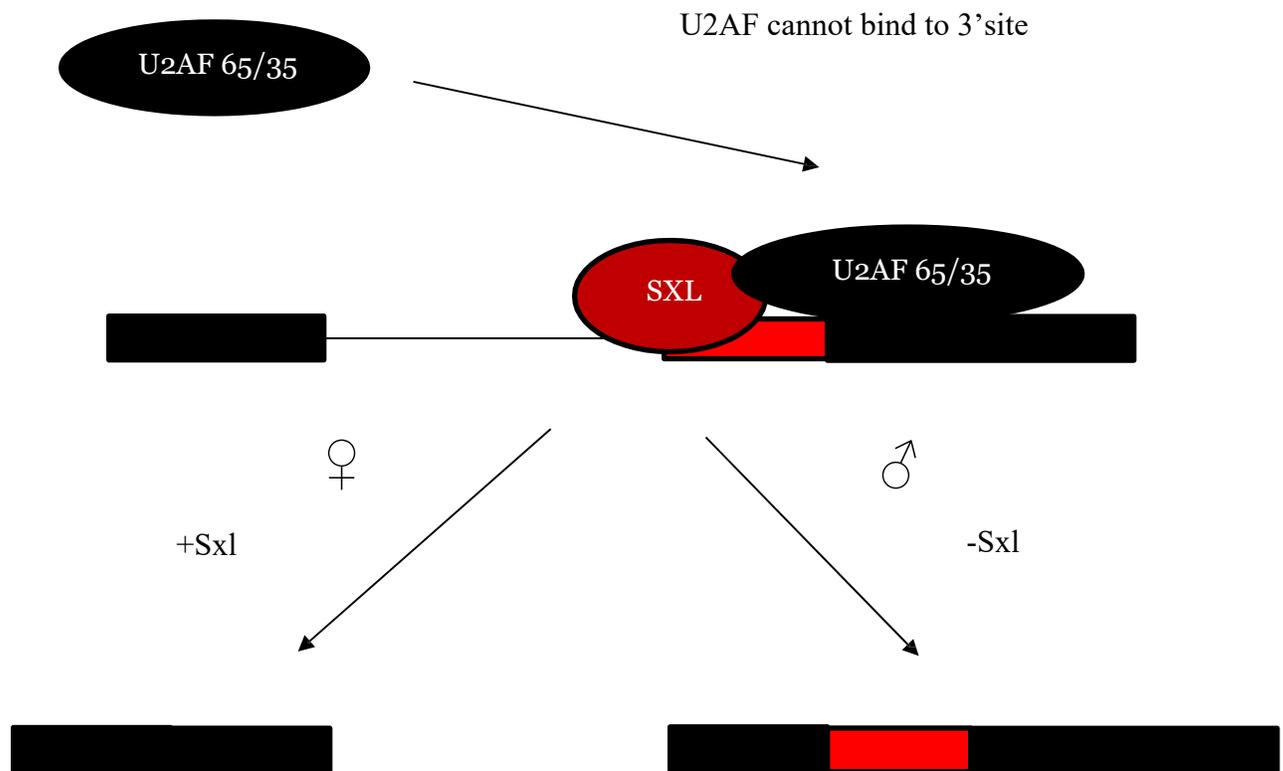
Figure 8. Regulation of the splicing in *Transformer (Tra) (D.B. 2003)*.

The *Transformer* gene produces the TRA protein, which is present in female flies. TRA is a positive regulator of splicing that activates the female-specific pattern, compared to SXL protein (Lopez 1995). In male flies, TRA is not produced—only TRA-2, which triggers male-specific splicing of *dsx* and *fru* transcripts (D.B. 2003). In female flies, TRA proteins accumulate and trigger female-specific splicing of *dsx* and *fru* transcripts as well. The *doublesex (dsx)* gene is alternatively spliced in a sex-specific manner and produces transcripts that differ in their carboxyl terminal domains. The female transcript has a shorter COOH domain, while the male transcript has a longer COOH domain.

TRA-2, which is an SR-related protein, regulates *dsx* and *fru* but also regulates its own splicing in the male germ lines. Splicing of the *tra* gene by its own protein produces four different forms, one of which plays a crucial role in spermatogenesis and which is named TRA-2[226], which suppresses production of *tra-2* mRNA in the M1 intron (Mattox, McGuffin et al. 1996). The start codon of TRA-2[226] is a part of the 3' splice site of the M1 intron. After 3' processing of the M1 intron, splicing of the M1 intron occurs but without expression of TRA-2[226].

The TRA-2[226] protein performs an especially important negative feedback on regulation of *Tra* mRNA. In the presence of TRA-2[226], the intron M1 is suppressed, which is located between exon 3 and exon 4. Exon 4 contains an AUG codon, as a result of which the long form of protein is produced, but if intron M1 is spliced, the start codon AUG is erased, and as a result, a short-truncated form with 179 amino acids is produced (Mattox and McGuffin et al. 1996).
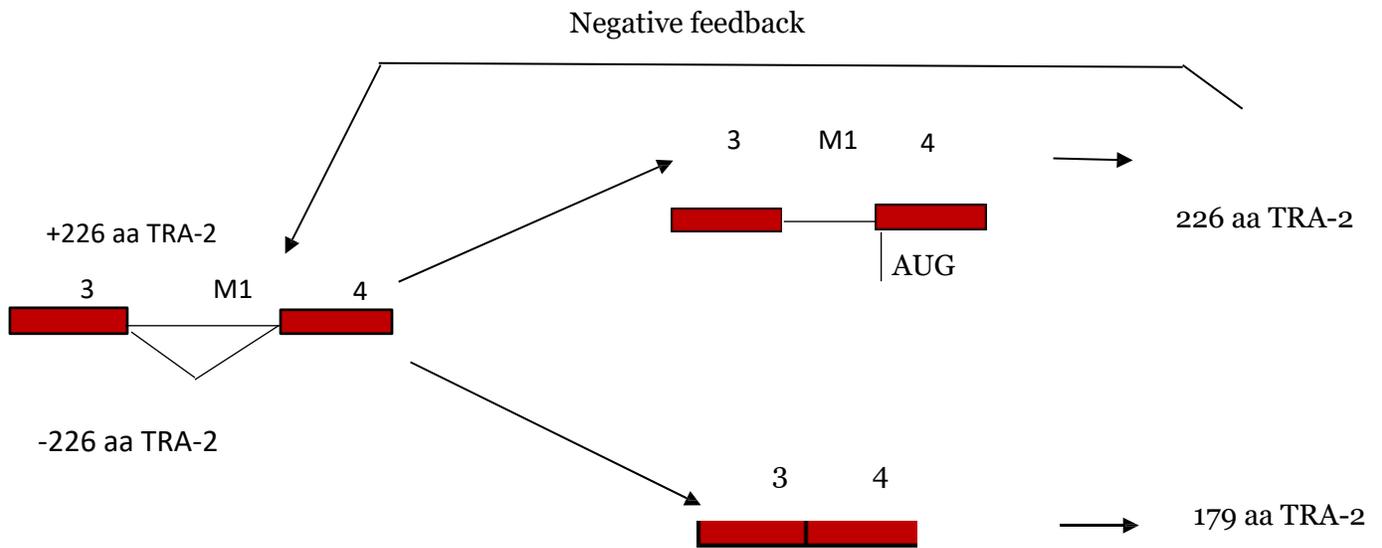
Figure 9. Alternatively, spliced *Tra* pre-mRNA in the germ lines of male individuals produces TRA-2^{226} protein. Through negative feedback, *Tra* pre-mRNA is spliced, and 226 aa TRA-2 protein is formed. This isoform suppresses splicing of the M1 intron, which is located between exon 3 and 4, and results in retention of the AUG codon on exon 4, which is in frame.
Consequently, splicing of the M1 intron erases the start codon for the 226-amino acid isoform, which is not produced. The downstream initiator generates the 179-amino acid form of TRA-2 (Mattox, McGuffin et al. 1996).

Both TRA and TRA-2 harbor RS-rich domains. They belong to a large family of proteins, SR proteins (Manley and Tacke 1996). TRA-2 has two RS domains, one responsible for protein-protein interaction as well as RNA-protein interaction. The other domain is not essential. SR-proteins bind to the purine-rich exonic enhancer and promote use of the weak 3' splice site (Xu, Teng et al. 1993).

1.3.3 DSX

Dsx splicing is based on strengthening of a weak female-specific 3' splice site in exon 3 (Hoshijima, Inoue et al. 1991). TRA and TRA- 2 are made only in female flies. TRA and TRA-2 work together to achieved female splicing, and both proteins bind to exonic sequences which have six 13-nucleotides and a purine-rich element (Hedley and Maniatis 1991). TRA-2 binds to the 3' end of the repeat and, with the help of TRA, which lacks an RNA binding domain, stabilizes interaction of an SR-protein named RBP1 with the 5' end of the repeat. Another group of SR protein exists that binds to the purine-rich element (Tian and Maniatis 1993). The *dsx* enhancer is located at a significant distance from the female-specific site, while the purine-rich enhancer is downstream from exon 3 and regulates the 3' splice site. The purine-rich element is located close to the 3' splice site and can operate independently from TRA and TRA-2 (Tian and Maniatis 1994). Regulation is achieved by making the 3' splice site dependent on TRA and TRA- 2, which helps in the recruitment of SR-protein to the site which then acts at a distance to stimulate splicing. In male flies, TRA is not produced; consequently, this leads to splicing exclusion of exon 4 as well as intron skipping between exon 5 and exon 6 in the mature mRNA of the *dsx* gene and the emergence of male-specific differentiation. In female flies TRA is present, which leads only to intron exclusion between exon 3 and exon 4. The alternative splicing of *dsx* is based on the recognition of a very weak 3' splice site on exon 1. This splice site cannot bind to U2AF splicing factors. Alternative splicing requires efficient splicing machinery to proceed. Exon 4 has evolved as a splicing enhancer and is under the control of TRA protein. Exon 4 has six 13-nucleotide sequences plus a purine-rich sequel. The pre-splicing machinery is composed of three proteins, two of which bind to the repeat element: an SR protein named RBP1, and an SR-related protein TRA-2. TRA binds to TRA-2 and stabilizes the whole complex, which is erected on exon 4 (Lynch and Maniatis 1996)

Figure 10. shows the interaction between TRA, TRA-2 and *Dsx* pre-mRNA.
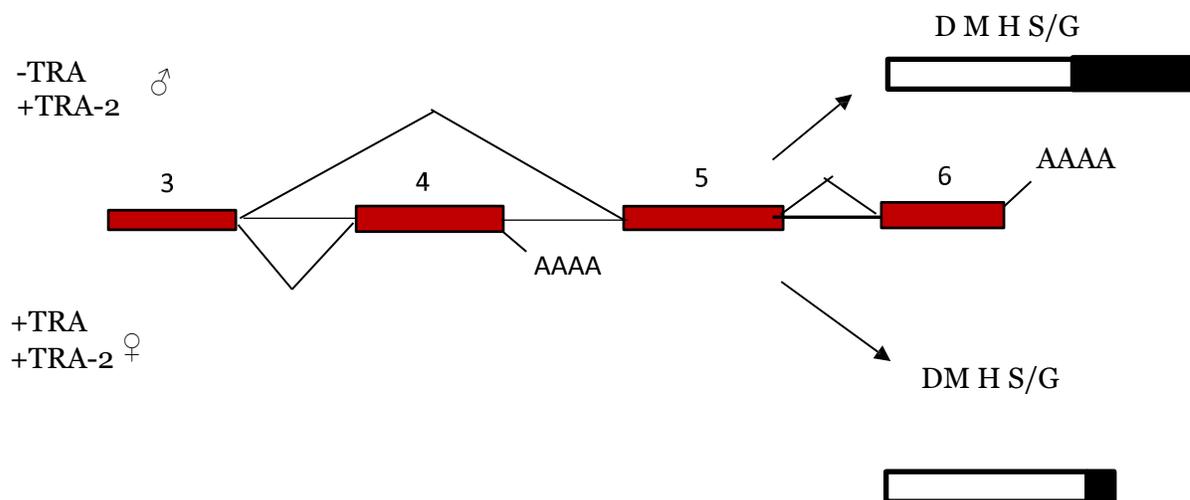
Figure 10. Alternative splicing of the *dsx* gene. The *dsx* gene is under the control of TRA and TRA-2 proteins. In male flies, only TRA-2 exists, which is insufficient to activate the female-specific alternative splicing pattern. (Tians and Maniatis 1994)

Another complex is assembled at the purine rich-element with the help of another SR-protein named dSRp30, along with TRA and TRA-2. TRA-2 has an important function because it mediates the assembly of two complexes through the interaction with TRA and paves the path for the assembly of the splicing machinery on exon 4. The dSRp30 protein stabilizes the bonds between U2AF and the weak 3' splicing site and leads to the alternative splicing of *dsx* (Lynch and Maniatis 1996).
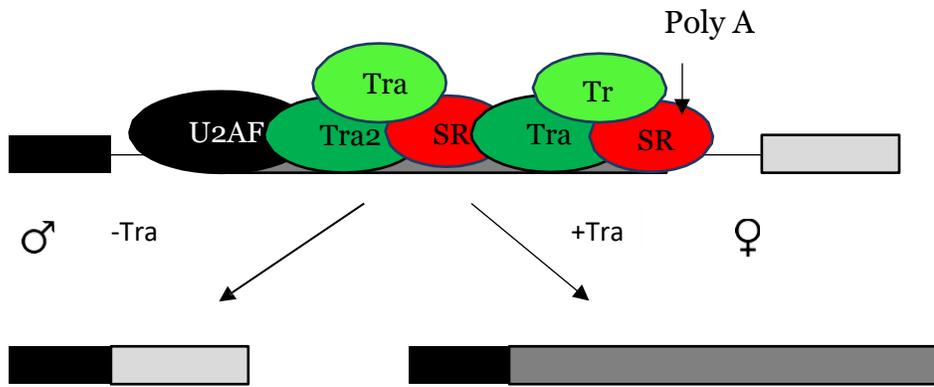
Figure 11. Alternative splicing of *Dsx* pre-mRNA. The spliceosome is assembled on exon 4, which has weak 3' splice sites, but before the spliceosome can be assembled, some steps are necessary to strengthen the bond between the splicing factor U2AF and the 3' splicing site. The reinforcement of the weak bond is achieved by an enhancer located in exon 4 and which consists of six 13-nucleotide long elements, and a purine-rich element. All of these elements can be regulated by TRA-2/TRA proteins with the assistance of an SR-protein named RBP1. This complex binds to the six 13-nucleoide elements and another compound that is formed to support the weak connection between the enhancer and TRA-2/TRA with the help of another SR-protein named dSRp30, which binds to the purine-rich element. TRA-2 protein is the connecting bridge between these two complexes (Lynch and Maniatis 1996).

1.3.4 FRU

The *fru* gene is a regulator of sexual behavior in *Drosophila melanogaster*. The TRA-2/TRA pair of proteins regulates the *fru* gene, and the FRU protein is a transcription factor containing BR-C, ttk, bab (BTB) and a zinc-finger domain. In male flies, the FRU protein determines male-to-male courtship behavior and is the main factor for developing the male-specific muscle of Lawrence (MOL). The absence of FRU leads to loss of function in *fruitless*, which leads to malformation of the male-specific muscle of Lawrence. In addition to the MOL, the gene is expressed in a limited number of neurons in the CNS. The *fru* gene contains four promotors, and promotor P1 is alternatively spliced in a sex-specific manner.
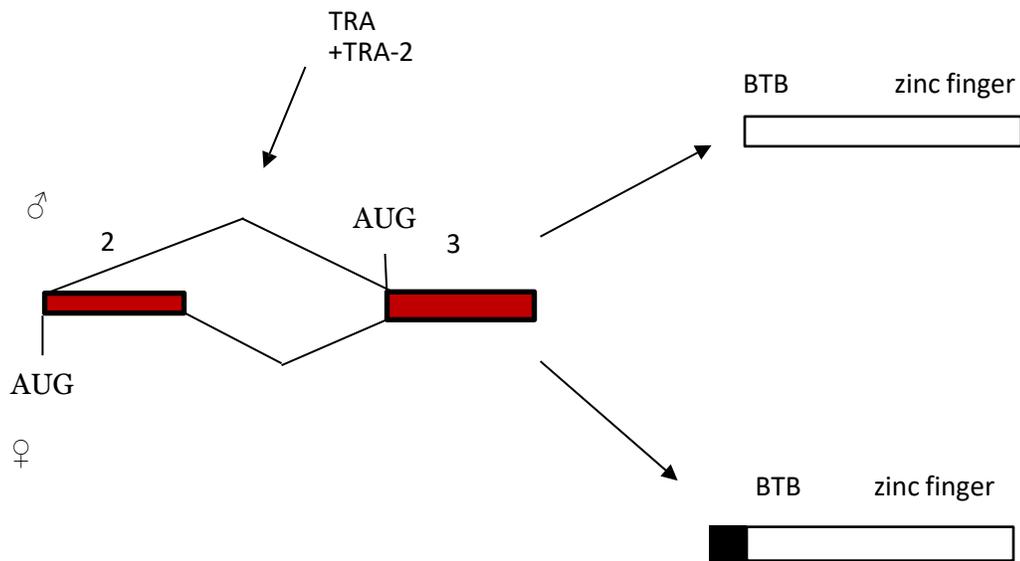
Figure 12. Alternative sex-specific splicing of the *fruitless* gene in *Drosophila melanogaster*. (adopted from Rynar, Goodwin 1996)

It has been confirmed that two 5' splice sites exist on *fru*, and these are crucial in determining the sex-specific alternative splice of the *fruitless* gene. These sites appear to be 1.5 kb apart. (Ryner, Goodwin et al. 1996). These two 5' splice sites engage in a contest for a common 3' splice site (Ryner, Goodwin et al. 1996).

The 5' splice site is located 70 kb apart from the 3' splice site. In females, splicing is activated by tandem TRA-2/TRA and follows the pattern of the *dsx* female-specific mechanism by engaging the 3' splice site in addition to the downstream 5' splice site with help from the RBP1 protein. Several copies of the repeat element are necessary for the alternative splicing which results in development of the female phenotype and development of female genitalia. In females, all repeats that are activated are found to be located upstream of the activated 5' splice site. TRA-2/TRA and RBP1 act through exonic

enhancer, which stimulates either the 5' splice site or 3' splice site.

The upstream 5' splice site is used for activating the male-specific splicing pattern and leads to development of the Muscle of Lawrence (MOL). These two 5' splice sites use different AUG codons, and as a result, the male transcript has 101 more amino acids at the amino terminus (Ryner, Goodwin et al 1996) than the female transcript.
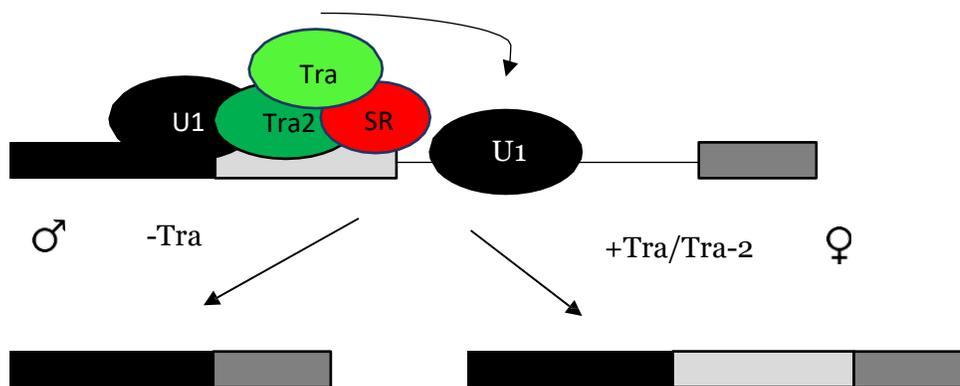
Figure 13. *Alternative splicing of the fruitless (fru) gene.* The two 5' splice sites can be seen clearly. The upstream 5' splice site determines male sex-specific splicing, while the downstream 5' splice site determines female sex-specific splicing (adopted from D.B. 2003).

1.3.5 Dosage compensation

Dosage compensation is a process that regulates expression of X-linked genes and balances the expression of the one X chromosome present in male individuals. The U1 snRNP splicing factor cannot bind to the 5' UTR due to the presence of a significant number of uridine rich elements. For dosage compensation to occur, another protein called TIA-1 is needed. This protein binds U1 snRNP to the 5' UTR, which strengthens the bonds between the weak 5' UTR and U1 snRNP. Binding SXL to the pyrimidine-rich sequence at the 3' UTR blocks the binding of U2AF65 and consequently U2 snRNP, which prevents production of the MSL-2 protein in female flies.

There is another prerequisite for effective SXL inhibition of the 3' splice site--the distance from the pyrimidine tract to the 3' splice site AG must be very great. The 3' splice site AG contains 13 nucleotides in *msl-2*. This leads to lack of interaction between the 3' splice site AG and U2AF65. As a result, SXL displays the U2AF65 splicing factor (Merendino, Guth et al.1999).
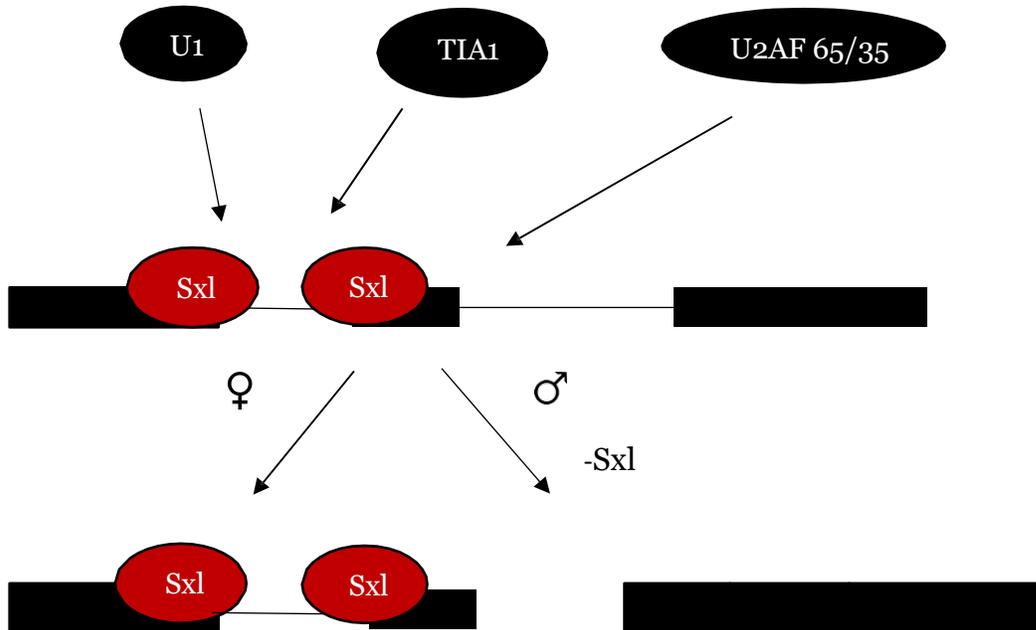
Figure 14. MSL-2 is a target of SXL. The SXL in females binds to two sites of the first intron of *Msl-2* pre-mRNA. One of the two sites is located in the pyrimidine tract, and the other binding site is located at the 5' splice site. SXL binds to the pyrimidine tract and blocks U2AF. The SXL also binds to the 5' splice site which blocks TIA-1 and also suppresses binding of U1 snRNP. Both binding sites of SXL are needed for retention of the intron of the *Msl-2* mRNA. The retained intron is in the 5' splice site (Forch, Puig et al. 2000).

Figure 15. summarizes the hierarchy of sex-specific alternative splicing in *Drosophila melanogaster*. The SXL protein stimulates, due to the presence of two X chromosomes, the skipping of exon 3 in pre-mRNA where the stop codon is found and helps to align exon 3 with the start codon on exon 2. This alignment leads to additional production of SXL protein, which further determines the female morphology in *Drosophila melanogaster*. In case of only one X chromosome, the SXL protein is not produced, and exon 3 is retained, introducing a stop codon, which leads to production of an inactive truncated protein. The SXL protein determines the retention of the intron at the 5' splice site of *msl-2;* as a result, production of female mRNA is inhibited. The SXL suppresses the use of the 3' splice site in *tra,* near which a stop codon is located. As a result, this stop codon is suppressed and TRA protein is produced, further determining the female phenotype. TRA regulates alternative splicing of *doublesex (dsx)* and *fruitless (fru)*. TRA-2 is expressed in both males and females but only determines the female specific alternative splicing in combination with TRA protein. This occurs only in female individuals. This process determines the female sex-specific alternative splicing and participates in the polyadenylation signal in *dsx* combined with the use of the downstream 5' splice site in exon 2 of *fru*. The result is a blend between DSX and FRU isoforms of these two proteins, and the expression of different proteins in the amino- and carboxyl terminals determines the sex of the flies (Forch and Valcarcel 2003).
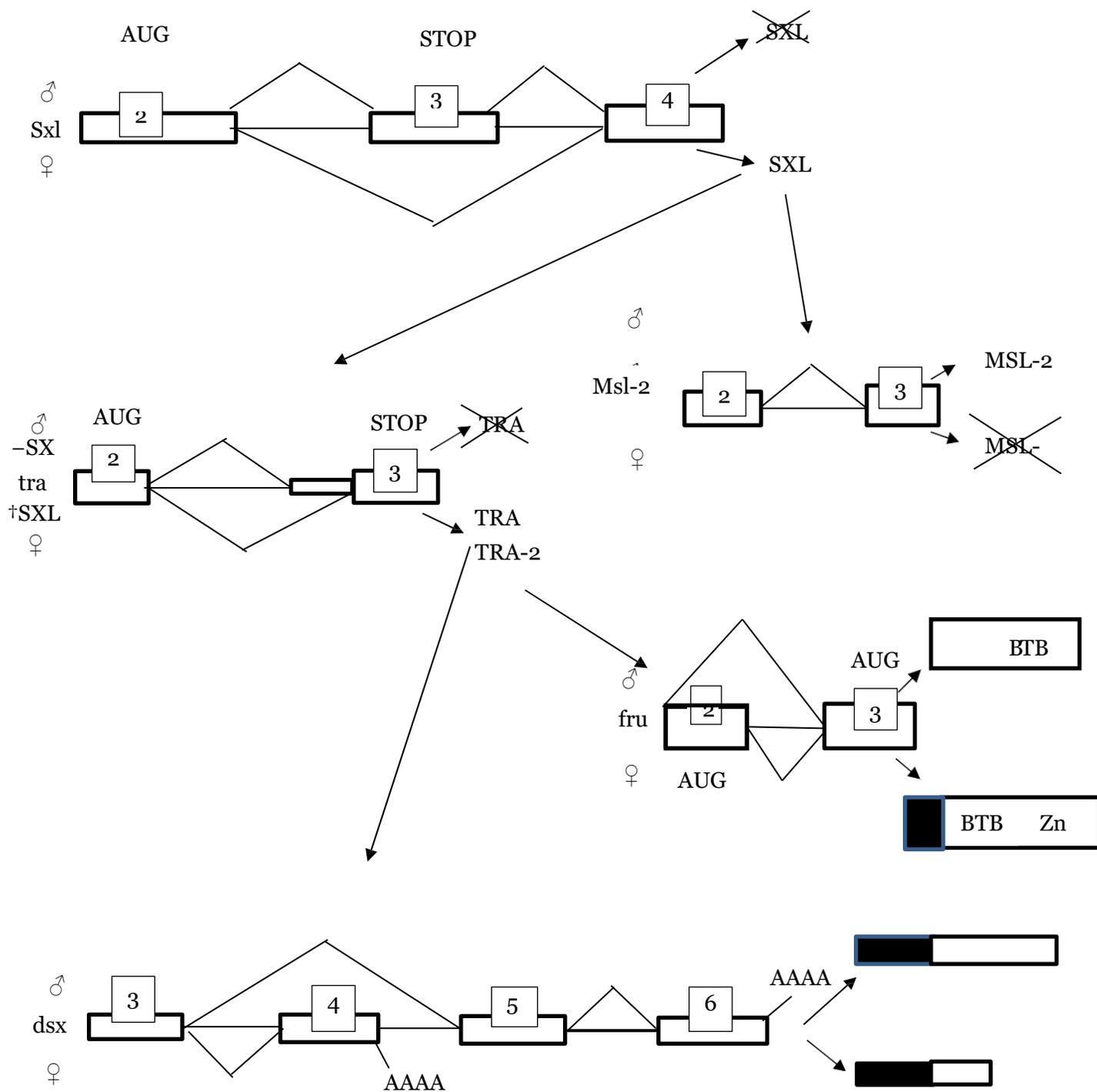
Figure 15. Hierarchy of splicing in *Drosophila* sex determination. (Forch, Valcarcel 2003)

1.4 Post-transcriptional modifications of mRNA

Post-transcriptional modifications of mRNA are processes that govern the stability of mRNA and which are essential for transmission to the progeny of genetic information with less interference from pressure outside the mRNA. Post-transcriptional modifications of mRNA will be concisely explained in the next paragraph.

Post-transcriptional modifications of mRNA can be divided into two categories. The first is RNA editing, which includes A-to-I editing (adenosine-to-inosine editing), C- to-U editing (cytosine-to-uracil editing), and conversion of uridine to pseudo-uridine (Ψ). The second category is internal modification, which includes m^6A methylation, that is, adding a methyl group to the sixth position of the adenine; and m^5C, which is methylation of the cytidine at the fifth position. The purpose of this paper is to elucidate the biological role of m^5C in the cell. In addition, cap1 and cap2 methylation of mRNA occurs and is interwoven simultaneously with the mRNA processing. Cap1 and cap2 methylation is the process of adding a methyl group to the cap at its 5' end at the 7th position of the guanosine (m^7G-cap) (Meister 2011).

1.4.1 A-to-I editing

A-to-I editing (adenosine-to-inosine editing) is a process where individual adenosine (A) bases are substituted to give rise to inosine (I) bases. A-to-I editing is catalyzed by ADAR (Adenosine Deaminase Acting on RNA). Schematically the reaction can be presented as follows:
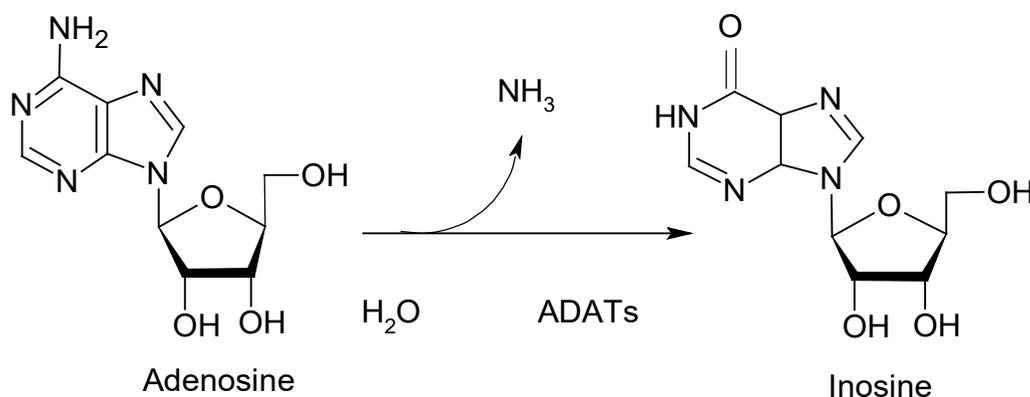
Figure 16. Alteration of adenine to inosine by ADATs. Deamination of adenine occurs at the sixth position where the amino group is substituted with a hydroxyl group.

Inosine is recognized as guanosine during translation, which leads to a change in the amino acid sequence which alters protein function and structure. The ADAR gene family has three identified members which belong to a discrete family of genes and have been found in vertebrates. They are referred to as ADAR1-3. ADAR1 and ADAR2 are expressed omnipotently in almost all tissues of vertebrates, while ADAR3 is expressed principally in the brain (Maas, Kawahara et al. 2006). All three proteins consist of one to three repeats of double stranded RNA binding domains (Meister 2011). These domains are required for interaction of ADAR with double stranded RNA substrates. The C- terminal domain conducts the catalytic deaminase activity.

The ADAR1 protein is expressed in two isoforms which are the product of alternative splicing (Maas, Kawahara et al. 2006). The ADAR1 mRNA is transcribed from two promotors, one IFN-inducible promotor that synthesizes the full form of the protein 150 kDa, and which is referred to as ADAR1 protein (p150). ADAR1 protein (p150) is interferon inducible. The other isoform is transcribed from a constitutive promotor and is known as ADAR1 protein 110 kDa (p110), which is shorter than p150 and does not have an N-terminal domain (Nishikura 2016). The IFN-inducible protein p150 is expressed principally in the cytoplasm, while the constitutive protein p110 is expressed principally in the nucleus (Maas, Kawahara et al. 2006). The expression of ADAR2 is regulated by the transcription activator CREB (cyclic adenosine monophosphate response element- binding protein). CREB serves as a repressor for the constitutive promotor ADAR1 p110 (Nishikura 2016).

Homodimerization activates the ADAR1 and ADAR2 and is a prerequisite for A- to-I editing. The third dsRNA domain of ADAR1 and the first dsRNA domain of ADAR2 participate in homodimerization and trigger the active form of the ADAR protein (D.B. 2003). ADAR3 is unable to dimerize. This may explain the lack of A-to-I editing activity in ADAR3 (D.B. 2003). It has been confirmed that both ADAR1 p150 and ADAR1 p110 move

back and forth between the nucleus and the cytoplasm. They are both exported from the nucleus with the help of nuclear export factors. The export of ADAR1 p150 mRNA transcript, which consists of the *AU-rich element* (ARE), is achieved by binding to the export receptor CRM1 (exportin1) or XPO1 together with the nuclear export signal (NES), whereas the export of the ADAR1 p110 is mediated by XPO5-RAN GTP and binds to TRN1 (transportin1).

The most important manifestation of mRNA editing is the production of a mixture of edited and non-edited species of protein which are co-expressed in the cell. This fine tune expression and regulation of cell specificity in a time-dependent manner (Maas, Kawahara et al. 2006). This process occurs in the genes' coding regions through a process called recoding-type editing (Nishikura 2016). A-to-I editing has been identified in the non-coding regions of the genes as well (Morse, Aruscavage et al. 2002; Rueter, Dawson et al. 1999) and within viral RNA after infection (Bass 1997; Polson, Bass et al. 1996).

Alu repeat elements have been studied extensively, and their participation in alternative splicing is validated. *Alu* elements participate in creating splice sites by A-to-I modification (Athanasiadis, Rich et al. 2004). A-to-I editing of *Alu* elements creates splice donor and acceptor sites, and all of these new sites are located predominantly in introns at the 3' UTR of *Alu* dsRNA. ADAR2 can self-edit the intronic sequences of its own pre-mRNA, which results in creating an alternative 3' splice acceptor (Nishikura 2016).

Unarguably, some precursors of microRNA (miRNA) have been edited, which leads to reduced expression of mature miRNA. ADAR1 edits miRNA and forms a complex with Dicer which further stimulates miRNA processing (Nishikura 2016). The binding of ADAR1 to Dicer promotes Dicer activity. Both isoforms of ADAR can form a complex with Dicer, but only ADAR1 p110 is able to activate the complex. ADAR1 can recognize the difference between RNA editing and RNA interference in that it can assemble in two distinct complexes. The ADAR1-ADAR1 homodimer has editing functions, whereas the ADAR1-

Dicer heterodimer participates in the RNA interference pathway (Nishikura 2016). ADAR suppresses miRNA maturation at different stages either directly or by editing.

Editing of primary miRNA-142 at position +4 and +5 by ADAR1 and ADAR2 is pre-requisite to keeping the Drosha-DGCR8 complex active (Nishikura 2016). RNAase III protein Drosha processes pre-miRNA. Drosha is folded into a dsRNA structure called a "hairpin" with help from the recognition factor DGCR8. After being processed in the nucleus, pre-miRNA is exported to the cytoplasm through the action of the export receptor Exportin 5. "In the cytoplasm [the pre-mRNA] has been processed by another RNAase III protein Dicer with the help of the TAR RNA-binding protein (TRBP known as TARBP2) to form a double-stranded intermediate complex, which is processed further to mature miRNA that has an average length of 22nt," (Nishikura 2016).

Dicer further participates in formation of the RNA-induced silencing complex (RISC) with the help of the Argonaut (AGO) protein. During formation of the RISC complex, the dsRNA is separated into two strands. One strand, called "functional," is paired with Argonaut (AGO) protein. The other strand, named "guide," directs the RISC complex onto the targeted miRNA (Nishikura 2016). Nucleotides 2-8 are the most important in the guide strand, because they are where the guide strand binds to the selected miRNA, which results in gene silencing. miRNA and Dicer form miRNP, which is transferred to Argonaut protein, where the dsRNA is further unwound, and only the guide strand is kept as a part of the RISC complex. The other strand named miRNA* (miRNA star) is typically removed from the cell, but in some cases miRNA* is retained and accumulates to form miRNPs which are capable of acting as functional miRNA, as well. The bi-functional miR-9/miR-9* pair is a good example, and both work independently in human neuronal cells (Meister 2011).

In addition to these three ADAR proteins, two more ADAR-like proteins have been

verified. One of these participates directly in spermatogenesis and is present in the testis and its named TENR. This protein possesses an RNA-binding domain. The other is TENR-like protein (TENRL) which is found in the brain (McKee, Minet et al.2005).

A small number of adenosines is edited in partially base-paired dsRNAs, for example, the glutamate receptor GRUA2, which is a precursor mRNA at the Q/R site. dsRNA is necessary for successful editing, which occurs in the area between the exonic sequence at the editing site, and a downstream intronic sequence named ESC (editing site complementary sequence) (Nishikura 2016). Adenine modification in A-to-I editing occurs most easily where a 5' uridine site is close to a 3' guanosine site (Nishikura 2016). A-to-I editing can contribute to pathologies. For example, editing of G-protein-coupled serotonin receptor 5-HT$_{2c}$R (Burns, Chu et al. 1997), potassium channel Kv1.1 (Bhalla, Rosenthal et al. 2004), and the α3 subunit of the GABA$_A$ (Y-aminobutyric acid type A) receptor (GABA3), are all implicated in negatively altering the structure and function of proteins.

Deficient A-to-I editing can also lead to the disease known as sporadic amyotrophic lateral sclerosis (ALS), which leads to the loss of the motor neurons in the spinal cord (Kawahara, Ito et al. 2004) and neuronal death in the CNC, based on forebrain ischemia (Peng, Zhong et al.2006). All these symptoms are based on insufficient editing of Q/R sites. The alteration of GLUR2 Q/R-site editing increases Ca^{2+} influx, which has debilitating overall effect (Nishikura 2016).

Mutations that lead to unedited sites or partial editing of the targeted proteins, have been identified with other pathologies such Aicardi-Goutières syndrome (ASG), in which nine ADAR1 mutations have been found. A mutation connected to a failure of the editing system of endogenous dsRNA in ALU elements is connected to sudden bursts of interferon in the body. ADAR1 and ADAR2 participate in editing of the 5-HT$_{2c}$R pre- mRNA and are involved in editing of 5' sites which are responsible for the production of

24 isoforms of the protein. In the case of unedited 5-HT receptor, the mutants demonstrate symptoms of decreased fat mass and increased expenditure of energy. In contrast, the decrease of editing at the same sites leads to development of Prader-Willi syndrome, with increased obesity and mental disorders. A-to-I editing is a vital post- transcriptional modification that leads to appearance of many pathologies but can also bring some positive effects such as protecting the organism from sudden and uncontrolled production of interferon in the blood stream after viral invasion.

1.4.2 C-to-U editing

C-to-U editing (Cytidine-to-Uridine editing) is a post-transcriptional modification which transforms cytidine into uridine. The deamination reaction is carried out by a large protein complex known as the editosome. The editosome consists of the catalytic subunit APOBEC1 (*apo*lipoprotein B mRNA-*e*diting enzyme *c*atalytic polypeptide 1), a cytidine deaminase. APOBEC1 is part of a large protein family with cytidine deaminase activity.

In humans, this family of proteins contains activation-induced deaminase (hAID) and has four members: APOBEC1, APOBEC2, APOBEC3AH, and APOBEC4. These proteins are active not only on the RNA substrate; they also participate in deaminase activities in DNA. All of these proteins contain a defined conserved cytidine deaminase motif (H-X-E-X_{23-28}-P-C-X-X whereas X is any amino acid). The amino acid chain in this motif contains a Zn ion, which guides a nucleophilic attack as part of the cytidine deaminase activity (Meister 2011). The APOBEC3 protein can inhibit the mobility of various mobile genetic elements including retrotransposons and retroviruses. An example of this is APOBEC3G, which consists of a factor that inhibits infection by the human immunodeficiency virus (HIV) (Meister 2011). APOBEC proteins have a wide range of

targets including B-cell maturation and antibody diversification, lipid metabolism, and editing viral RNA, such as HIV (Meister 2011).

Schematically, C-to-U editing can be presented as follows:
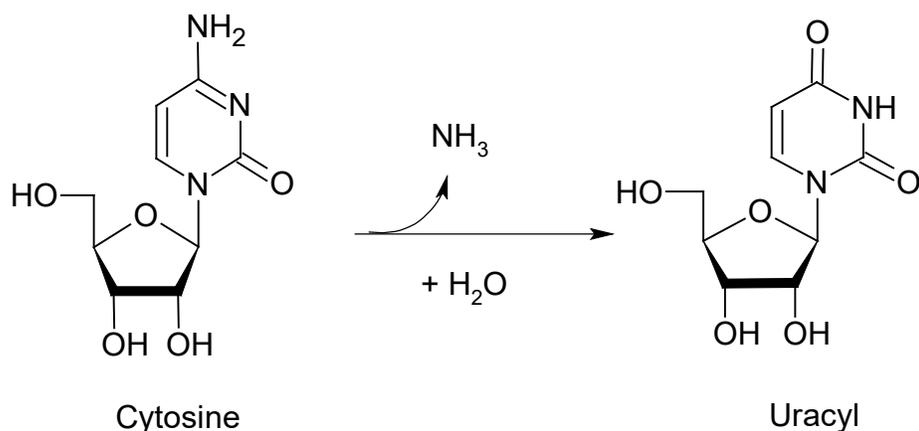
Figure 17. Cytosine deamination. The mechanism of C-to-U editing involves a hydrolytic deamination reaction at the C4 position of the cytidine. The *ApoB* mRNA has a specific motif that is embedded in its secondary structure and is surrounded by a 5' efficiency element and a 3' mooring element (Meister 2011). A water molecule carries out the nucleophilic attack on the amino group of the cytidine base, resulting in release of ammonium and transformation of the cytidine to uridine. The enzyme complex that participates in the conversion is referred to as the editosome, and its catalytic subunit is APOBEC1.

In APOBEC1 deaminase, APOBEC1 is paired with another protein named A1CF, which is an RNA-binding protein. Together, they facilitate the modification of C-to-U. The tandem APOBEC1-A1CF is highly expressed in tissues with intensive C-to-U editing. RMB47 has a structure like A1CF and forms a complex with APOBEC1 which is fully active and performs deaminase activity (Fossat, Tourle et al. 2014).

As mentioned above, APOBEC1 forms a complex called the editosome with the RNA-binding protein APOBEC1-Complementation Factor (ACF1, or ACF). *Apolipoprotein-b*

(Apob) mRNA has been investigated for many years. *ApoB* mRNA is edited in the nucleus. *ApoB* mRNA has been described as RNA that contains three regions that are important in the catalytic process and their distribution inside of the catalytic center. Two of these elements are distal flanking elements which appear to play a supportive function for the catalytic center. These are called the 5' and 3' efficiency elements (Chester, Scott et al. 2000). The 5' efficiency element is referred to as a "mooring" region and consists of 11 nucleotides. This element is located 4-5 nucleotides downstream from the edited C site and consists of an AU-rich element which is crucial for editing. The 3' efficiency element is referred to as "spacer" and is located between cytidine and the mooring sequences. Spacer consists of 2-8 nucleotides, and its function is intricately connected to identifying and finding the optimal size of nucleotides, which case is four (Chester, Scott et al. 2000).

The third element is referred to as the enhancer element. Alternative splicing is responsible for the emergence of two new proteins named APO100 and Apo48 which participate actively in lipid metabolism (Fossat and Tam 2014). The production of APO100 is mainly in the liver, where editing is achieved at position 2153 in exon 26, where a Glutamate codon (CAA) is transformed into a stop codon (UAA) (Chester, Scott et al. 2000). APO48 is made in the small intestines, in which a stop codon (UAA) is present at position 6666 of *ApoB* mRNA (Fossat, Tourle et al. 2014).

APOBEC is found in the nucleus as well as the cytoplasm. Structural characteristics of APOBEC1 reveal that this protein contains two important structures that contribute to equal distribution of the protein in the nucleolus as well as cytoplasm. These are: N-terminal domain, which has 56 amino acids in addition to a nuclear localization signal (NLS); and C-terminal domain, which consists of 24 amino acids and contains a cytoplasmic retention signal (CRS) as well as a nuclear export signal (NES) (Chester, Scott et al. 2000).

Alternative splicing is observed for *APOBEC1* mRNA, which is spliced to exon 2, which in turn gives rise to a new protein named APOBEC-T, which is expressed in normal, adenomatous and cancerous gastrointestinal tissue (Chester, Scott et al. 2000). The APOBEC1 protein binds to an AU-rich sequence located at the 3' UTR that exists in *c-myc* mRNA, *inteleukin-2* mRNA (IL-2), *granulocyte-microphage colony-stimulating* mRNA, and *tumor necrosis factor α* mRNA (TNF-α), all of which may suggest that the APOBEC1 protein participates in cell growth regulation, proliferation and tumorigenesis through these mRNAs (Chester, Scott et al. 2000).

Three APOBEC1-like proteins are described, named as follows: phorbolin-1, APOBEC2, and activation-induced deaminase (AID) (Chester, Scott et al. 2000). These proteins all participate in tumorigenesis and in some pathologies, such as psoriatic lesions, in which the phorbolin-1 is abundant. In contrast, APOBEC2 is expressed in involuntary striated muscle, which is heart muscle, and voluntary striated muscle, which is skeletal muscle (Chester, Scott et al. 2000) For the editing to proceed, the participation of other proteins such as RBP is necessary. These proteins participate as substrates for A1CF, including a protein named RBM47 (Fossat, Tourle et al. 2014).

RBM47 has three RNA recognition motifs (RRM) and is expressed in organs such as liver, lungs and the small intestine. RBM47 is first expressed in the foregut endoderm in 8.5-day-old mice embryos (E8.5).

Several more proteins are identified as suppressors of C-to-U editing, named as follows: KHSRP (Lellek, Kirsten et al. 2000), HNRNAB (known as ABB1) (Lau, Zhu et al. 1997), SYNCIP (GRY-RBP, HNTNPQ) (Blanc, Navaratnam et al. 2001), HNRNPC1 (Greeve, Lellek et al. 1998), and CELF2 (CUGBP2) (Anant, Henderson et al. 2001). (Fossat, Tourle et al. 2014). All of these proteins possess a common characteristic, in that they all bind to APOBEC1. In conclusion, C-to-U editing plays a crucial role in B-cell maturation, lipid metabolism, defense against retroviruses, and tumorigenesis.

1.4.3 Pseudouridylation (ψ)

The pseudouridylation RNA modification (ψ) is unusual because there is physical detachment of the uridine base followed by a 180° rotation, then followed by reattachment of the newly formed base. Compared to the rest of the modifications, where a basic exchange of functional groups exists, pseudouridylation is a process of rearrangement including detachment and rotation of the existing pyrimidine ring without further exchange of active functional groups and with the formation of a unique carbon-carbon glycosidic bond in a process called isomerization of uridine to ψ (Yu and Meier 2014).

Schematically, pseudouridylation can be represented as follows:
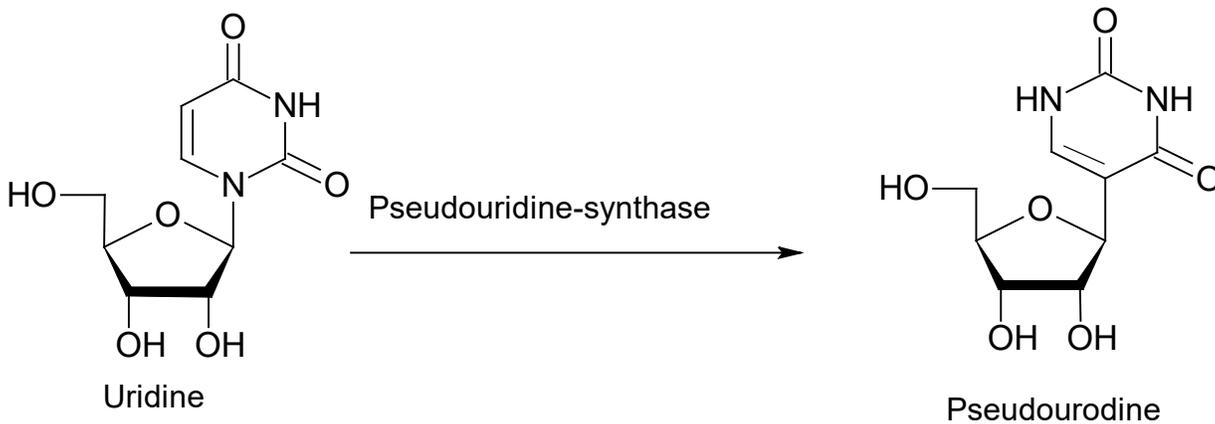
Figure 18. Pseudouridylation (ψ) of uridine.

Pseudouridylation is performed by enzymes named pseudo-uridine synthases with the help of auxiliary proteins and a class of non-coding small RNA called small nucleolar RNAs (snoRNAs). All of these components are assembled to form an RNA-protein complex. 200 unique snoRNA genes are known to exist. Many snoRNA genes are encoded in the introns of protein-coding genes, and the existence of non-coding snoRNA host genes is confirmed as well (Meister 2011). In both cases, the introns are spliced out

from pre-mRNA, and the snoRNAs are subsequently processed from the intronic RNA.snoRNA genes are frequently derived from independent genes. The independent genes give rise to polycistronic transcripts which are formed in clusters of independent genes. Individual snoRNAs have been assembled from such transcripts using endoribonucleolytic cleavage (Meister 2011).

Pseudouridylation can be accomplished in two ways. One way is with the help of the snoRNA-dependent pseudouridine synthases. The other occurs with site-specific pseudouridine synthases guided by consensus sequence (Schwartz, Bernstein et al. 2014). Pseudouridylation-guided RNAs identify uridines by site-directed pairing. These RNAs belong to a category called H/ACA RNAs, one of the two major classes of snoRNAs. The other major class of snoRNAs is referred to as C/D snoRNAs. These snoRNAs are responsible for methylation of targeted RNA residues at the 2' carbon of the ribose (Meister 2011).

H/ACA RNA has an average length of 133 nucleotides and folds into a stem-loop structure with a stem interrupted by a large bulge. This bulge is important for targeted RNA recognition and for pseudouridylation and is therefore called the pseudouridylation or pseudoU packet. The targeted RNA, which is complementary to the bulge, binds to the pseudoU packet, and an ACA or H box is located downstream of the stem-loop approximately 15 nucleotides away from the pseudouridylation site.

At the 3' end of the H/ACA RNA, there is a triplet called the ACA box because it consists of ACA nucleotides and is located only three nucleotides from the 3' end. Next to the ACA triplet is an H box which contains a hinge region with the nucleotide consensus ANANNA. All of these elements are part of the secondary structure of H/ACA RNA. The bulge that recognizes the targeted uridine is formed by two 3-10 nucleotides-long antisense components and is one of the two hairpins directly situated at the 5' and 3' end of the upper stream (Yu and Meier 2014).

Ribonucleoproteins (RNPs) that participate in assembling of the isomerization complexes are: H/ACA box, pseudouridine synthase and three other core proteins, one of which has a variety of names because it was discovered simultaneously in different species: NAP57 and dyskerin in mammals; Cbf5 in yeast; Nop60B in flies; and TruB in bacteria (Yu and Meier 2014). The structure of NAP57 is confirmed and is as follows: an N-terminal domain rich in lysine, followed by a central catalytic domain, an RNA-binding pseudouridylase, an archeosine transglucosylase (PUA) domain, and finally a C-terminal domain which is also rich in lysine. The other proteins that participate in formation of active H/ACA RNP are NOP10, NHP2 and GAR1. GAR1 has N- and C- terminal glycine- arginine domains which are important for assembling the active particle and which are directly connected to substrate turnover. GAR1 binds directly to the active center of NAP57 and works independently of NOP10 and NHP2 (Yu and Meier 2014). NAP57 forms the core of the particle that catalyzes uridine to pseudouridine, to which the two other proteins NOP10 and GAR1 bind independently, while NHP2 docks to NOP10 (Yu and Meier 2014).

How does the pseudouridylation catalytic site work? The ACA triplet of the H/ACA RNA binds to the PUA domain of NAP57, whereas the loop of the hairpin connects with NHP2, which subsequently expands the localization of the pseudouridylation pocket over the central catalytic domain. H/ACA RNA covers all the above-mentioned proteins. (Yu and Meier 2014). As mentioned above, pseudouridylation is a process that breaks down the N-glycosidic bond and rotates the base 180°. The mechanism involved is based on the conservation of aspartic acid and involves a nucleophilic attack by the aspartate (Yu and Meier 2014). In the machinery of pseudouridylation, at least six components are involved simultaneously in the nucleus and cytoplasm.

The correct expression and folding of the H/ACA box depend on the existence of

two H/ACA box chaperones named NAF1 and SHQ1. Stable expression of H/ACA RNA

cannot be achieved without help from these two chaperones, which do not participate in the mature particles or as part of the Cajal bodies (Yu and Meier 2014). The SHQ1 binds tightly to NAP57 and protects NAP57 from degradation and aggregation, which leads to the prevention of non-cognate RNAs from joining to NAP57. The covalent bonds between NAP57 and SHQ1 are so strong that an additional chaperone named R2TP is needed for their separation. The structure of the R2TP is composed of hexameric rings, which are made from AAA+ATPase, pontin (PIH1D1) and reptin (RPAP3) (Yu and Meier 2014). The separation of NAP57 from SHQ1 is based on disruption of the charged C-terminus with the help of ATP-ase (Yu and Meier 2014).

The chaperones are assembled with the help of RNA polymerase II and are part of the phosphatidyl inositol 3-kinase related kinase (Yu and Meier 2014).

In humans, 23 proteins bear the pseudouridine synthase domain, but most of these proteins have not yet been investigated, and their functions are unclear (Schwartz, Bernstein et al. 2014). The most studied are PUS1, PUS4, and PUS7. Any alteration in the structure of these proteins leads to a different manifestation of pathology, varying from affecting sensitivity to bacteria, through lethality in *Cerevisiae* when ablation of rerun pseudouridylation occurs.

In humans, the most prominent pathology is dyskeratosis congenita, a disorder characterized by failure of ribosomal biogenesis and accelerated risk of cancer resulting from malfunction of *DKC1*/Diskerin, the mammalian ortholog of *CBF5*. Deletion of *pus1* in *S. cerevisiae* results in growth defects. The mutation of *PUA1* proceeds to mitochondrial myopathy and sideroblastic anemia (Schwartz, Bernstein et al. 2014).

1.4.4 Methylcytidine

Methylcytidine is a post-transcriptional modification of mRNA that remains elusive. Until recently, 5-methylcytosine was detected using purified RNA, after which separation techniques such as liquid chromatography and mass spectroscopy were used (Hussain, Aleksic et al. 2013). Use of labeled H in living cells enables detection of m^5C in the mRNA. Advances in techniques to determine methylation of mRNA, such as high- throughput screening, in combination with next-generation sequencing (NGS) have enabled the discovery of widespread modifications in coding and non-coding RNAs (Hussain, Aleksic et al. 2013). The regulatory functions of m^5C are questionable, but it has been confirmed that m^5C facilitates binding of Mg^{2+} to Tran, which brings stability to the secondary structure of Tran through its impact on the anticodon stream loop (Hussain, Aleksic et al. 2013).

The methylation on the fifth position of cytosine, by itself or combined with other modifications, brings stability and protects RNA from degradation and cleavage. All of the modifications mentioned above can play a role in the regulation of the translation (Hussain, Aleksic et al. 2013). Methylation at the fifth position of the cytidine (m^5C) is described as follows:
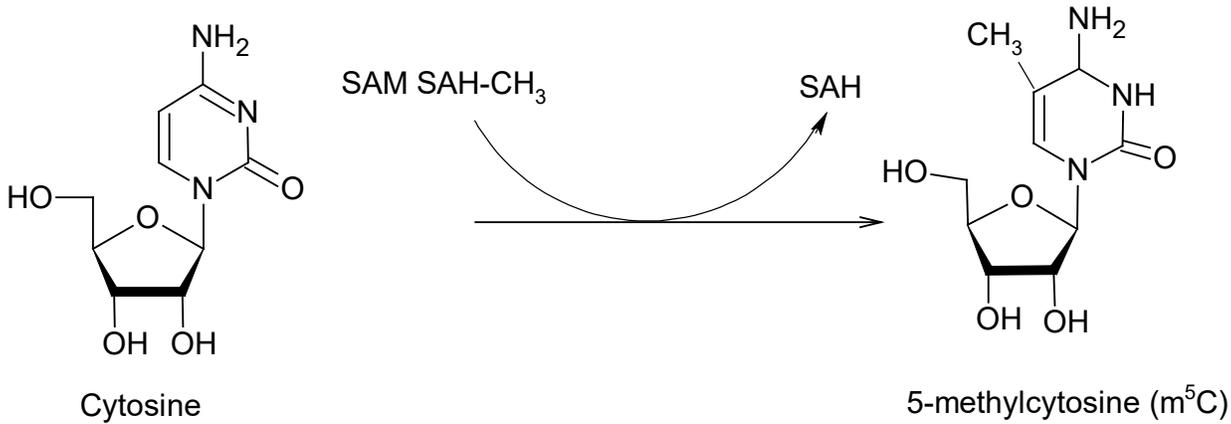
Figure 19. Methylation of m^{5}C proceeds using methyl donor S-Adenosyl Methionine (SAM) with methyltransferase as the enzyme.

The enzymes responsible for producing m^{5}C are called methyltransferases. These enzymes belong to a large group of proteins, with six members, divided based on their structural and functional properties:

1. RsmB/Nol1/NSUN1

2. RsmF/Yeb/NSUN6

3. RImI

4. Ynl022

5. NSUN6

6. DNMT2

Only the DNMT2 enzyme is of the single-cysteine type that catalyzes similarly to DNA-Mtases, with the cysteines localized in their catalytic pockets (King and Redman 2002), while the rest of the methyltransferases utilize two cysteines (King and Redman 2002). The NSUN2 enzyme is the best studied and is characterized by m^{5}C RNA-Mtase, which is involved in targeting specific tRNAs together with DNMT2 (DNA methyltransferase 2). NSUN2 together with DNMT2 are more highly conserved enzymes than methylate-specific tRNAs.

The methylation reaction starts with formation of a covalent bond between the

conserved cysteine and the RNA (King and Redman 2002). Bisulfite analysis of NSUN2 reveals that this enzyme has a broad spectrum of targets, including various tRNAs, that are methylated at positions 34, 38, 49, and 50 (King and Redman 2002). NSUN2 methylate extremely specific targets in vault RNAs and mRNAs. The structure of methyltransferase is as follows: The N-terminal domain, which includes both pseudo- uridine synthase and an archaeosine transglucosylase referred to as the PUA domain, followed by the RNA binding domain in the middle of the structure; and finally, the C- terminal domain in which is nested the S-adenosyl methionine (SAM) binding domain.

Figure 20. and Figure 21. represent the structure of NSUN6, which has two isoforms—A and B. The difference between the two isoforms is in the length of the molecule, which likely is linked to the difference in their function. As the schematic representation shows, isoform A, AdoMet_Mtase, is located in the interval from 149-285, whereas in isoform B the same domain is in the interval 230-366. The PUA domain, responsible for RNA binding, is located differently in both isoforms. I n Isoform A, the PUA domain is in the 21-90 interval, whereas in isoform B, the same domain is in the 102- 171 interval.

NSUN6 is part of the RsmB/RsmF family. C^5 methylase has a different length in both isoforms. In isoform A, C^5 methylase is in the 126-358 interval, whereas in isoform B, the C^5 methylase is in the 207-439 interval. There is still insufficient clarity about the differences in function of both isoforms.
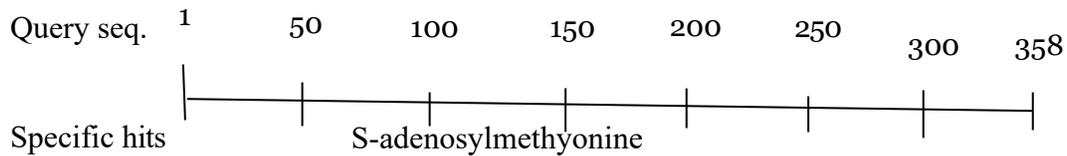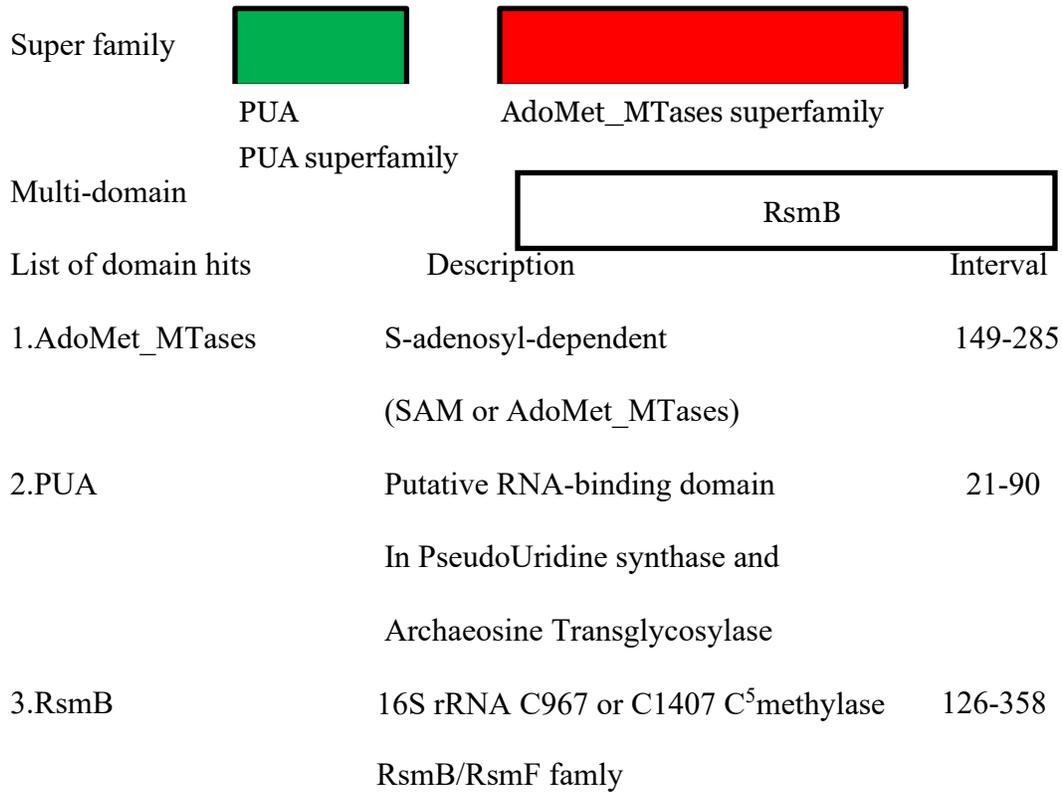
Figure 20. Schematic representation of CG11109 (NSUN6) isoform A. (ncib.nlm.nih.gov), next representation is the isoform B, which continue on the next page.

Position	1	75	100	150	225	300	375	439

Specific hits   S-adenosylmethionine binding site

Superfamilies

PUA   AdoMet_MTases

PUA superfamily   AdoMet_MTases superfamily

Multidomains

List of domains	Description	Interval
1. AdoMet_MTases	S-adenosyl-dependent (SAM or AdoMet_MTases)	230-366
2. PUA	Putative RNA-binding domain In PseudoUridine synthase and Archaeosine Transglycosylase	102-171
3. RsmB	16S rRNA C967 or C1407 C^5methylase RsmB/RsmF famly	207-439

Figure 21. Schematic representation of CG11109 (NSUN6) isoform B (ncib.nlm.nih.gov).

Targets of the RNA of m^{5}C-RMTs can be identified through deep RNA bisulfate sequencing of tissue and cell lines, but this method requires pre-size and effective enzyme knockout. A new method named Aza-IP (5-azacytidine-mediated RNA immunoprecipitation) was developed to circumvent the disadvantages of using deep RNA bisulfate sequencing. This method enriches the RNA target of the specific m^{5}C-RMTs directly, by identifying the exact cytosine targeted by the enzyme (Khoddami and Cairns 2013). The m^{5}C-RMT forms a covalent complex-enzyme-substrate intermediate with its targets. The sulfur atom of a cytidine residue is the m^{5}C-RMT catalytic domain that covalently binds to the C^5 position of the targeted RNA. Formation of the covalent bond is the first step, followed by methylation that occurs by enamine methylation of the C^5 position of the targeted cytosine, utilizing the methyl donor S-Adenosyl Methionine (SAM). Beta elimination is the preferred method to free the used enzyme.

The mechanism depicted above can be disrupted by the inhibitor 5-azacytidine (5-aza-C) and 5-aza-2'deoxycytidine (5-aza-dC). These substances are incorporated by RNA polymerase to the nascent RNA. When nitrogen substitution is achieved at C^5, the RMT enzyme remains bound to the targeted RNA. This results in the depletion of the endogenous enzyme, leading to hypomethylation of RNA (Khoddami and Cairns 2013). Consequently, the presence of 5-aza-C, even with the over-expression of an m^{5}C-RMT, results in only a small amount of free active enzyme (Khoddami and Cairns 2013).

The identification of methylated cytidine using this method led to the recognition of novel methylated RNAs including NSUN2-ribosomal RNA; the RNA subunit of RNAseP (RPPH1); and two mRNAs (CIN and NAPRT1). Eight candidates: 5SrRNA, RPPH1; vault RNAs (VITRANA1-1, VITRANA1-2 and VITRANA1-3); small Cajal body-specific RNAs (SCARNA2 and C/D box SNORNA); and Y RNA (RNY1), have been identified as potential substrates for the NSUN2 enzyme (Khoddami and Cairns 2013).

NSUN6 is a protein of interest, and this paper has only scratched the surface of a complex interaction between NSUN6 and other proteins responsible for participating in different processes in the cell. The process of methylation of the cytidine goes through formation of an intermediate complex with a covalent bond between the enzyme and the conserved cysteines of RNA. NSUN6 is confirmed to be an m5C RNA methyltransferase, identified with the help of sequence and structure prediction based on the N-terminal domain. NSUN6 includes pseudouridine synthase and PUA domain; RNA biding middle part; and a C-terminal S-adenosyl methionine (SAM) binding domain (Haag, Warda et al. 2015). UV crosslinking and analysis of cDNA (CRAC) have identified enzymes that interact predominantly with tRNAs-tRNA^{cys}, tRNA^{thr} and tRNA^{arg}. UV crosslinking followed by pull down confirms that NSUN6 methylates tRNA^{thr} and tRNA^{cys} at position C72 (Haag, Warda et al. 2015). NSUN6 also methylates tRNA, which contains a 3' site CCA. Deletion of 3' CCA leads to the inability of NSUN6 to methylate RNA (Haag, Warda et al. 2015). NSUN6 methylation is not well understood, and further research is needed in this area.

1.4.5 Methyladenosine

The most common post-transcriptional modification is the methylation of adenosine at the N^6 position (m^{6}A). This type of methylation is dependent on the presence of the conserved sequence "RRACH." The mechanism of m^{6}A modification can be represented as follows:
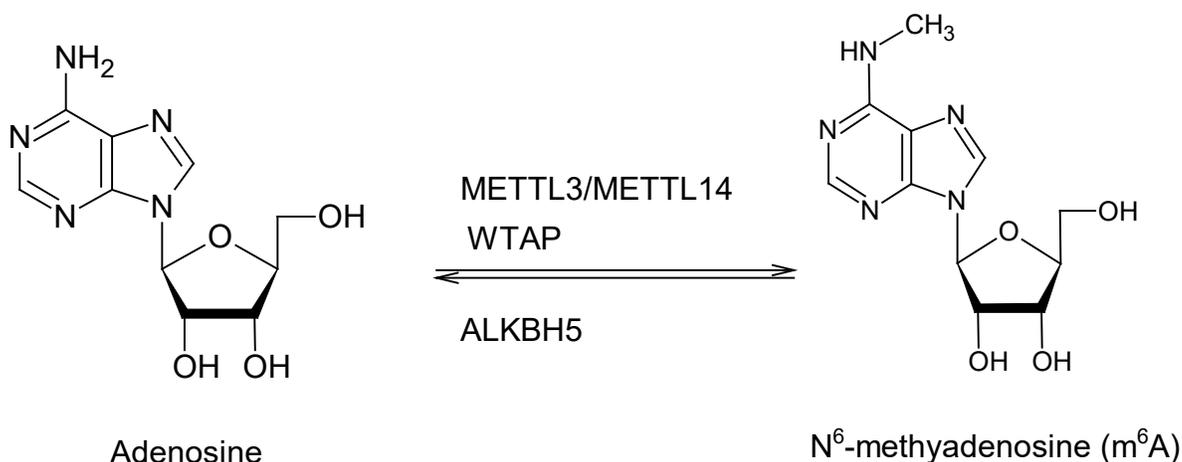
Figure 22. Mechanism of m^{6}A modification

Structural analysis reveals that m^{6}A methylation is implemented by a large methyltransferase complex composed of three subunits: METTL3 and METTLE14, both of which appear to function as catalytic subunits, and Wilms' tumor 1-associated protein (WTAP), which is the regulatory subunit (Adhikari, Xiao et al. 2016; Ping, Sun et al. 2014). Use of *in vitro* synthetic RNA substrate allows prediction of the capability and spatial capacity of the methyltransferases. m^{6}A is used as a further fraction *in vitro*. As a result, the methyltransferase separates into three fractions: Methyltransferase A1 (MT-A1), MT-A2, and MT-B. When these proteins are isolated, their molecular mass is confirmed as they emerge. MT-A2 has a molecular mass of 200-kDa, and MT-B has a mass of 800- kDa. A 70-kDa protein identified initially as MT-A was later renamed methyltransferase- like protein 3 (METTL3) (Adhikari, Xiao et al. 2016).

In *Drosophila,* the IME4 protein is a homologue of METTL3 and is expressed in ovaries and testes. This leads to the conclusion that this protein is part of the Notch

signaling pathway c. As mention above, MRTTL3 has two conserved functional domains. In addition, METTL14 has two conserved functional domains, one of which is called EPPL (Glu-Pro-Pro-Leu) and which seems to form part of the active catalytic domain as well as forming part of the SAM binding motif. METTL14 binds to the GAC sequence. METTL3 binds to the GAC and AAC sequences, which methylate preferentially. METTL3 and METTL14 form a stable heterodimer, and methyltransferase activity is enhanced as a heterodimer. METTL14 is also localized on nuclear speckles. The structure of METTLE14 is as follows: N-terminal domain, captured as coiled-coil domain, followed by a C- terminal domain containing a glycine-rich (G-ch) sequence.

WTAP is another component of the methyltransferase complex, although it functions as a splicing factor and binds to Wilms' tumor 1 (WT1) protein, and its activity includes methylation of mRNA. Knockout of WTAP reduces the level of m^6A. WTAP interacts with the METTL3-METTL14 dimer and colocalizes with them on the nuclear speckles. WTAP is highly conserved and, together with METTL3, forms a complex that methylates mRNA in *S. Cerevisiae,* called Mum2-Ime4- Slzl (MIS). Increased apoptosis is identified in the case of knockdown of this complex in zebra fish (Adhikari, Xiao et al. 2016). KIAA1429 is a newly discovered protein that participates in the methylation of mRNA. Its localization is on the nuclear speckles along with the other proteins responsible in the methylation of mRNA.

Methyltransferases are opposed by demethylases of m^6A erasers. The existence of two m^6A demethylases is confirmed. The first of these is called the fat mass and obesity-associated protein (FTO), and the second is called α-ketoglutarate-dependent dioxygenase alkB homolog 5 (ALKBH5). Bioinformatic analysis reveals that FTO is one of the Fe (II)-dependent and α-ketoglutarate -dependent AlkB superfamily of proteins and that it catalyzes oxidative demetylation (GERken, Girard et al 20017). FTO is expressed in the brain and predominantly in the hypothalamus in mice. FTO, as can be predicted from its

name, is connected to human obesity and energy expenditure. A single nucleotide polymorphism in the first intron of FTO determines the development of different types of body mass index and the risk of obesity in humans. Loss of function and over-expressed FTO mutation are reported. A mutation in Arg316Gln of FTO is connected to loss of function and causes postnatal retardation and malfunction (Boissel, Reish et al. 2009; Gao, Shin et al. 2010). In addition, this mutation causes increased postnatal lethality, growth retardation and altered body weight in knockout mice (Fischer, Koch et al. 2009). Over-expression of FTO in mice is connected to altered body weight and food intake (Church, Moir et al. 2010). *In vitro* functional assays show that single-stranded RNA can be demethylated at the 3-methyluridine (3mU) position by FTO (Gerken, Girard et al. 2007).

FTO is observed through immunofluorescent analysis in a dot-like partial speckle localization pattern. This implies a connection between FTO and m^6A- dependent splicing (Jia, Yang et al. 2008). The catalytic oxidation carried out by the FTO enzyme results in formation of two intermediated compounds, one called N^6-hydroxymethyladenosine (hm^6A), and the second called intermediate N^6-formyladenosine (f^6A) (Fu, Jia et al. 2013).

ALKBH5 (α-ketoglutarate-dependent dioxygenase alkB homolog 5) is a member of the AlkB family (Zheng, Dahl et al. 2013). Structural analysis reveals that human ALKBH5 has an N-terminal domain fold as a coiled-coil structure, and an alanine-rich C- terminal domain. Knockdowns of ALKBH5 result in increased levels of total m^6A and accumulation of polyadenylated RNA in the cytoplasm. ALKBH5 resides in the nucleus speckles (Zheng, Dahl et al. 2013).

The biological function of m^6A modification is important for increasing the stability of RNA. The importance of m^6A modification is reaffirmed in the evolutionary process by ensuring that this modification prevails and is correctly passed on to future generations. Evolution has developed the necessary tools to ensure that correctly modified information is not only passed on to future generations but that this modified information is correctly

read. The function of reading this information correctly is performed by a group of proteins called RNA readers.

The readers proteins are divided into two main groups, depending on their localization: cytoplasmic YTH domain containing families (YTHDF1, YTHDF2 and YTHDF3); and nuclear domain containing families (YTHDC1 and YTHDC2). These proteins selectively bind to RNA where m^6A exists and preferentially bind to a single-stranded RNA, which is long and contains 160 nucleotides (Dominissini, Moshtich-Moshkovitz et al 2012).

In humans, the YTH domain family contains five members as described above, all of which contain at their N-terminal region a Pro-Gln-Asn-rich domain, which is responsible for subcellular localization. The C-terminal domain of the proteins in this family is responsible for binding to m^6A methylated RNAs (Dominissini, Moshitch-Moshkovitz et al. 2012).

YTHDF2 has extremely high specificity, which helps to regulate RNA stability in the cytoplasm. YTHDF2 has an N-terminal domain which is Pro-Gln-Asn-rich and is responsible for taking the targeted mRNA to processing bodies (P-bodies), where the targeted mRNA is processed further for degradation. YTHDF2 competes directly with the ribosome for binding to mRNA. (Sheth and Parker 2003). YTHDF2 binds to its target RNA at the same time as the late stage of readenylations. This process forms part of the mRNA decay pathway (Wang and He 2014). In conclusion, YTHDF2 is transferred to the RNA degradation path after reading the m^6A substrate, and this pathway is reversible.

YTHDF1 and the YTHDF3 are identified as readers, but there is not enough information yet for a conclusive determination about their function in the cell. YTHDC1 and YTHDC2 are confined to the nucleus. It is confirmed that YTHDF1 participates information of YT (YT521-B) bodies, which are active transcript sites, localized close to the nucleus speckles (Nayler, Hartmann et al. 2000), and which can interact with other

splicing factors. YTHDF1 could function as a potential tumor suppressor in endometrial cancer (Zhang, zur Hausen et al. 2010).

YTHDC1 and YTHDC2 have the same binding motif, GGAC, which is used the most by both proteins (Yu and Meier 2014). YTHDC1 is found adjacent to stop codons, which may suggest that YTHDC1 is localized in the nucleus. YTHDC2 functions as an ATP-dependent RNA helicase and could be involved in cancer susceptibility (Fanale, Iovanna et al. 2014). In conclusion, m^6A plays an extremely important role in guarding and increasing the stability of mRNA.

1.4.6 Cap1 and Cap2 methylation

Eukaryotic mRNA is capped at the 5' end with a 7-methylguanosine nucleotide. The chemical structure of guanosine and N^7-methyl guanosine (m^7G) is as follows:
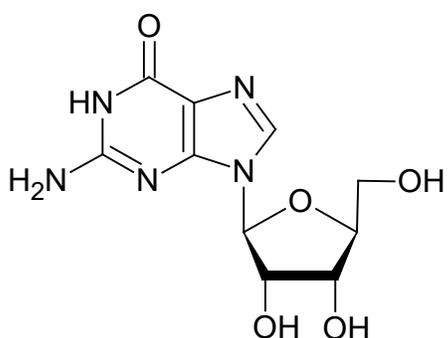
Guanosine
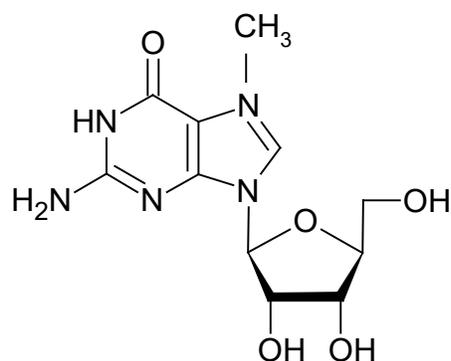
N^7-methylguanosine (m^7G)

The connection between N^7- methyl guanosine and the first base is unique in that there is a 5'-5' triphosphate bridge formed between the 2'-O-ribose, which is methylated on the first and second transcribed nucleotide, and N^7-methyl guanosine (m7G), which is inverted and which is connected by the 5'-5' triphosphate bridge with the first transcribed nucleotide (Werner, Purta et al. 2011).
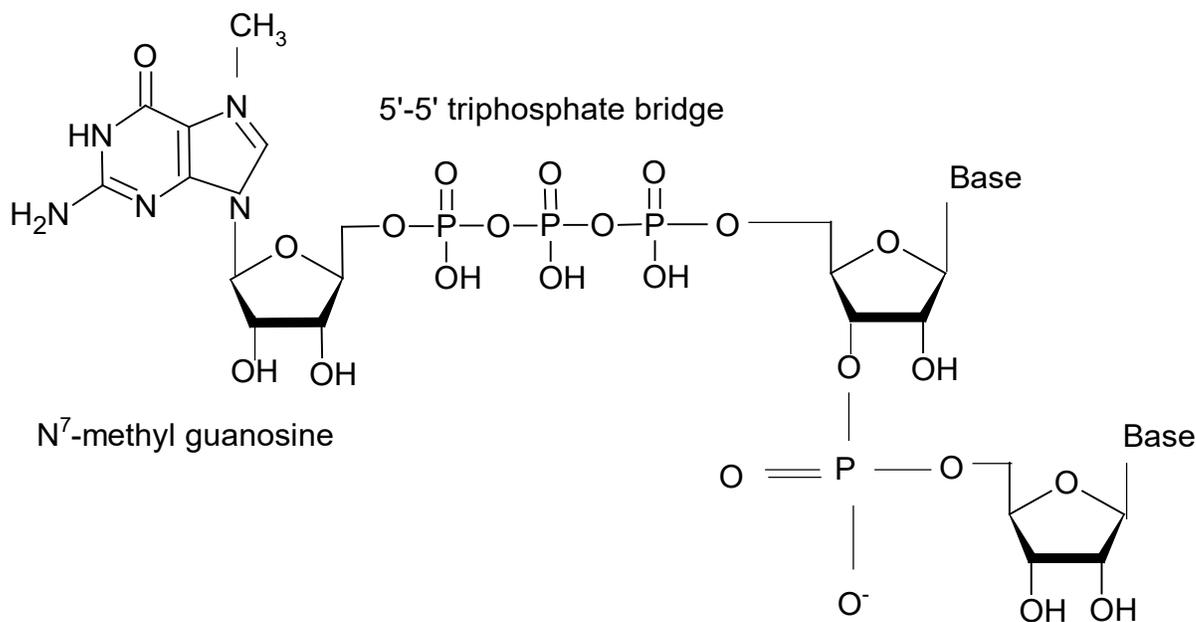
Figure 23. N^7- methyl guanosine linked by 5'-5' triphosphate bond or bridge to the pre-mRNA. (Meisner 2011)

This modification was established and retained during evolution because it may increase the stability of mRNA. The enzymes that carry out the reaction in humans are named respectively Cap1 2'-O-Ribose methyltransferase (hMTr1) and Cap2 2'-O-Ribose methyltransferase (hMTr2). An inverted m7G indicates the beginning of the transcription

site. The m7G cap protects the mRNA from the 5' and 3' exoribonucleases. Splicing of the pre-mRNA occurs more efficiently when the 5' cap is present. mRNA 3' processing is influenced positively by the 5' cap structure, and the export to the cytoplasm is more efficient and increased by the presence of m7G. After processing of the mRNA, none of the uncapped mRNAs is exported from the nucleus. In the cytoplasm, the presence of a functional m7G cap is necessary for correct translation. The translational machinery recognizes the 5' cap structure of the mRNA and binds to it in order to begin translation. The 5' cap is necessary to initiate translation (Meister 2011).

Three enzymes are required to process the 5' end of the mRNA: RNA triphosphatase, guanylyl transferase, and guanine-N7-methyltransferase (Meister 2011). RNA triphosphatase is the enzyme that catalyzes the first step of the capping reaction by hydrolyzing the γ-phosphate of the 5' terminal nucleotides of the pre-mRNA, creating a pre-mRNA with a diphosporylated 5' end. RNA triphosphatase structurally resembles cysteine phosphates and contains cysteine in its active center.

The second step of the formation of a 5' cap is the hydrolyzing of the GTP to GMP by guanylyl transferase, which takes place as a result of two sequential steps. The first step is removing the γ- and β-phosphates sequentially, as a result of which, newly formed GMP is covalently bound to gyanylyltransferase. In the second step, the enzyme transfers GMP to the dephosphorylated 5' end of the pre-mRNA. The guanylyl transferase is structurally remarkably close to DNA and RNA ligases.

The third step in formation of a 5' cap is methylation of the guanine base at the 7th position by guanine-N7-methyltransferase, with the enzyme utilizing S- adenosyl methionine (SAM) as a donor for methyl groups. As the result of methyl transfer, the SAM is transformed into S-adenosyl-homocysteine (SAH). In yeast, RNA triphosphatase is known as Cet1, and guanylyl transferase is known as Ceg1. These two independent enzymes form a complex (Meister 2011).

In mammals, Mce-1/Hce1 consists of both RNA triphosphatase and guanylyl transferase (Belanger, Stepinski et al. 2010). The 5' cap in yeast does not have additional modifications and it is referred to as capO. In humans, the cap structure consists of 2'-O-ribose methylation at either the first or the second nucleotide (Belanger, Stepinski et al. 2010).

In humans, snRNA participates in splicing of co-transcriptional capO, which is responsible for further 5'-end maturation consisting of hypermethylation of m^7G to $m^{2,2,7}G$ (trimethyl guanosine or TMG). Hypermethylation takes place in the cytoplasm by

the action of trimethylaniline synthase 1 (Tgs1) (Mattaj 1986). U1 snRNA is required for TMG formation and for proper snRNA trafficking (Mouaikel, Verheggen et al. 2002). In humans, capO and cap1 motifs are present in all mRNA molecules.

Two proteins have RNA 2'-O-ribose MTase activity: FTSJD1 and FTSJD2 (Werner, Purta et al. 2011). In FTSJD2, a second functional domain, which is remarkably similar to the GTase, is characteristic for capping enzymes in mRNA. Some other domains exist in the same protein based on sequence alignment: a G-patch domain may be involved in RNA binding; a conserved triad K-D-K, which is distinct from 2'-O-ribose; a WW domain which may be involved in protein-protein interaction; a GTase-like domain which does not have any catalytic residue. FTSJD2 has a nuclear localization signal (NLS) (Wernre, Purta et al. 2011).

FTSJD1 is also referred to as hMTr2 and has two RFM domains with conserved SAM-binding sites. hMTr2 has a catalytic K-D-K triad, similar to hMTr1 (Werner, Purta et al. 2011). S-adenosylmethionine (SAM) is used as a methyl donor.

Human capO MTase and hMTr1 function equivalently on both non-methylated and methylated RNA, whereas hMTr2 prefers capO RNA (Werner, Purta et al. 2011). The N-terminal domain of hMTr1 is predicted to have three motifs (80KKKRK, 103KKRKR, and 126KRK) responsible for nuclear interaction with importin-α and which promote RNMT

transcription and which selectively bind to the guanosine cap. As mentioned above, capping occurs co-transcriptionally within the first 20-30 transcript nucleotides of the nascent synthesized pre-mRNA.

These capping enzymes are recruited to the Pol II elongation complex by directly binding to the C-terminal domain (CTD) of Pol II. They subsequently phosphorylate serine 5 of the CTD, which serves as the container for the repeated sequence motif YSPTSPS (Meister 2011). This mechanism has evolved for precise control of cap formation. As the length of synthesized pre-mRNA reaches 20-30 nucleotides, specific proteins induce pausing of Pol II. These enzymes are NELF and DSIF. Pausing is a mechanism for efficient recruitment of the rest of the capping enzymes to the 5'-end of the pre-mRNA. After all the machinery has been built, the capping enzymes release the Pol II and stimulate transcription. Serine 5 is dephosphorylated during the elongation, which leads to dissociation of the capping enzymes (Meister 2011). The cap structure is constructed from specific, so-called *cap binding proteins* (CBP). The initiation of translation in the cytoplasm demonstrates the importance of the cap structure. The eukaryotic translational initiation factor eIF4E is a cap-binding protein which actively participates in recognizing the cap structure during translational initiation.

The cap structure is involved in pre-mRNA splicing, pre-mRNA 3' end formation, and mRNA export with the assistance of the *cap-binding complex* (CBC). The CBC is a heterodimer composed of two subunits, CBP80 (80 kDa) and CBP20 (20 kDa) which must bind to each other to express their function. The CBC is exchanged for eIF4E before export through the nuclear pore. Capping of the mRNA is likely to participate in and connect different stages of RNA maturation.

1.4.7 Aims

This paper seeks to investigate and observe the correlations that arise from the intimate interaction of the NSUN6 protein with genes responsible for sexual differentiation in *Drosophila melanogaster*. In order to shed light on the role of methylated mRNA in sexual differentiation in these flies, the first aim is to observe the influence of m^5C modification on alternative splicing in the *sxl, tra, tra-2, fru, dsx and msl-2* genes. Of particular interest is the role of the *sxl* gene, and how *sxl* gene splicing is influenced by m^5C in determining sexual differentiation in both sexes. The second aim is to observe the way in which modified s*xl*, in turn, influences the four other genes responsible for sexual differentiation: *tra* and *tra*-2; *fru*; *dsx*; and *msl*-2. The third aim is to observe the tight integration of NSUN6 mRNA with the proteins of all five genes.

Chapter 2

Material and methods

2.1 Drosophila genetics

Flies were kept in plastic vials with 10 ml standard agar food (8.5% dextrose, 6% cornmeal, 2.5% nipagin, 1% agar in tap water). Live yeast was added on the top of the food to encourage female flies to lay their eggs. There was strictly regulation of light following the circadian cycle of 12 hours daytime and 12 hours nighttime at 25°C and 70% humidity. Fly crosses were transferred into a new vial every 48 hours. Stocks were maintained continuously and flipped every 3-4 weeks to obtain the correct mutant.

2.2 Molecular Biology

2.2.1 RNA extraction.

A process of homogenization was carried out with three heads and three thoraxes of the correct mutant in 50 μl of Trisol in a tube. The pestle was washed with 450 μl of Trisol. The homogenized material was moved unto an Eppendorf tube, and the contents were spun in a vortex for two minutes. Next, 250μl of chloroform was added to the solution, and the contents were spun in a vortex again for 2 minutes. The Eppendorf tube was left for five minutes on ice until the phases in the tube were separated. The tube was then put in the centrifuge to spin for 10 min at 16 000 rmp. After that, 1μl of glycogen was added to a new tube. The aqueous phase from the old tube was transferred to the new tube with the glycogen. Then, one volume of isopropanol (450 μl) was added to the tube and spun in a vortex for two minutes. After that, the tube was put into the centrifuge and spun for ten minutes at 16 000 rmp. The tube's cap was opened and allowed to dry for

approximately five minutes. After that, 20 μl dH₂O-DEPC was added, or in case of performing reverse transcription (RT), only 7.4 μl of water was added. The last step was to freeze/thaw to dissolve RNA.

2.2.2 Reverse transcription (RT)

To convert RNA into cDNA, a RT reaction was performed. 13 μl of master mix was prepared for a total of 20 μl. The mix contained isolated RNA (9.5 μl), 10X RT buffer (200mM Tris-HCL, 500 mM KCL, pH 8.3) (1 μl), 100 mM DTT (1 μl), ribonuclease inhibitor (RNAsin-20 units) and 20 μM gene specific primer (1 μl). The samples were kept at 70°C for 15 min in a thermal cycle. After this, another 6 μl of master mix containing 10X RT buffer (1 μl), 100 mM DTT (1 μl), 100mM MgCL₂ (1 μl), 10 mM dNTPs (1 μl) and RNAsin (1 μl) was prepared and added to the previous reaction mix just before the temperature reached 50°C. The mixture was incubated at 50°C for 10 minutes. RT reaction was carried out at 46°C for 1 hour and 20 minutes utilizing Superscript II Reverse Transcriptase (Invitrogen). The final extension was at 70°C for 15 minutes.

2.2.3 Polymerase chain reaction (PCR).

For my samples, I prepared a master mix solution with a total volume of 50 μl. The mix contained 1.5 μl template (cDNA, genomic DNA or plasmid), 5 μl 10X Taq buffer (including 20 mM MgCL₂, Fermantas), 1 μl forward and return primers (20 μl), 1 μl dNTPs (10 mM), and 0.25 μl DreamTaq DNA Polymerase (5 U/μl, Fermentas). dH₂O was added to reach the 50 μl total volume. Eppendorf tubes with the samples were placed in a thermal cycler (Eppendorf) with the following procedure: initial denaturation at 94°C for 30 seconds, 30-37 cycles of denaturation at 94°C for 30 sec, an annealing step at 56°C for 40 seconds, and synthesis at 72°C for 45 seconds, followed by a final extension for one minute (annealing temperature and synthesis times differ with the primer type and length of PCR

128

product). A proofreading enzyme Phusion DNA Polymerase (Finnzymes) and 1X Phusion buffer was used for PCR cloning

For the single fly, PCR used one male fly transferred into the PCR tube and frozen at -80°C for 10 minutes. 200 µl of isopropanol was added the tube and incubated at room temperature for 30 min. The isopropranolol was removed via SpeedVac for one hour, after which a PCR mix was added to the tube and PCR reaction was carried out as explained above.

2.2.4 Agarose gel electrophoresis.

PCR products were visualized on 1% and 2% agarose gel. 1 or 2 grams of agarose were fully dissolved in 100 ml of 1X TAE buffer (Tris 40 mM – EDTA 1mM – glacial acetic acid 20 mM – pH 8.0). 5 µl of ethidium bromide was added and mixed well. The mixture was poured into a gel tank, and combs were placed. The gel was solidified at room temperature. 10 µl of PCR product was mixed with 9X DNA loading buffer containing glycerol (50%), EDTA (10% 0.5 M – pH 8.5), Tris (1% 50 mM – pH7.5) xylene cyanol (0.05%), and bromophenol blue (0.05%). The samples were loaded and run at 150-180 V for 45 min.

2.2.5 Competent cells preparation.

Frozen glycerol stock of DH5α cells were brought out and set up at 37°C. A single colony was inoculated in flasks with 10 ml of Luria Bertani (LB) medium. They were then continuously shaken vigorously and grown at 37°C overnight. The next morning, the pellet that grew in the flask was harvested and processed in a centrifuge at 4°C at 3000 rmp for 15 minutes. The supernatant was discarded, and the pellet was resuspended in 1 litre of ice cold 10% glycerol. The suspension was centrifuged at 5000 rmp for 15 minutes. The step was then repeated with 0.5 lt. and 250 ml of ice cold 10% glycerol. Following this, the pellet was resuspended in a total volume of 3.5 ml of ice cold 10% glycerol to produce a

concentration of cells at $1 - 3x10^{10}$ cells/ml. 100 µl aliquots were produced and then stored at -80°C.

2.2.6 Digesting PCR products.

PCR products were digested with 1 unit of enzyme/1 µl per hour. All digested mixtures were prepared as follows: 1 µl DNA added to 2 µl 10x buffer with 1 µl of each restricted enzyme, after which 16 µl of dH_2O was added. The enzymes were inactivated with phenol/chloroform/isoamyl alcohol (at a ratio of 50:49:1) and centrifuged at 16,000g for 10 min. 50 µl of supernatant was run on 1 % agarose gel. I used EcoRV and Xbal, two DNA restricted enzymes.

2.2.7 Media preparation

Luria Bertani (LB) medium was prepared by dissolving 1% peptone, 0.5% yeast extract and 0.5% sodium chloride in ddH_2O. The pH was adjusted to 7.0 with 5 m sodium hydroxide. The prepared mixture was autoclaved. SOC media was used in the final step of bacterial transformation and was prepared as follows: 2 % tryptone was combined with 0.5% yeast extract and added to 10mM sodium chloride, 2.5mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulfate and 20mM glucose in 1 lt of ddH_2O and then sterilized. This medium obtains maximum transformation efficiency of *E. coli*.

2.2.8 Induction of protein.

From all 16 falcon tubes, 1.5 ml of liquid was separated into new falcon tubes. Inoculation of 4-5 individually transformed bacterial colonies was performed into 2 LB+antibiotic, and they were then grown at 37°C for three hours (in order to reach $OD_{600}=0.5 - 1.0$). 1 ml of the amount was set aside for uninduced control. Multiple colonies were essential, since different transformants have different expression levels. 1 µl of 0.4M IPTG was added to the tubes, then left for three hours to grow at 37°C. Cell pellets

were processed in a microfuge for 15-20 seconds, and supernatant was aspirated.

Cells were then resuspended in 100 µl of ddH$_2$O, after which was added 1 µl of 200 mM PMSF. 50 µl of hot 3x SDS buffer was then added (pre-warmed in a boiling water bath for one minute before use), and the mixture was rapidly shaken and then boiled for 3-5 minutes. The mixture was then processed in a microfuge for 30 seconds. 5 µl of supernatant including appropriate controls, non-transformed cells and uninduced (transformed) cells was then loaded into an SDS-PAGE gel. A combination of 10 % running gel and 4% stacking gel was used for the SDS-PAGE gel to visualize protein production. The gel was stained with Coomassie blue. The resulting image is presented in Fig. 46.

2.2.9 Gibson assembly

The Gibson cloning master mix consisted of different enzymes using one buffer. Each enzyme had a unique function. The electrocompetent cells were thawed on a bed of ice. 50 µl of electrocompetent cells were then transferred to a pre-chilled electroporation cuvette with a 1 mM gap. The assembly products were then diluted three-fold with dH$_2$O prior to electroporation. This was achieved by mixing 5 µl of assembled products with 10 µl of dH$_2$O, after which 1 µl of assembly product was added to the electrocompetent cells. The cells were then mixed gently by pipetting.

Once DNA was added to the cells, electroporation was carried out immediately. 950 µl of SOC media was added to the cuvette immediately after electroporation. The solution was then placed in a tube at 37°C and was vigorously shaken (250 rmp) for 60 minutes. 100 µl of cells were placed on warm plates at 37°C and then incubated overnight at 37°C.

Chapter 3

Results

3.1 NSUN6 gene

The localization of CG11109 (NSUN6) in *Drosophila melanogaster* is found to be in the

3$^{rd.}$ chromosome L arm:

Figure 24. CG11109 (NSUN6) Gene localization. CG11109 consists of two isoforms, CG11109-RA and CG11109-RB. Their respective mRNA consists of 1445 bp for isoform A and 1415 bp for isoform B (Hoskins, Carlson et al. 2015)

The structure of the NSUN6 gene in *Drosophila melanogaster* is shown in Figure 25:

TAAAATCACAGAGCAAAATAAAACTTGACCGAGAATACACTTGCTGTTGAAAAAACTGCAAACTCACCAACATAAACGTTGCCGATA

1,745 1,750 1,755 1,760 1,765 1,770 1,775 1,780 1,785 1,790 1,795 1,800 1,805 1,810 1,815 1,820 1,825

NSUN6 - AUTO (100%)

917

Unknown

GAGATGGGCGCAGCGCATGAGCTAGAGCGGGATGCCGATGGTTGTGTAGTCCCGTGGTGAAAACTCGGGAAATGCTGCCACGTCTCTT

1,830 1,835 1,840 1,845 1,850 1,855 1,860 1,865 1,870 1,875 1,880 1,885 1,890 1,895 1,900 1,905 1,910

NSUN6 - AUTO (100%)

Nsun6 CI F7 - AUT... | P{wHy}CG...9DG30804 - AUTO (100%) | Nsun6 CI R7 - AUTO (100%)
end seq result 25.09. P2 D3 - AUTO (100%)
start seq result 25.09. P2 ...

Nsun6 CI F3 · misc_feature | Unknown | end seq result 25.09. P2 D4 · misc_feature

CTCCGCCACCTCGGCTAGAGTTGCCAACGATTTGGATTAATCGATTCATCATTGTCGCACAATGCTCACTCGATAAAGTTCACTATC

1,920 1,925 1,930 1,935 1,940 1,945 1,950 1,955 1,960 1,965 1,970 1,975 1,980 1,985 1,990 1,995 2,000

BalI

NSUN6 - AUTO (100%)

Unknown | misc_feature

CAGGTGATTTGAACTAAGTTAAATGTTAATATGTTTTACCCAAAAACACCATTTTTGGTGAACCCGTTGCTGGAAGCCGATATCCTT

2,005 2,010 2,015 2,020 2,025 2,030 2,035 2,040 2,045 2,050 2,055 2,060 2,065 2,070 2,075 2,080 2,085

EcoRV

NSUN6 - AUTO (100%)
Nsun6 CI R0 - AUTO ...

Unknown

AAAGTGAATGTATGTACTTTCAATGTGCACAAATACGTATTTACAACAAAAACTGGCTTGCAAATTTTATTAACTGTTAATTCCTGG

AAACCTACCCGAGATGTTGTGCATTGATCCGCTGGACTCACAGCTCACCAAAGCTGTGGCCGACTCCGAGCTCAAGGAGGIAGIGGT

2,355 2,360 2,365 2,370 2,375 2,380 2,385 2,390 2,395 2,400 2,405 2,410 2,415 2,420 2,425 2,430 2,435

SacI

NSUN6 - AUTO (100%)
Nsun6 CI F1 - AUTO (100%) | PUA, RNA binding doma

TGACACTAACTGCGGAGCTGCACTACTGCGAGGCGCCCACATATACGCCCCCGGTGTTTTAGCTATGGAGTCAAACACTGAGCGGGA

2,440 2,445 2,450 2,455 2,460 2,465 2,470 2,475 2,480 2,485 2,490 2,495 2,500 2,505 2,510 2,515 2,520

NSUN6 - AUTO (100%)
PUA, RNA binding domain - AUTO (100%)

GGAACTCGTCAATGTCTACGCAGATCTGGATGGAAAATGCAAGCGCGGGACGGTCAAGCGGTACGAAAGCCCAAATAAGGTTTTCCT

2,525 2,530 2,535 2,540 2,545 2,550 2,555 2,560 2,565 2,570 2,575 2,580 2,585 2,590 2,595 2,600 2,605

BglII

NSUN6 - AUTO (100%)
PUA, RNA binding domain - AUTO (100%)
Nsun6 CI R1 - AUTO (100%)

GGGTACAGGCAAGGTGTTAATGCAGCGGTATCAGCTGTTCAACAACGCGGAGAAGCCCGCCTCTGGAGTGGCCGTGGAAATGCAGTC

2,615 2,620 2,625 2,630 2,635 2,640 2,645 2,650 2,655 2,660 2,665 2,670 2,675 2,680 2,685 2,690 2,695

NSUN6 - AUTO (100%)
PUA, RNA binding domain - AUTO (100%)

CAACGTCAGCGGAGTGCCCACGCTGGGGGATCTAAGCAGCGAAGACGGCCTGTTGCAAAACCTGCCTTCTATTGTTTGCGTCCGTGT

2,700 2,705 2,710 2,715 2,720 2,725 2,730 2,735 2,740 2,745 2,750 2,755 2,760 2,765 2,770 2,775 2,780

NSUN6 - AUTO (100%)

NSUN6 - AUTO (100%)

misc_feature

CCTGGATCCGCAGCCCGGCGAGCGCATACTGGACATGTGTGCCGCTCCGGGTAACAAAACAAGCCACATAGCCGAGTTGATGGGCGA

2,790 2,795 2,800 2,805 2,810 2,815 2,820 2,825 2,830 2,835 2,840 2,845 2,850 2,855 2,860 2,865 2,870

BamHI

NSUN6 - AUTO (100%)

Nsun6 CI F4 - AUTO (100%)

TCGGGGATCTGTGGTGGCCCTAGACAATTCGGCCAGTCGCGTCCGCAGCATGCTGCCCAAATTGGGACACTACAAAAGCATAACGGC

2,875 2,880 2,885 2,890 2,895 2,900 2,905 2,910 2,915 2,920 2,925 2,930 2,935 2,940 2,945 2,950 2,955

SphI

NSUN6 - AUTO (100%)
AdoMet MTase - AUTO (100%)

GCATGTTTTTAACTCCACCGAGGCAGTGGCAGCGGACGCTCCATCTGCACCCGTGGGAGAATTCACTGGGCCTCCGTTTCCATGTGA

2,960 2,965 2,970 2,975 2,980 2,985 2,990 2,995 3,000 3,005 3,010 3,015 3,020 3,025 3,030 3,035 3,040

EcoRI

NSUN6 - AUTO (100%)
AdoMet MTase - AUTO (100%)

CTTGTCCTCATATCCCCATATTCAGCGCCGGCTATTTGAACAGGCAGTGCAGCTCGTTCGTCCTGGCGGCATCCTTGTATATAGCAC

3,135 3,140 3,145 3,150 3,155 3,160 3,165 3,170 3,175 3,180 3,185 3,190 3,195 3,200 3,205 3,210 3,215

NSUN6 - AUTO (100%)
AdoMet MTase - AUTO (100%)
Nsun6 CI F2 - AUTO (100%)

GTGCACCGTCACCGAGGATGAGTGCGAATGTCTTGTGGCCTGGGCACTAAGAAAATTTGTCGAGCTGCGGCTAACAGATGCCACTCC

3,225 3,230 3,235 3,240 3,245 3,250 3,255 3,260 3,265 3,270 3,275 3,280 3,285 3,290 3,295 3,300 3,305

NSUN6 - AUTO (100%)
Nsun6 CI R3 - AUTO (100%)

CCGGTGGGGCGGTCCCGGACTTCCGCTTCCCGGTTTCGAGGCATGTAAATCTCGTCTGTTGCAGCGCTTCGGTCCAAGCGGAGCAAA

3,310 3,315 3,320 3,325 3,330 3,335 3,340 3,345 3,350 3,355 3,360 3,365 3,370 3,375 3,380 3,385 3,390

NSUN6 - AUTO (100%)
Nsun6 ... | Nsun6 CI R3 - AUTO (100%) | Nsun6 CI R2 - AUTO (100%)
Unknown

133

Figure 25. Structure of NSUN6 gene in *Drosophila melanogaster*. 5' and 3' UTR are not shown. The blue color represents the P-element, which was inserted in the gene, the position of which is determined later. The red color represents the PUA binding domain, and the green color is the AdoMet MTase domain. (Generated with the help of Genomecompiler.com)

The NSUN6 methyltransferase is a conserved protein that is responsible for methylation of the specific mRNA in the cytidines at the 5th position. Applied UV crosslinking as well as chemical crosslinking are used to identify NSUN6 and its function. Methyltransferase NSUN6 is part of a substantial family of proteins named NSUN. Specifically, this protein is part of the Nol1'Nop2/SUN domain (NSUN) proteins. NSUN6 is found in the cytoplasm and colocalizes with a protein that is a marker for the Golgi apparatus, as well as with proteins from the pericentriolar matrix (Haag, Waeda et al. 2015).
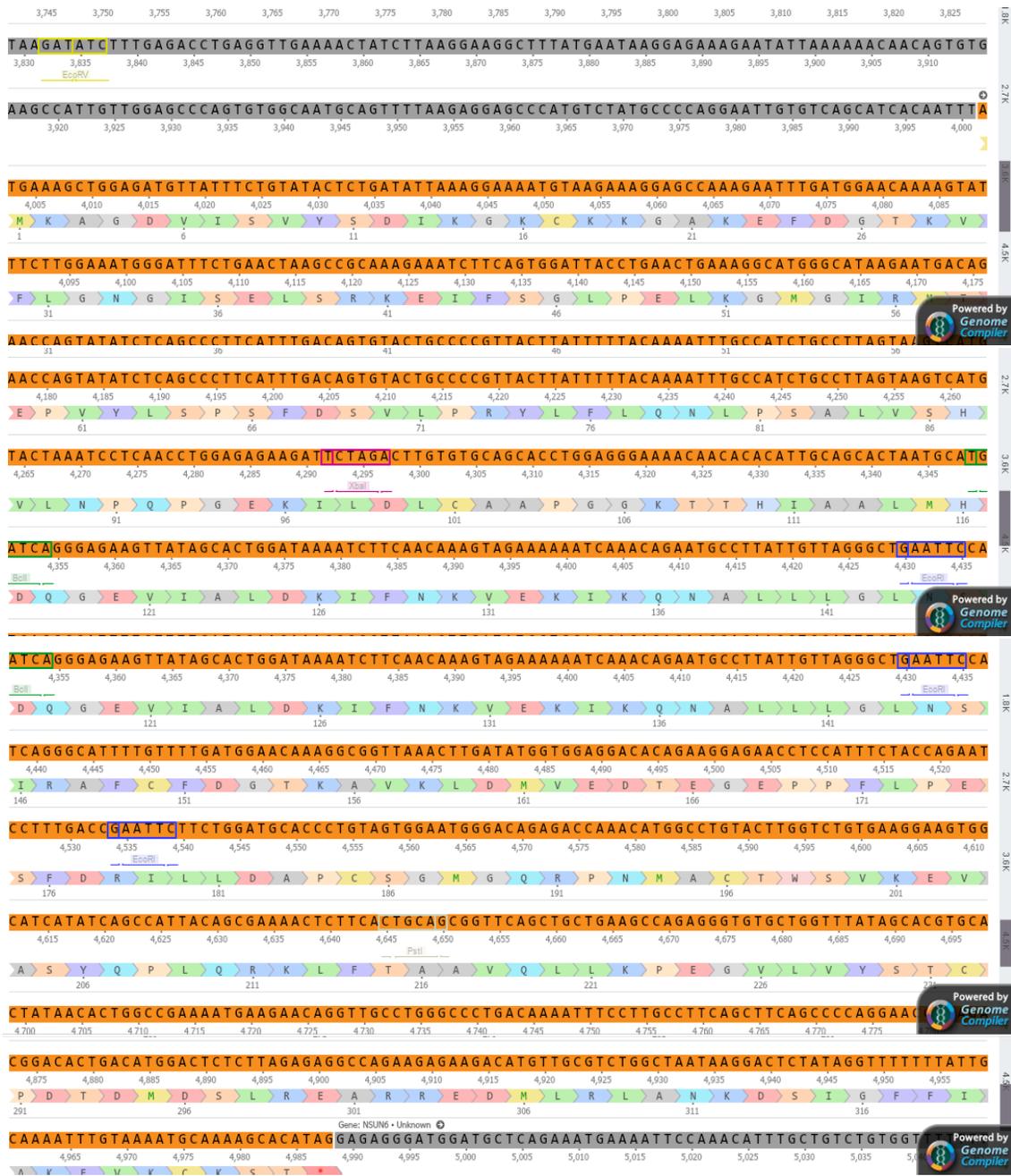
Figure 26. Predicted structure of *Homo sapiens* NOP2SUN methyltransferase. (Generated with the help of genomecompiler.com)

The *Drosophila melanogaster* NSUN6 gene consists of two exons. In *Homo sapiens,* the NSUN6 gene is made up of two exons, as well. To investigate the role of NSUN6 methyltransferase in this project, it was necessary to develop a null mutant for NSUN6 and an overexpressed mutant in *Drosophila melanogaster.*

3.2 Localization of inserted P-element

Before developing the null NSUN6 mutant, it was necessary to localize and insert a P-element in the CG11109 (NSUN6) in *Drosophila melanogaster*. Localization of the P-element took three steps that can be presented as follows:

First step

Second
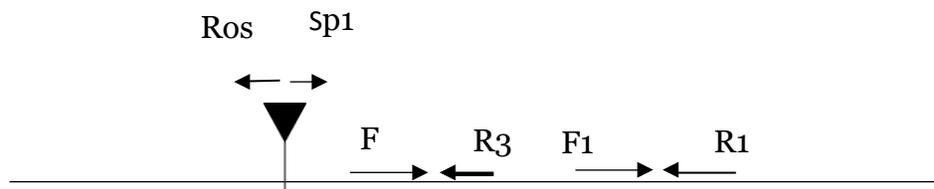
Third step
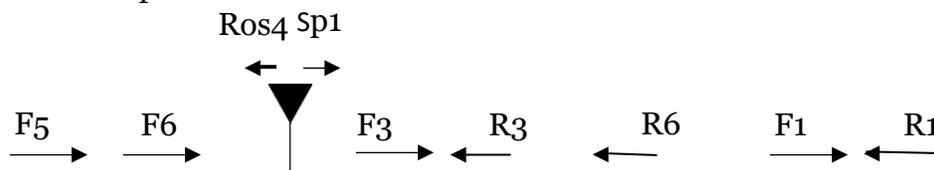
Figure 27. Steps taken to localize the inserted P-element in the NSUN6 gene.

Primers were used to localize the inserted P-element in the NSUN6 gene. They are as follows:

1. F3-CCGTCACCGAGGATGAGTGCGAATGT

2. R3-CGAGGCATTGTAAATCTCGTCTGTTGCAG

3. F1-CTGCTTGCCAAACCTACCCGAGATGTTG

4. R1-CTGGATGGAAAATGCAAGCGCGGGACGG

5. F4-CCGGGTAACAAAACAAGCCACATAG

6. R4-GCAGGGTGCATCGGAGCAAGATTCGGTCA

7. F2-CAGCGCCGGCTATTTGAACAGGCAGTG

8. R2-GTCCAAGCGGAGCAAATGCGGACAC

9. F5-gcgcgagtgttcgtggatagcccACAC

10. F6-gcggcttccaatactaatacttccact

11. R6-TCTCTGCCTTGACATACCACCCCC.

The first step in localizing the inserted P-element was to confirm the position of the 5'
and 3' ends of the P-element using F1R1 primers. The second step was to determine the
correct localization of the 5' end using Ros4 and F3 primers. The third step was to localize
the correct 3' site using Sp1R3.

The next figure represents the results for the localization of the inserted P-element:
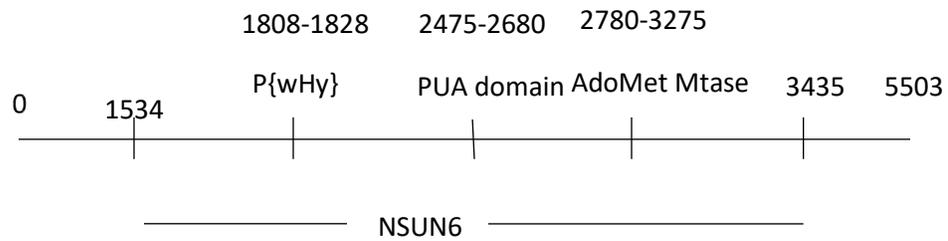
1. 1534-3435 NSUN6 domain
2. 1808-1828 P{wHy}
3. 2475-2680 PUA domain
4. 2780-3275 AdoMetMtase

Figure 28. Structure of the NSUN6 gene with precise localization of the inserted P-element and PUA domain as well as the AdoMetMtase domain.

3.3 Hobo element

Another transposable element, called a hobo element, was inserted inside of the P-element. The Hobo element was used to easily recognize the correct mutant along with using a pair of primers that would eventually show the jump out. The Hobo element has the tendency to catalyze local transposition and recombination between the elements of the same chromosome. The *P/Hobo* element was designed and can easily generate deletions that can be extended easily both ways from the site of insertion (Ashburner 1989). The construct *P{wHy}* is a *P-element* carrying w^+ and y^+ marker genes with *Hobo* element placed between them. The *Hobo* element in this case is used to generate a series of deletions. Exposure to *Hobo* leads to transposition of the internal element near the site which, as a result, leads to the deletion of chromosomal material between the insertion that was presented and the next insertion, which can be expressed as *P{5'wHy}* and *P{3'wHy}*. The insertions varied in size from 216 bp to 400 bp (Ashburner 1989).

Schematically, the jump out of the Hobo element can be represented as follows:

P {wHy} element

white[+]   Hobo   yellow[+]

NSUN6

NSUN6

Associated Deletion

NSUN6

NSUN6

Associated Deletion
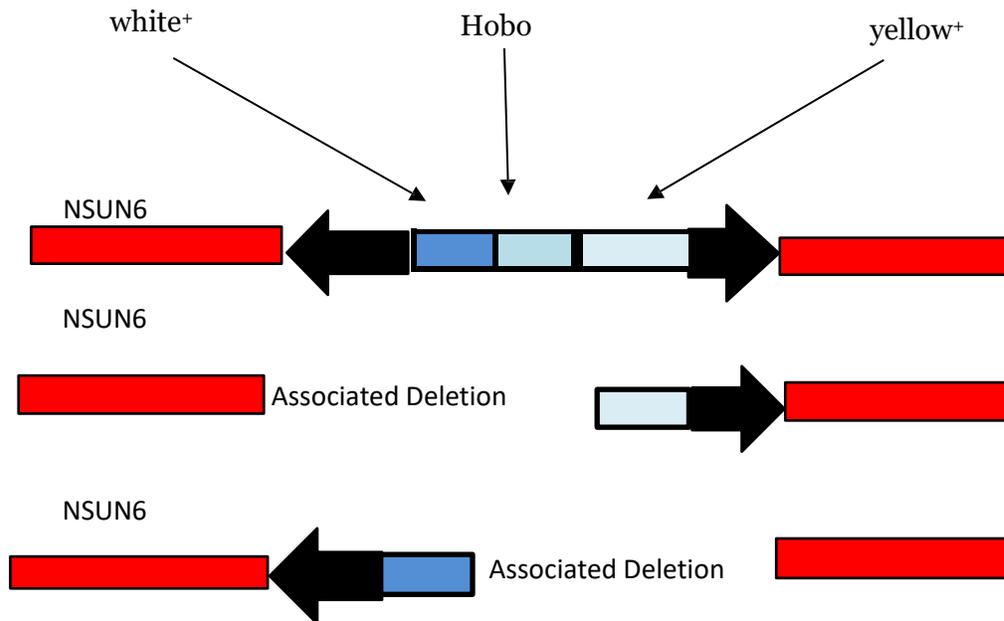
Figure 29. Hobo element inserted in the P-element. (Venken, Belen 2005) Hobo element inserted in the P element between two traits for white eyes (w^+) and normal body color (y^+), and associated deletion forward to downward from the Hobo element localization.

As can be seen in Figure 29, the Hobo element is locked out between two traits w^+ for white eyes and y^+ for normal body color. The Hobo element is highly unstable and, when it jumps out, it erases one of the traits. The erased trait is easy to trace and recognize because only individuals with one trait would be found in the cross. To validate the mutant with the Hobo element that has jumped out, this individual was crossed with a deficient mutant. The progeny of this cross was examined and used to validate the jump out. I carried out more than 1200 crosses and isolated six mutants which were incorrect. This validation of the Hobo element is also a part of the over-expressed NSUN6 mutants.

Two lines of mutants were developed with the precise jump out:

1. 3A NSUN6/TM 3Sb x Df w$^+$/TM 3Sb cross

2. 4C NSUN6/TM 3Sb x Df w+/TM 3Sb cross

Individuals from these two lines were used as a template for the PCR and sequenced with the forward primer F0-ACTAAACACTCACACGAACATCTACA to visualize the jump out. Agarose gel was developed:
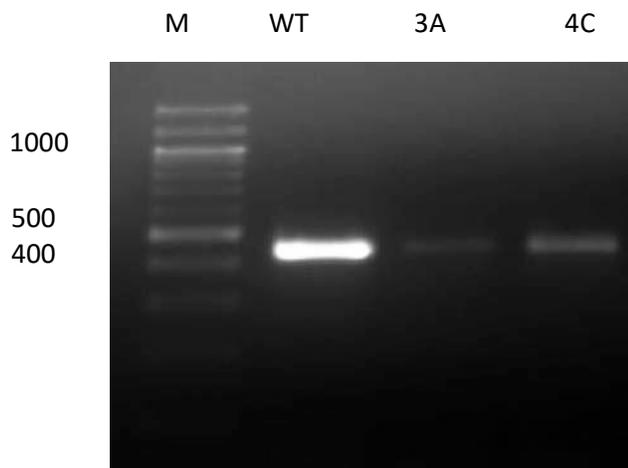

Figure 30. Agarose gel with WT, 3A and 4C primers which shows the precise jump out. The fragments are 400 bp.

To confirm the jump out, the PCR product was sequenced, and the confirmation of the actual jump out was visualized.

3AF0 P{wHy}

CAGCGCATGAGCTAGAGCGG

Figure 31. Jump out in the 3AF0 PCR product, observed in the 277-278 position. The blue region is the actual P-element in the gene.

# 4CF0 P{wHy} element

CAGCGCATGAGCTAGAGCGG
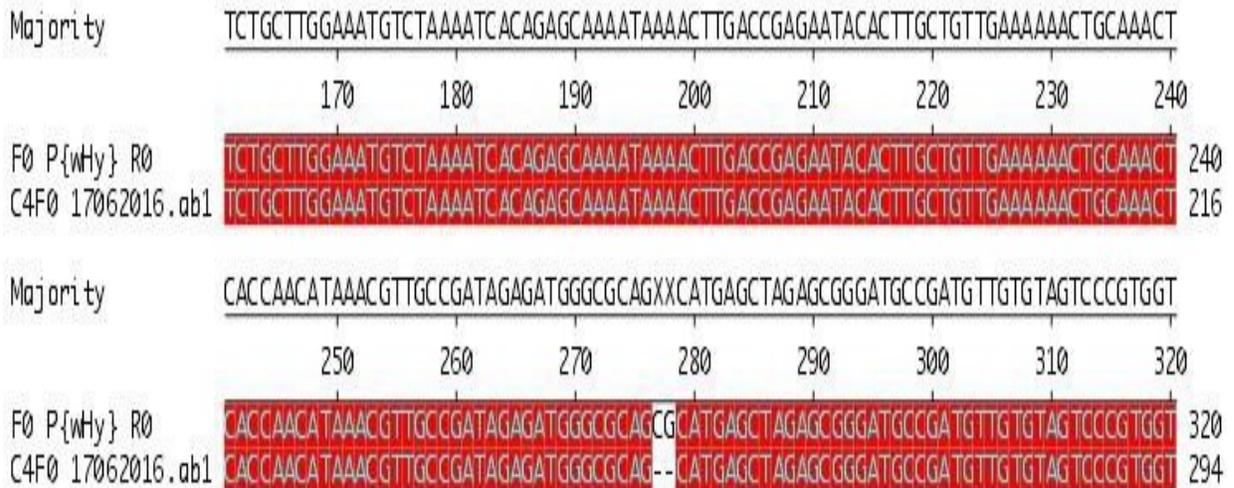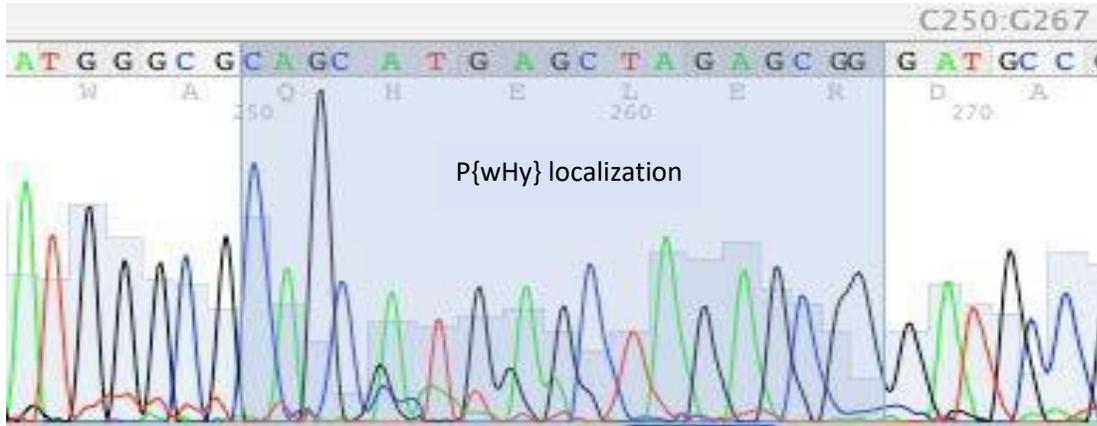
Figure 32. Jump out in the 4CF0 PCR product with the same results as 3AF0. This is definitive confirmation that the jump out was achieved, and again the blue area in the upper part of the figure presents the inserted P-element.

3.4 Null NSUN6 mutant

In theory, the null NSUN6 mutant should be clearly developed from different mutants using traits that are easily spotted. In this experiment, the first two crosses did not produce the desired mutant. For this reason, it was necessary to concentrate on a better strategy. But even the third cross did not produce the correct mutant.
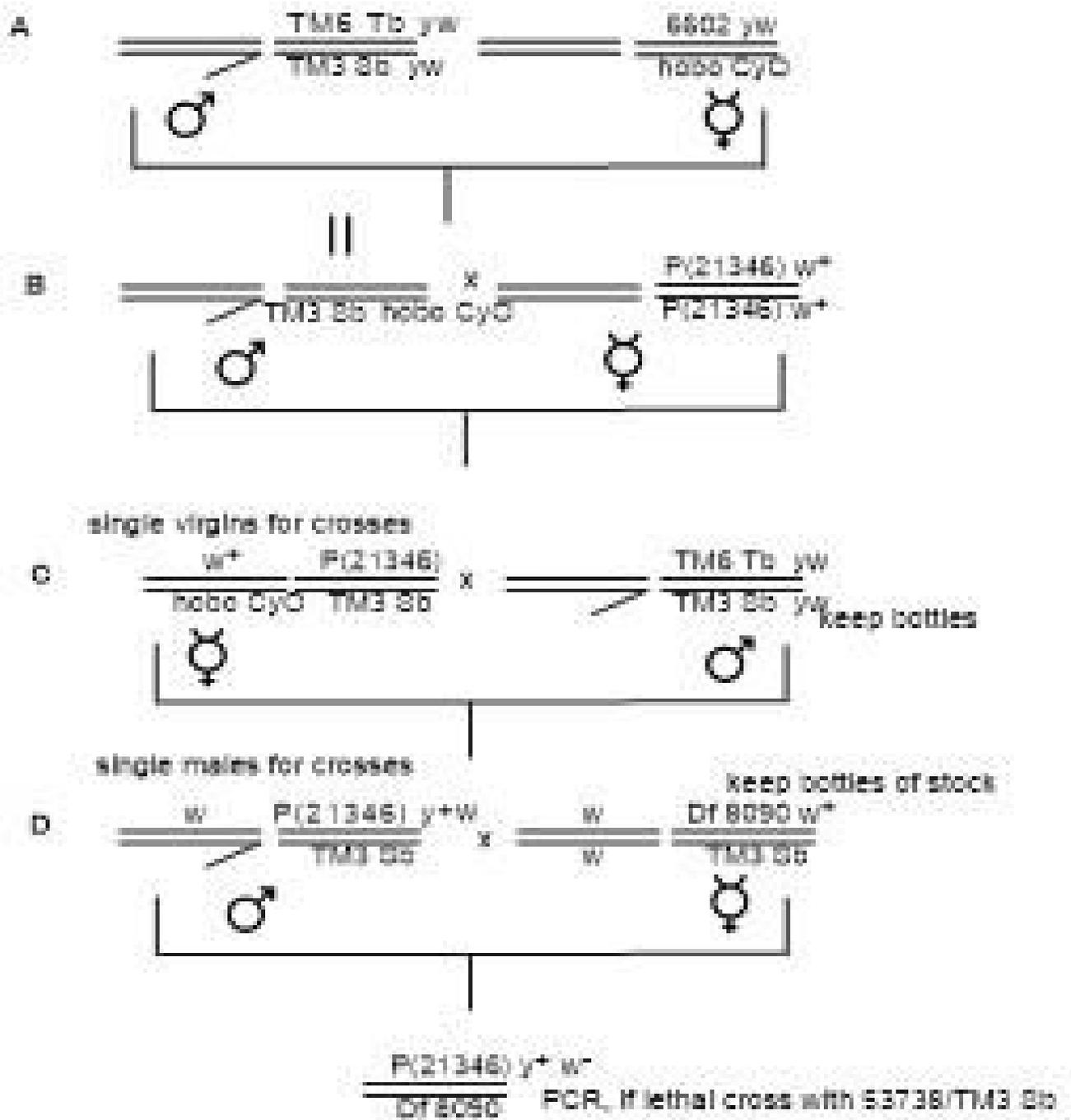
Figure 33. Crossing scheme of a mutant with inserted P-element, represented as P (21346).

Before continuing to describe the experiment, it is necessary to explain the role of chromosome balancers. In this crossing scheme, $y+w^+$ was used, together with *TM3* and *TM6* chromosome balancers at the 3rd chromosome to support the stubble-stubbloid gene (Sb) and tubby body appearance. The reason for introducing the chromosome balancer in the *Drosophila melanogaster* genetics was to maintain the lethal mutations and keep them in the progeny in a way that does not require special selection.

The mechanism, introduced by the balancer, is the following: a cross of l^l /+ x l^l/+, where l^l is a recessive-lethal mutation. This produces separation of the viable progeny into two types—one wild-type homozygous, and the other lethal heterozygous. If a heterozygous individual with two different but linked lethal mutations is crossed with another individual of the same genotype, it will produce a self-contained stock. The majority of the viable progeny will be identical to the parents, but this raises the problem that introducing a non-virgin female in the crossing scheme would generate breakdown. This could be avoided by introducing a crossover suppressor, that is, an inversion into one of the chromosomes. This would have the great advantage of suppressing the exchange over a long strand of genetic material (Ashburner 1989).

Flies must be recognized easily. There was a need to introduce a visibly recognized dominant mutation to mark the balancing chromosome (Ashburner 1989). In the second chromosome, the main chromosome balancer is based on the dominant *Curly* mutation, which has two inversions that are spontaneous in origin of appearance (Ashburner 1989). *CyO* was introduced as a third inversion to stop the breakdown. *Cy* is a good and effective suppressor of the exchange over the whole chromosome (Ashburner 1989).

Members of the third multiple (TM) series of balancers are good and effective

balancing elements for the third chromosome. *TM3* is one of the most used balancers for chromosome 3. *TM3* usually carries *Sb* and *Ser* as dominant markers, and in case of double exchange in *TM3*, *Ser* can be lost if both *FM6* and *SM5* are present (Ashburner 1989). The small terminal duplication y^+ is useful but can easily be lost. One way to ensure that *TM3* retains the y^+ is to keep it with homozygous chromosomes (Ashburner 1989).

TM6 is an X-ray-induced inversion. The *TM3/TM6* heterozygous is a viable mutant (Ashburner 1989). T h e *Hobo* element, linked to *CyO,* presents as a trait for curly wings. The first step was to cross a tubby male with normal body color and white eyes with a virgin female with curly wings. The progeny of this cross, which was a young male with curly wings and stubble, was used in the next step to cross with a virgin female carrier of the inserted P-element with orange eyes.

The next step was to isolate the progeny from this cross, which was a virgin female with orange eyes and curly wings. This female was a carrier of the inserted P-element in which the Hobo element was locked between two traits—normal body color and white eyes, and the male with a tubby body and stubble, which had two chromosome balancers in the third chromosome. The progeny from this cross was a male with white eyes, which was crossed with a virgin deficient with red eyes. The resulting progeny should have been the correct mutant, but multiple crossings failed to produce this correct mutant. All PCRs were negative for the correct mutant.

In this case, after changing the cross twice, the third cross did not work. It appears that there is a fundamental error in setting up the whole crossing scheme.

3.5 Over-expressed NSUN6 mutant

The *GAL4-UAS* transcriptional regulatory system is adapted from *Saccharomyces cerevisiae* into *Drosophila melanogaster*. The system was constructed from two separate transgenes. The first transgene has an *Adh* promotor and drives the expression of the yeast *GAL4* transcriptional activator. The second transgene, which carries the β-gal coding region, is located downstream from the *GAL4* binding sites and is known as the upstream activating sequence (UAS). Both elements were located close to the cross that was introduced. The expression of *lacZ* occurred in the tissues where *Adh* expression was visible. The expression requires both a *GAL4* driver and a *UAS* repeat upstream of the reporter (Ashburner 1989).

This two-part system was upgraded to a complex that would allow expression of *GAL4* driven by specific promotors or by enhancer traps, and a *UAS* vector that would be able to accept the coding sequence of a gene of choice (Ashburner 1989). The *GAL4-UAS* system has a variety of experimental applications, such as the expression of transgenes for cell-specific rescue of mutant phenotypes that are expressed in a mosaic manner, functional suppression of the synaptic activity in the nervous system as well as participation in the cellular signaling pathways or cell-to-cell communication (Osterwalder, Yoon et al. 2001). Over-expressed NSUN6 mutants demonstrate some abnormalities. For example, female flies possess defects in their ovaries, which could suggest that this alteration leads to infertility. My experimental results verified this to be the case.
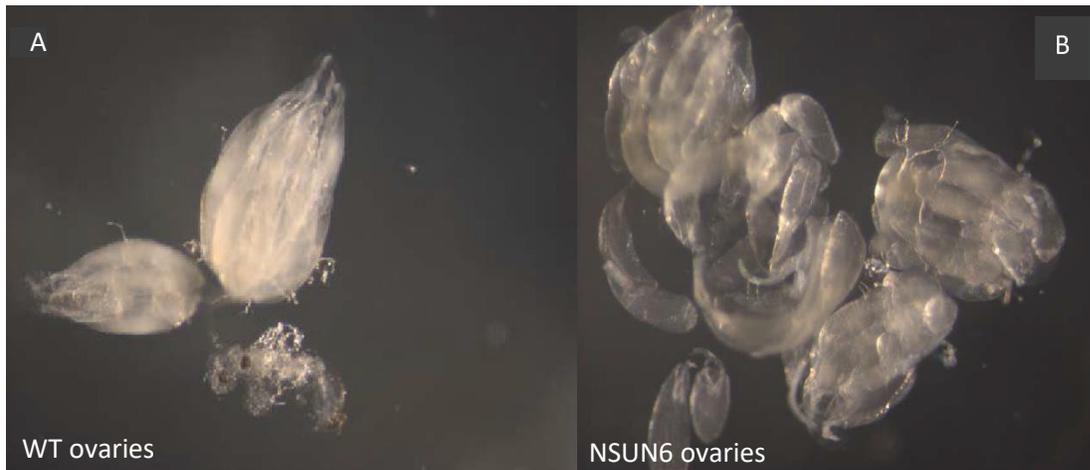
Figure 34. Ovaries of WT *Drosophila melanogaster*. In figure 34(A), the structure of the organ is complete, whilst the ovaries of the NSUN6 female mutants in figure 34(B) show structural deformity and abnormalities which lead to infertility.

As mentioned above, the *GAL4-UAS* system randomly inserts *GAL4*. In this way, *UAS* activates the gene of interest when one of the two parents possess the promotor-driver *GAL4* and the other possesses the *UAS* gene of interest. *GAL4* helps for tissue-specific expression in a restricted region in the fly so that the *UAS* gene of interest is expressed only in this region. The *UAS*- NSUN6 mutant is expressed in the NSUN6 protein, which is localized in the base of the brain of the larvae. Confocal microscopy could potentially show clear expression in the base of the larvae brains, but the expression of the NSUN6 protein in this mutant was connected to red fluorescent protein (RFP) which would interfere with the localization of the NSUN6 protein.

I was in the process of removing the RFP but did not have enough time to finish it. I used the NSUN6 mutant for immunochemistry. Developed antibodies against NSUN6 protein do not exist. The NSUN6 protein is tagged with HA sequences, which are localized in the C-terminal domain. HA sequence is a hemagglutinin protein, which is isolated first from the influenza virus and is the glycoprotein that is responsible for viral invasion into the cell and particularly regulates membrane fusion, as well as carrying out a receptor-binding function. HA sequence is interesting because it can tag the protein of interest in

a recombinant expression vector. Proteins produced from this system cause stable fusion, and the resulting products provide stable complexes. The anti-HA antibodies are specific for N- and C-terminal HA-tagged proteins. NSUN6 can be visualized using this antibody, and the procedure was completed in the following way:

Anti-HA rat antibody was mixed with dapi in a ratio 1:500 in 10% goat serum. A secondary antibody was used as anti-rat FITC green fluorescence in a ratio 1:200 PBTB.

As mentioned above, NSUN6 is tagged with red fluorescent protein (RFP), which was interfering strongly with the green channel. This was why it became clear that this protein should be removed.

The next step was to analyze the progeny resulting from a cross between a female with the genotype NSUN6-*UAS* and a male dsx-*GAL4*. The first-generation progeny resulted in NSUN6/dsx males and females, but notably the males of this cross had morphologically female genitalia. Genotypically they had XY chromosomes, and the females from this cross were sterile. To investigate fully these results, I had to determine the alternative splicing that occurs in *sxl, dsx, tra, fru and msl-2*. The introduction of this paper laid out in great detail the mechanisms that occur during alternative splicing in all these genes theoretically.

A control was set up using *ewg* primers to check whether the RNA extraction was done correctly. The two *ewg* primers are as follows:

1. *Egw* forward primer: 4F1-ATGGTACAACTGCCCAGGTTC
2. *Egw* reverse primer: 5R1-TGACATCACATTGCTCACCGA

I conducted PCR using these two primers and extracted RNA from *WT-male and female, and dsx/NSUN6-male and female.*
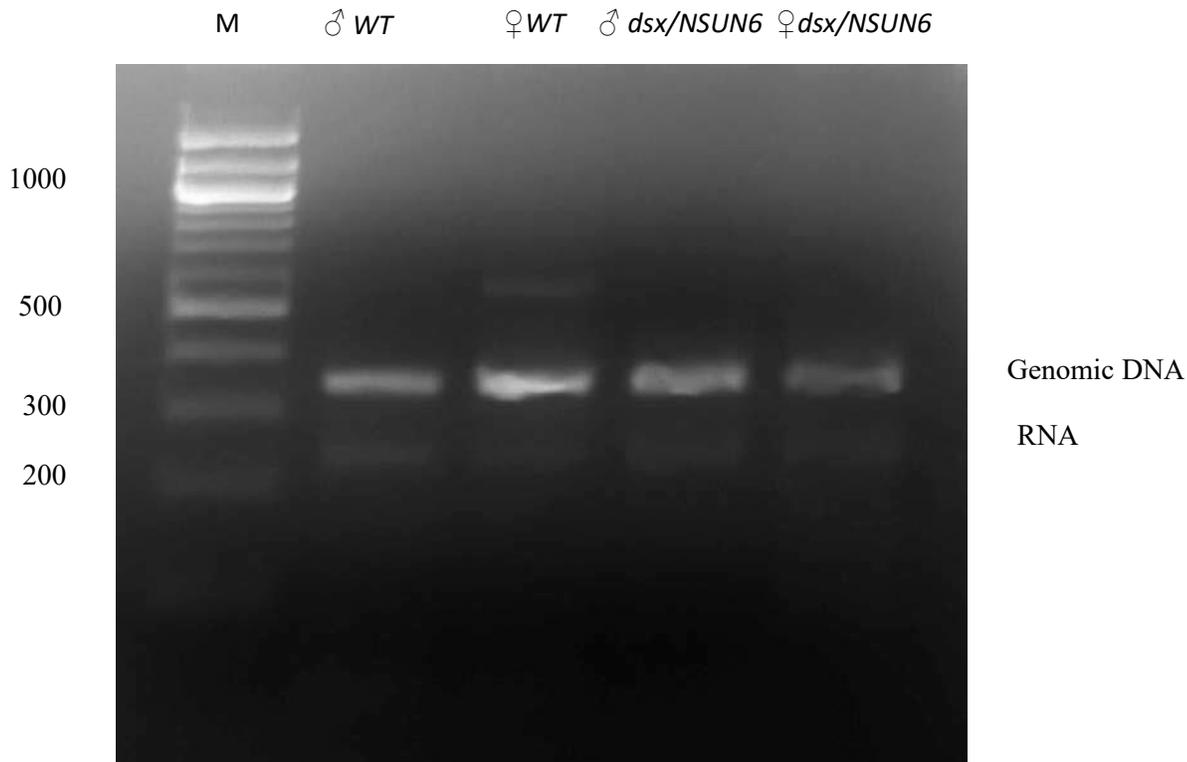
Figure 35. *Control for correct extraction of RNA from the desired mutants.* The amount of RNA is small, which caused difficulty with further extraction of the RNA. A sterile environment should be strictly followed because RNA degrades very quickly. The RNA has a very weak signal, which could suggest omissions in the techniques of extraction of the RNA, underestimation of the protocol used, or old primers. The length of the product is 202 bp, and the genomic DNA represents a significant part of the extracted components.

The alternative splicing of *sxl, tra, dsx, fru and msl-2* was done using a specific primer named *Sxl RT* along with the extracted RNA from four types of mutant: *WT*-male and female, and *dsx/NSUN6* male and female. I conducted PCR on the extracted RNA, and the product from this PCR was used with the specific primer to show the alternative splicing. All the results were inconclusive. The specific primer *Sxl RT* is made by mixing the following primers:

1   Oligo DT primer 5 pm/µl

2   *Sxl* R2 primer 10 fm/µl

3   *Msl-2-GSP* primer 10 fm/µl

4   *Fru* C2 primer 10 fm/µl

The following primers were used to visualize the alternative splicing in all genes:

1. *Sxl* F2-GTACGGCAACAATAATCCGGGTAG; this primer is localized in the sxl ORF

2. *Sxl* R2-CATCGTCGCGTCGTGGTTACAATG; this primer is localized in the main ORF

3. *Dsx* F2 EX-gcgcgagtgttcgtggatagcccACAC

4. *Dsx* MR1 EZ-CTCACCCACAATGGTCACCACGATG; this primer is localized in the male spcific isoform exon

5. *Dsx* FR2 EZ-CACATATTAGCTTAATGCTTCGGTG; this primer is localized in the female specific isoform exon

6. *FruM* F1-CCGCACAAGCGGAACATCGAAACGGA; male alternative exon

7. *FruF* F2-CCCTCAAAAGCCCGGCAACCTAAAGTTAG; female alternative exon

8. *Fru* C1R1-CTATGCGACGTCACGCTCGCCTGCGAGG; c1 first common exon

9. *Tra* For-GGATGCCGACAGCAGTGGAAC

10. *Tra* Rev-CAGACGCACTCGCTCCAGATC

11. *Msl-2* F1- CACTGCGGTCACACTGGCTTCGCTCAG

12. *Msl-2* R2-GAGGAATTGCAGGTAACTAGCCCAGGAG.

Extraction of RNA is a decisive step to obtain good results from all these primers and which is necessary to visualize the alternative splicing. However, despite all my efforts, extracting RNA was not enough to obtain the correct results. Based on the knowledge provided by other publications, a prediction could be made about the expected fragments in this experiment. Here are the predictions of the fragments:

1. *Sxl*- male (400bp); female (200bp)

2. *Tra*- male un-spliced (310bp); female spliced (190bp)

3. *Msl*-2- male spliced (200bp); female un-sliced (300bp)

4. *Dsx*- male spliced (150bp); female (600bp)

5. *Fru*- male (150bp); female (300bp)

Further work will determine the right splicing length and visualization.

The cross between *dsx-GAL4* male and *NSUN6-UAS* virgin female gave the following results: The total number of individuals was 65, of which 53 were females, 14 were pseudo-females, another 22 were sterile females.

Genotype	Number of females from the whole number of females	Number of pseudo-females	Percent of pseudo-females from the whole number of females
♀dsx/NSUN6	22 (53)	22(0)	0%
♂dsx/NSUN6	14 (53)	14(14)	26%
♀NSUN6/GAL4	17 (53)	17(0)	0%
♂NSUN6/GAL4	12 (0)	12(0)	----

The first-generation progeny was divided into four genotypes:

1. Males with genotype *NSUN6/GAL4*

2. Females with genotype *NSUN6/GAL4*

3. Males (pseudo-females) with genotype *dsx/NSUN6*

4. Females (sterile) with genotype *dsx/NSUN6*

The genotype can be represented graphically as follow:

**Distribution of genotypes in the first generation**
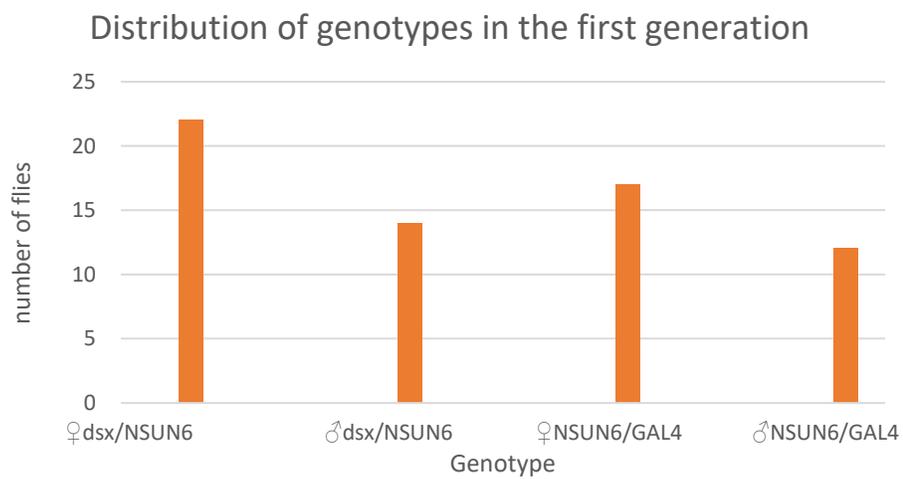
Figure 36. Representation of the number of individuals with each genotype.

The next figure represents the percentile of each genotype based on 100%:

Percentile of all genotypes to 100%
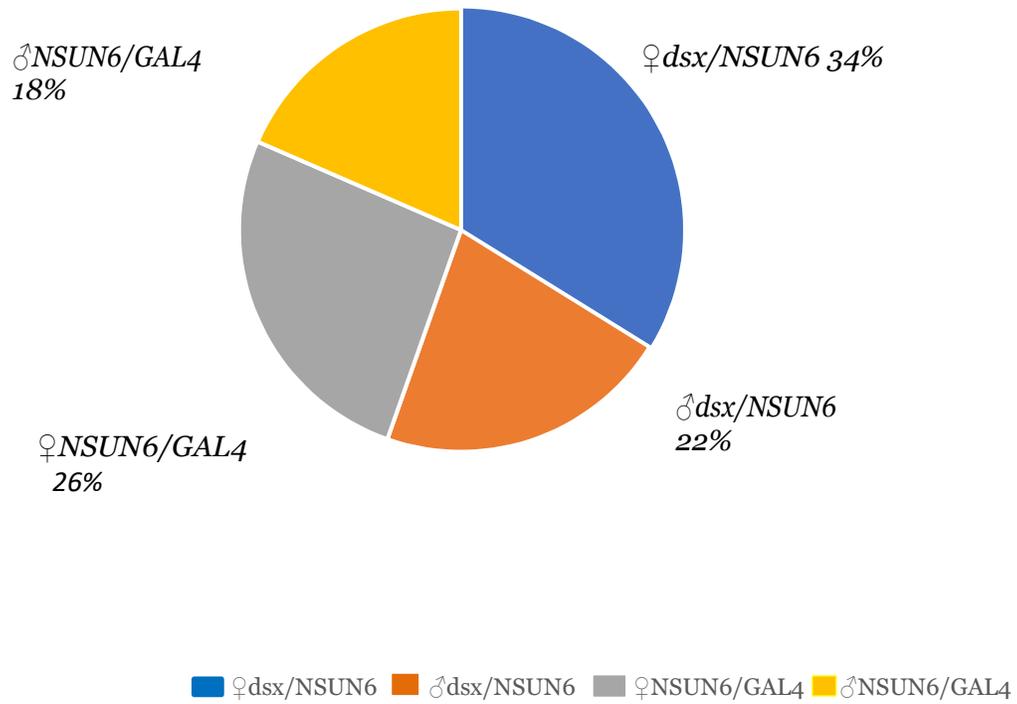
Figure 37. Percent of each genotype found in the first generation of the cross.

The most frequently encountered genotypes are:

♀*NSUN6/GAL4*-26%.

♂*NSUN6/GAL4*-18%.

♀*dsx/NSUN6* (sterile females)- 34%.

♂*dsx/NSUN6* (pseudo-females)- 22%.

One may conclude from Figure 37. that most of the progeny come from the two genotypes of interest mentioned above, and that the sum of the two genotypes accounts for 56% of the population of the first generation. This represents a challenge to maintaining this cross viable.

The number of females in the first generation includes all of the genotypically male individual which developed female genitalia. These males exhibit female behavior during copulation.
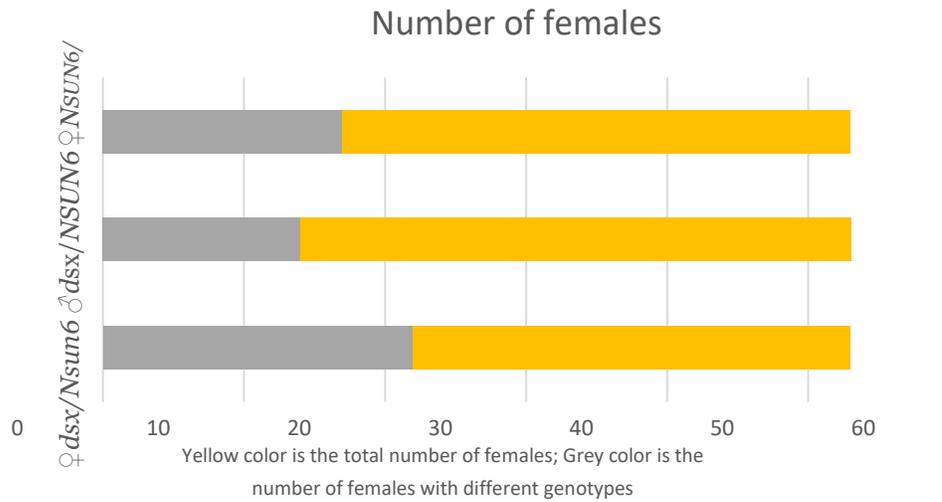
Figure 38. The females found in the first generation. 53 is the total number of females found in the first generation out of 65 individuals. 14 of the females were pseudo-females, and 22 of the whole female number were sterile females (grey color). The rest are normal females, which were responsible for the continuation of the cross.
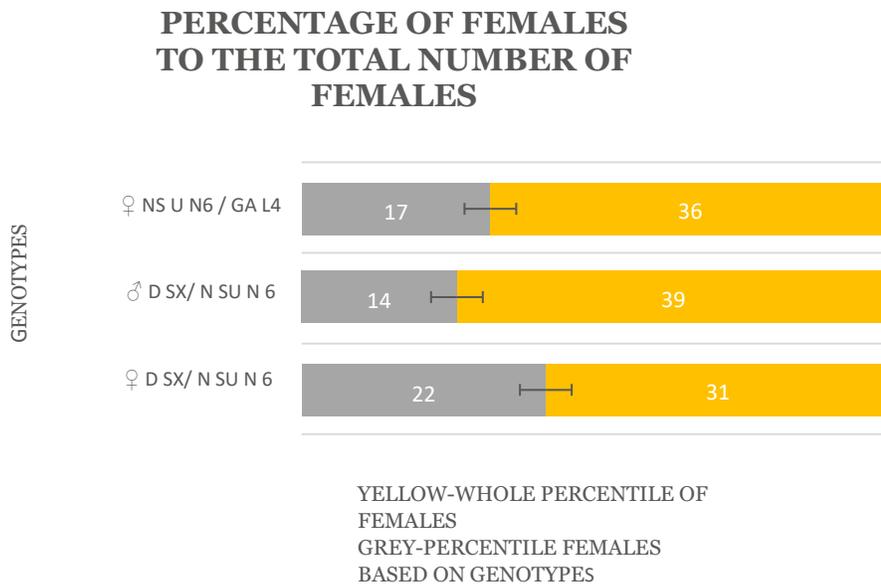
Figure 39. Percentile of different genotypes to be found in the first generation.

Most of the females are sterile, comprising 42% of the total number of females.

1. ♀*NSUN6/GAL4* (normal females)- 32.1%

2. ♀*dsx/NSUN6* (sterile females)- 41.5%

3. ♂*dsx/NSUN6* (pseudo-females)- 26.4%

The percentile numbers of the pseudo-females can be presented as following in the diagram:

♂*dsx/NSUN6*
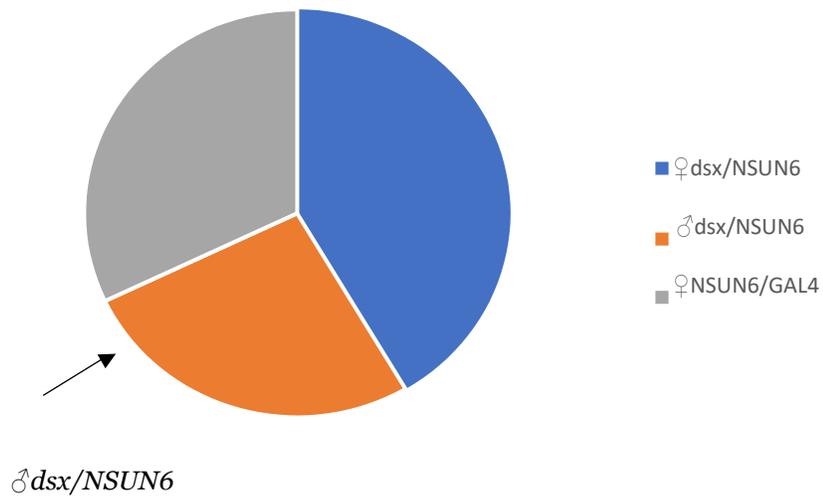
Figure 40. Percent of pseudo-females in the first generation. 26% of the females are pseudo-females. The other 74% are true females, but only ♀*NSUN6/GAL4* females can reproduce.

The cross between ♂*dsx/GAL4* and ♀ *NSUN6* gave the following pictures of individuals:
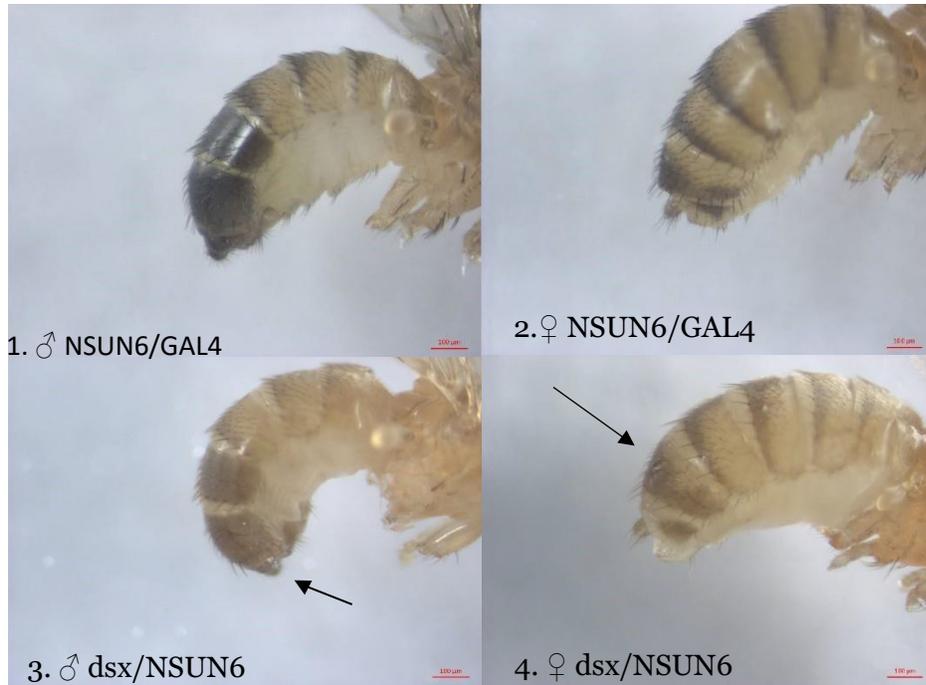
Figure 41. Representation of all phenotypically found individuals in the first generation. Picture 3 clearly shows female genitalia in male with XY chromosome; picture 4 represents the loss of pigmentation in females.

The next step was constructing the plasmid from the NSUN6 gene. The University of Zurich sent me already made constructs of this plasmid named pGW-Ha with a C-terminal tagged with a 3xHA epitope.
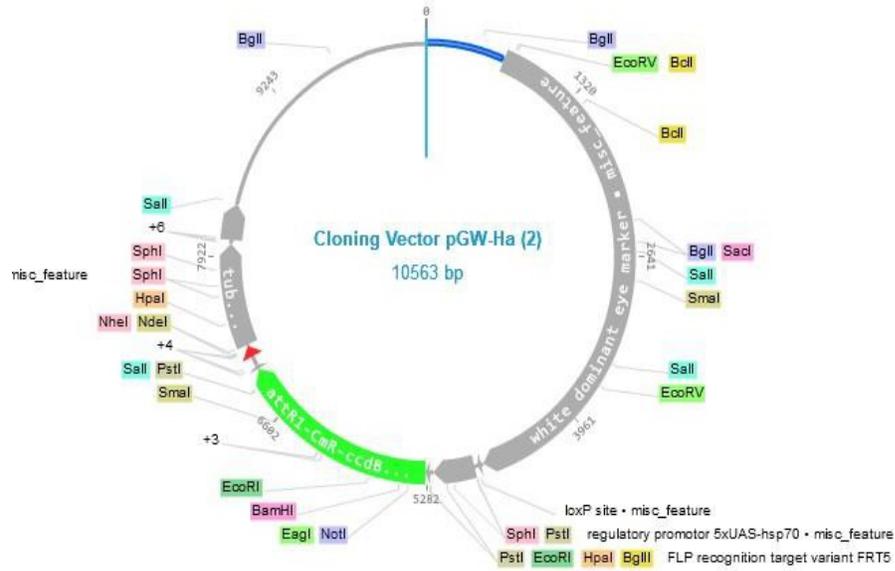
Figure 42. *pGw-3Ha vector*. The green section represents the position where NSUN6 and PUA domains will be incorporated next. The red dart tip represents the C-terminal which is tagged with 3HA sequences. This is the parent vector, which was used later.

The next step was to transform the NSUN6 and PUA domains into the plasmid, which can

be seen on fig.42 the parent plasmid.
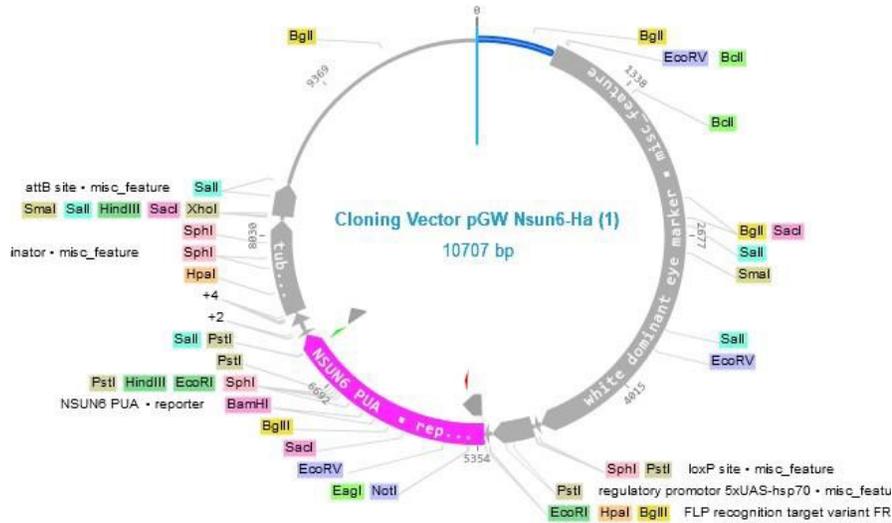
Figure 43. *pGw-NSUN6 PUA-3Ha vector*. This vector was used to create the cDNA library with different clones.

The next step was to use the new vector where some point mutation would be present. After introducing the point mutation, it was easy to examine the behavior of the mutated protein. My task was to determine whether, if the protein were binding to mRNA, where the binding site would be and how it would change after introduction of point mutations.
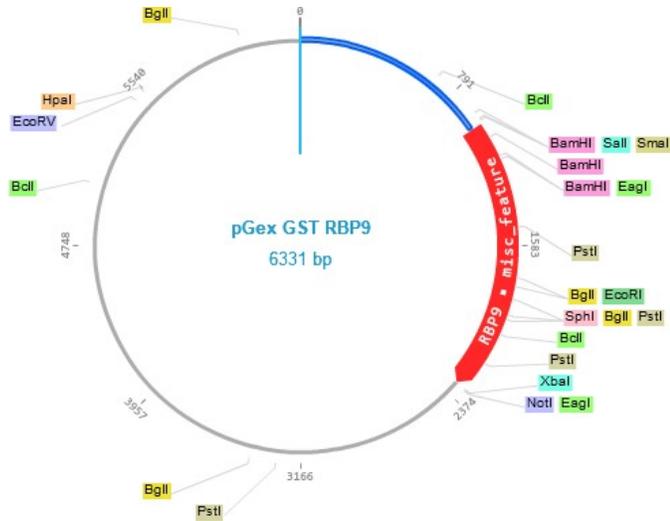
Figure 44. pGex RBP9 is the parent clone used later to install NSUN6 and PUA domains.

The NSUN6 and PUA domains were introduced into the parent vector using the Gibson assembly. A pair of primers was developed for each domain.

For the NSUN6 domain, the pairs of primers were:

1. NSUN6 F1 GA

   CAGGGGCCCCTGtcgactagtcccgggTTTTACCCAAAAACACCATTTTTGGTGAACC

2. NSUN6 R1 GA

   cacgatgaattgcggccgcTCTAGATTATATTTTCTGAAACTTCGCAATAAAAAAGCC

3. PUA F1 GA

   CAGGGGCCCCTGtcgactagtcccgggGATCCGCTGGACTCACAGCTCACCAAAG

4. PUA R1 GA

   cacgatgaattgcggccgcTCTAGATTATATTTTCTGAAACTTCGCAATAAAAAAGCC
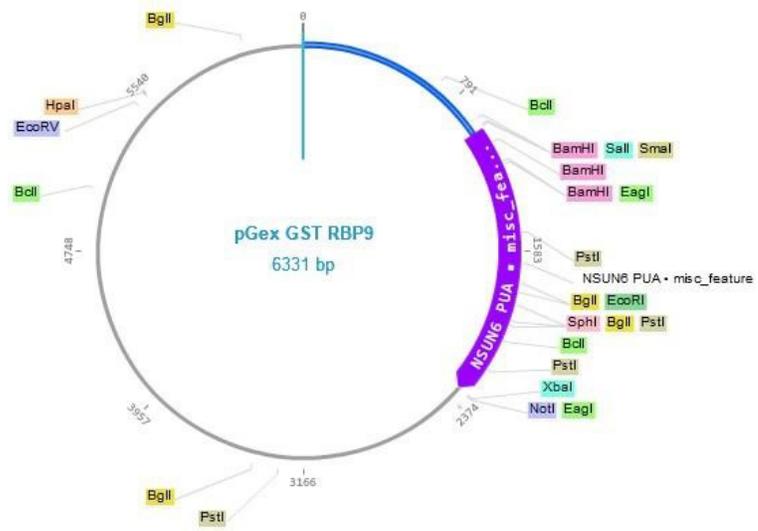
Figure 45. Transformed vector using the Gibson assembly to introduce the NSUN6 and PUA domain into the structure of the vector.
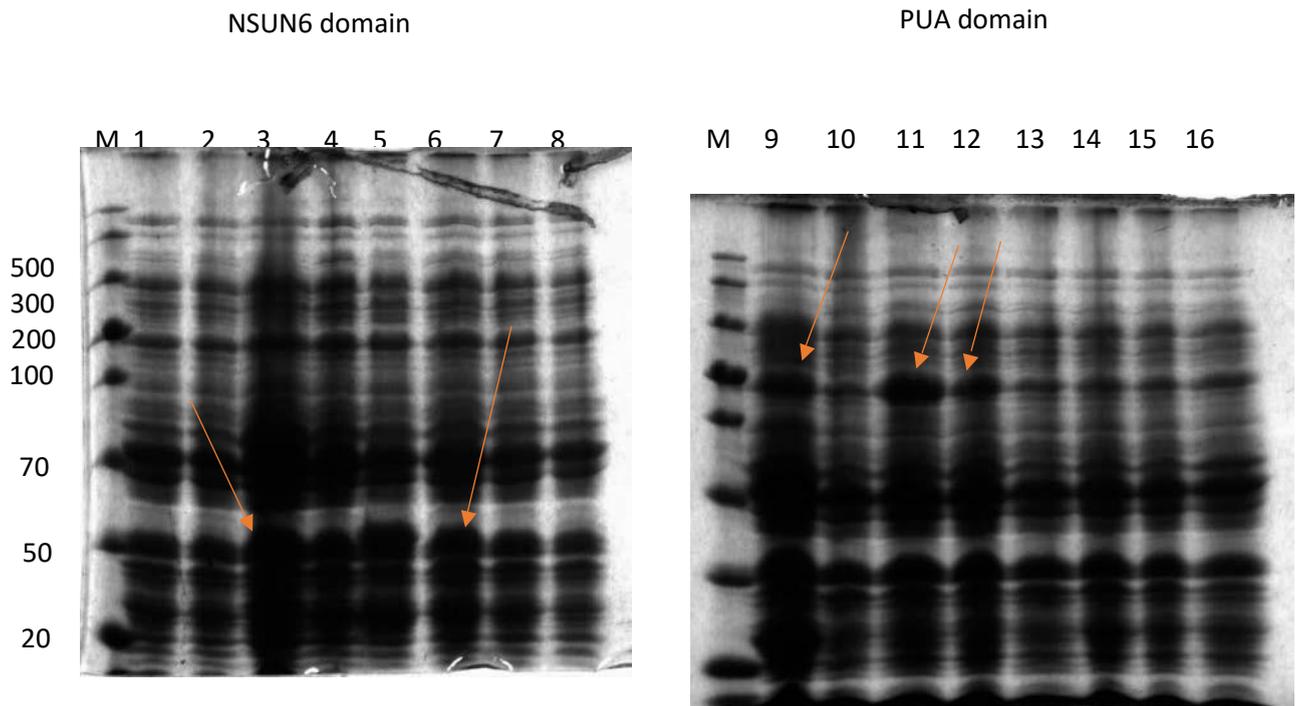
Figure 46. *Induction of protein from the bacteria*. Samples 3 and 6 from the NSUN6 domain, which responded to 49kDa molecular mass protein Anandand were the best induced samples and were chosen for mini preps. Samples 9, 11, and 12 from the PUA domain, which responded to 212 kDa molecular mass protein, were picked as the best samples for induced protein and were used to make mini preps. All samples were digested with EcoRV and Xbal restriction enzymes. These samples gave the correct fragments, and they were added to the cDNA library.

Chapter 4

Discussion and Conclusions

The purpose of this research was to observe the role of NSUN6 methyltransferase in a null mutant and an over-expressed mutant. In this research, I examined the role of methylated mRNA in the organism and the pathologies that could result from such methylation. I worked on the m5C modification in the third chromosome of *Drosophila melanogaster*. This modification made by the NSUN6 methyltransferase had profound effect on the first generation in mutants. This modification influenced the pathology in development of the ovaries in the NSUN6 overexpressed mutant, which led to further disorganization and deformity in the ovaries in the mutant, compared to the wild type and was the reason for later developed infertility. This was the reason for fewer number of individuals in the next generations.

Normally, wild type *Drosophila melanogaster* produces hundreds of progenies, but in this case the progeny of mutants was fewer. We are still uncertain as to how m5C modification interfered with formation of a female splicing pattern in SXL protein interaction with methylated mRNA of the *sxl* gene. We are also still unable to determine what the interaction was of the SXL protein in the process of forming the male splicing pattern with this modification. In addition, both the interaction of m5C modification with the alternative slicing pattern in the *sxl* gene, along with the way in which this modification influenced the alternative splicing of *tra-, dsx-, fru-* and *msl-2* genes in males and females is unclear.

We know from already published papers that NSUN6 is a protein in the family of Nol1/Nop2/SUN domain (NSUN proteins). Its role is distinguished by its ability to methylate cytosine of specific mRNA. The methylated mRNA has a profound effect on the proteins that are built up from it, and the stability of methylated mRNA is changed as well. However, we have to conduct more research to determine the interaction of between NSUN6 and all of the genes that contribute to the sexual development of *Drosophila melanogaster.*

As can be seen, methylation of mRNA by NSUN6 causes interruption in the correct structure of the *Drosophila melanogaster* ovaries, which results in infertility in later generations. Over-expressed NSUN6 mRNA in the CNS of larvae of *Drosophila melanogaster* stimulates the expression of NSUN6 protein in a specific area at the base of the brain. Over-expressed NSUN6 protein interferes with the genes responsible for sex determination of *Drosophila melanogaster.*

LIST OF
REFERENCES

[1] S. Adhikari, W. Xiao, Y. L. Zhao, and Y. G. Yang. m (6)A: Signaling for mRNA splicing. *RNA Biology,* **13**(9): 756–759, 2016.

[2] H. Amrein. Multiple RNA-protein interactions in Drosophila dosage compensation. *Genome Biology,* **1**(6): REVIEWS 1030, 2000.

[3] A. Anand, A. Villella, L. C. Ryner, T. Carlo, S. F. Goodwin, H. J. Song, D. A. Gailey, A. Morales, J. C. Hall, B. Anant, S., J. O. Henderson, D. Mukhopadhyay, N. Navaratnam, S. Kennedy, J. Min, and N. O. Davidson. Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing. CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1 complementation factor. *Journal of Biological Chemistry,* **276**(50): 47338–47351, 2001.

[4] M. Ashburner. *Drosophila: a laboratory handbook*, 1989.

[5] A. Athanasiadis, A. Rich, and S. Maas. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol,* **2**(12): e391, 2004.

[6] G. J. Bashaw and B. S. Baker. The regulation of the Drosophila msl-2 gene reveals a function for Sex-lethal in translational control. *Cell,* **89**(5): 789–798, 1997.
uj
[7] B. L. Bass. RNA editing and hypermutation by adenosine deamination. *Trends in Biochemical Sciences,* **22**(5): 157–162, 1997.

[8] F. Belanger, J. Stepinski, E. Darzynkiewicz, and J. Pelletier. Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase. *Journal of Biological Chemistry,* **285**(43): 33037–33044, 2010.

[9] L. Bell, R., J. I. Horabin, P. Schedl, and T. W. Cline. Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in Drosophila. *Cell,* **65**(2): 229–239, 1991.

[10] T. Bhalla, J. J. Rosenthal, M. Holmgren, and R. Reenan. Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nature Structural and Molecular Biology,* 11(10): 950–956, 2004.

[11] V. Blanc, N. Navaratnam, J. O. Henderson, S. Anant, S. Kennedy, A. Jarmuz, J. Scott, and N. O. Davidson. Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing. *Journal of Biological Chemistry,* **276**(13): 10272–10283, 2001.

[12] R. T. Boggs, P. Gregor, S. Idriss, J. M. Belote, and M. McKeown. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. *Cell,* **50**(5): 739–747, 1987.

[13] S. Boissel, O. Reish, K. Proulx, H. Kawagoe-Takaki, B. Sedgwick, G. S. Yeo, D. Meyre, C. Golzio, F. Molinari, N. Kadhom, H. C. Etchevers, V. Saudek, I. S. Farooqi, P. Froguel, T. Lindahl, S. O'Rahilly, A. Munnich, and L. Colleaux. Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *American Journal of Human Genetics,* **85**(1): 106–111, 2009.

[14] C. M. Burns, H. Chu, S. M. Rueter, L. K. Hutchinson, H. Canton, E. Sanders-Bush and R. B. Emeson. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature,* **387**(6630): 303–308, 1997.

[15] A. Chester, J. Scott, S. Anant, and N. Navaratnam. RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. *Biochimica et Biophysica Acta,* **1494**(1-2): 1– 13, 2000.

[16] C. Church, L. Moir, F. McMurray, C. Girard, G. T. Banks, L. Teboul, S. Wells, J. C. Bruning, P. M. Nolan, F. M. Ashcroft, and R. D. Cox. Overexpression of Fto leads to increased food intake and results in obesity. *Nature Genetics,* **42**(12): 1086–1092, 2010.

[17] S. M. Crowder, R. Kanaar, D. C. Rio and T. Alber. Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proceedings of the National Academy of Sciences USA,* **96**(9): 4892–4897, 1999.

[18] D.B. Mechanisms of alternative pre-mRNA splicing. *Biochemistry,* 2003.

[19] G. Deshpande, M. E. Samuels and P. D. Schedl. Sex-lethal interacts with splicing factors in vitro and in vivo. *Molecular and Cellular Biology,* **16**(9): 5036–5047, 1996.

[20] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, and G. Rechavi. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature,* **485**(7397): 201–206, 2012.

[21] D. Fanale, J. L. Iovanna, E. L. Calvo, P. Berthezene, P. Belleau, J. C. Dagorn, G. Bronte, G. Cicero, V. Bazan, C. Rolfo, D. Santini, and A. Russo. Germline copy number variation in the YTHDC2 gene: does it have a role in finding a novel potential molecular target involved in pancreatic adenocarcinoma susceptibility? *Expert Opinion on Therapeutic Targets,* **18**(8): 841–850, 2014.

[22] S. Farooqi, B. Sedgwick, I. Barroso, T. Lindahl, C. P. Ponting, F. M. Ashcroft, S. O'Rahilly, and C. J. Schofield. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science,* **318**(5855): 1469–1472, 2007.

[23] J. Fischer, L. Koch, C. Emmerling, J. Vierkotten, T. Peters, J. C. Bruning and U. Ruther. Inactivation of the Fto gene protects from obesity. *Nature,* **458**(7240): 894–898, 2009.

[24]   T. W. Flickinger and H. K. Salz. The Drosophila sex determination gene snf encodes a nuclear protein with sequence and functional similarity to the mammalian U1A snRNP protein. *Genes & Development,* **8**(8): 914–925, 1994.

[25]   P. Forch, L. Merendino, C. Martinez and J. Valcarcel. Modulation of msl-2 5' splice site recognition by Sex-lethal. *RNA,* **7**(9): 1185–1191, 2001.

[26]   P. Forch, O. Puig, N. Kedersha, C. Martinez, S. Granneman, B. Seraphin, P. Anderson and J. Valcarcel. The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Molecular Cell,* **6**(5): 1089–1098, 2000.

[27]   P. Forch and J. Valcarcel. Splicing regulation in Drosophila sex determination. *Progress in Molecular and Subcell Biology,* **31**: 127–151, 2003.

[28]   N. Fossat and P. P. Tam. Re-editing the paradigm of Cytidine (C) to Uridine (U) RNA editing. *RNA Biology,* **11**(10): 1233–1237, 2014.

[29]   N. Fossat, K. Tourle, T. Radziewic, K. Barratt, D. Liebhold, J. B. Studdert, M. Power, V. Jones, D. A. Loebel, and P. P. Tam. C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Reports,* **15**(8): 903–910, 2014.

[30]   Y. Fu, G. Jia, X. Pang, R. N. Wang, X. Wang, C. J. Li, S. Smemo, Q. Dai, K. A. Bailey, M. A. Nobrega, K. L. Han, Q. Cui and C. He. FTO-mediated formation of N6-hydroxymethyladenosine and N6- formyladenosine in mammalian RNA. *Nature Communications,* **4**: 1798, 2013.

[31]X. Gao, Y. H. Shin, M. Li, F. Wang, Q. Tong and P. Zhang. The fat mass and obesity associated gene FTO functions in the brain to regulate postnatal growth in mice. *PLoS One,* **5**(11): e14005, 2010.

[32]   F. Gebauer, L. Merendino, M. W. Hentze, and J. Valcarcel. Novel functions for 'nuclear factors' in the cytoplasm: The Sex-lethal paradigm. *Seminars in Cell and Devevelopmental Biology,* **8**(6): 561-566, 1997.

[33]   T. Gerken, C. A. Girard, Y. C. Tung, C. J. Webby, V. Saudek, K. S. Hewitson, G. S. Yeo, M. A. McDonough, S. Cunliffe, L. A. McNeill, J. Galvanovskis, P. Rorsman, P. Robins, X. Prieur, A. P. Coll, M. Ma, Z. Jovanovic, I.

[34]   J. Greeve, H. Lellek, P. Rautenberg, and H. Greten. Inhibition of the apolipoprotein B mRNA editing enzyme-complex by hnRNP C1 protein and 40S hnRNP complexes. *Biological Chemistry,* **379**(8-9): 1063–1073, 1998.

[35]   S. Haag, A. S. Warda, J. Kretschmer, M. A. Gunnigmann, C. Hobartner, and M. T. Bohnsack. NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA,* **21**(9): 1532–1543, 2015.

[36]   T. Haline-Vaz, T. C. Silva and N. I. Zanchin. The human interferon regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity. *Biochemical and Biophysics Research Communications,* **372**(4): 719–724, 2008.

[37]   J. C. Hartner, C. Schmittwolf, A. Kispert, A. M. Muller, M. Higuchi, and P. H. Seeburg.

Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *Journal of Biological Chemistry,* **279**(6): 4894–4902, 2004.

[38]   M. L. Hedley and T. Maniatis. Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell,* **65**(4): 579–586, 1991.

[39]   J. I. Horabin, J. I. and P. Schedl. Sex-lethal autoregulation requires multiple cis-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site. *Molecular and Cellular Biology,* **13**(12): 7734–7746, 1993.

[40]   K. Hoshijima, K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura. Control of doublesex alternative splicing by transformer and transformer-2 in Drosophila. *Science,* **252**(5007): 833–836, 1991.

[41] R. A. Hoskins, J. W. Carlson, K. H. Wan, S. Park, I. Mendez, S. E. Galle, B. W. Booth, B. D. Pfeiffer, R. A. George, R. Svirskas, M. Krzywinski, J. Schein, M. C. Accardo, E. Damia, G. Messina, M. Mendez-Lago, B. de Pablos, O. V. Demakova, E. N. Andreyeva, L. V. Boldyreva, M. Marra, A. B. Carvalho, P. Dimitri, A. Villasante, I. F. Zhimulev, G. M. Rubin, G. H. Karpen, and S. E. Celniker. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Research,* **25**(3): 445–458, 2015.

[42]   S. Hussain, J. Aleksic, S. Blanco, S. Dietmann, and M. Frye. Characterizing 5-methylcytosine in the mammalian epitranscriptome. *Genome Biology,* **14**(11): 215, 2013.

[43]   G. Jia, C. G. Yang, S. Yang, X. Jian, C. Yi, Z. Zhou, and C. He. Oxidative demethylation of 3- methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Letters,* **582**(23–24): 3313–3319, 2008.

[44]   R. Jove, and J. L. Manley. In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. *Journal of Biological Chemistry,* **259**(13): 8513–8521, 1984.

[45]   Y. Kawahara, K. Ito, H. Sun, H. Aizawa, I. Kanazawa, and S. Kwak. Glutamate receptors: RNA editing and death of motor neurons. *Nature,* **427**(6977): 801, 2004.

[46]   R. L. Kelley and M. I. Kuroda. The role of chromosomal RNAs in marking the X for dosage compensation. *Current Opinion in Genetic Development,* **10**(5): 555–561, 2000.

[47]   R. L. Kelley, I. Solovyeva, L. M. Lyman, R. Richman, V. Solovyev, and M. I. Kuroda. Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in Drosophila. *Cell,* **81**(6): 867–877,1995.

[48]   V. Khoddami and B. R. Cairns. Identification of direct targets and modified bases of RNA cytosine methyltransferases. *Nature Biotechnology,* **31**(5): 458–+, 2013.

[49]   M. Y. King and K. L. Redman. RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. *Biochemistry,* **41**(37): 11218–11225, 2002.

[50]   P. P. Lau, H. J. Zhu, M. Nakamuta and L. Chan. Cloning of an Apobec-1-binding protein that also interacts with apolipoprotein B mRNA and evidence for its involvement in RNA editing. *Journal of Biological Chemistry,* **272**(3): 1452–1455, 1997.

[51] H. Lellek, R. Kirsten, I. Diehl, F. Apostel, F. Buck, and J. Greeve. "Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex." *Journal of Biological Chemistry,* **275**(26): 19848–19856, 2000.

[52]   Y. Li, Yang, U. Dahal, X. M. Lou, X. Liu, J. Huang, W. P. Yuan, X. F. Zhu, T. Cheng, Y. L. Zhao, X. Wang, J. M. Rendtlew Danielsen, F. Liu and Y. G. Yang. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Research,* **24**(2): 177–189, 2014.

[53]   A. J. Lopez. Developmental role of transcription factor isoforms generated by alternative splicing. *Developmental Biology,* **172**(2): 396–411, 1995.

[54]   K. W. Lynch and T. Maniatis. Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer. *Genes & Development,* **10**(16): 2089–2101, 1996.

[55]   S. Maas, Y. Kawahara, K. M. Tamburr, and K. Nishikura. A-to-I RNA editing and human disease. *RNA Biology,* **3**(1): 1–9, 2006.

[56]   Manley, J. L. and R. Tacke. SR proteins and splicing control. *Genes & Development,* **10**(13): 1569–1579, 1996.

[57]I. W. Mattaj. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell,* **46**(6): 905–911, 1986.

[58]   W. Mattox, M. E. McGuffin, and B. S. Baker. A negative feedback mechanism revealed by functional analysis of the alternative isoforms of the Drosophila splicing regulator transformer 2. *Genetics,* **143**(1): 303–314, 1996.

[59]   A. E. McKee, E. Minet, C. Stern, S. Riahi, C. D. Stiles, and P. A. Silver. A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Developmental Biology*, **5**: 14, 2005.

[60]   G. Meister. *RNA biology: an introduction.* Weinheim, Wiley-VCH, 2011.

[61]L. Merendino, S. Guth, D. Bilbao, C. Martinez and J. Valcarcel. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature,* **402**(6763): 838–841, 1999.

[62]   D. P. Morse, P. J. Aruscavage, and B. L. Bass. RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. *Proceedings of the National Academy of Sciences USA,* **99**(12): 7906–7911, 2002.

[63]   J. Mouaikel, C. Verheggen, E. Bertrand, J. Tazi, and R. Bordonne. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved

methyltransferase that is localized to the nucleolus. *Molecular Cell,* **9**(4): 891–901, 2002.

[64]   O. Nayler, A. M. Hartmann and S. Stamm. The ER repeat protein YT521-B localizes to a novel subnuclear compartment. *Journal of Cell Biology,* **150**(5): 949–962, 2000.

[65]   G. Neubauer, A. King, J. Rappsilber, C. Calvio, M. Watson, P. Ajuh, J. Sleeman, A. Lamond and M. Mann. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genetics,* **20**(1): 46–50, 1998.

[66]   K. Nishikura. A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews Molecular Cell Biology,* **17**(2): 83–96, 2016.

[67]   T. Osterwalder, T., K. S. Yoon, B. H. White, and H. Keshishian. A conditional tissue-specific transgene expression system using inducible GAL4. *Proceedings of the National Academy of Sciences USA,* **98**(22): 12596–12601, 2001.

[68]   L. O. Penalva, M. F. Ruiz, A. Ortega, B. Granadino, L. Vicente, C. Segarra, J. Valcarcel, and L. Sanchez. The Drosophila fl (2)d gene, required for female-specific splicing of Sxl and tra pre-mRNAs, encodes a novel nuclear protein with a HQ-rich domain. *Genetics,* **155**(1): 129–139, 2000.

[69]   P. L. Peng, X. Zhong, W. Tu, M. M. Soundarapandian, P. Molner, D. Zhu, L. Lau, S. Liu, F. Liu, and Y. Lu. ADAR2-dependent RNA editing of AMPA receptor subunit GluR2 determines vulnerability of neurons in forebrain ischemia. *Neuron,* **49**(5): 719–733, 2006.

[70]   X. L. Ping, B. F. Sun, L. Wang, W. Xiao, X. Yang, W. J. Wang, S. Adhikari, Y. Shi, Y. Lv, Y. S. Chen, X. Zhao, A.

[71] A. G. Polson, B. L. Bass, and J. L. Casey. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature,* **380**(6573): 454–456, 1996.

[72]   L. M. Powell, S. C. Wallis, R. J. Pease, Y. H. Edwards, T. J. Knott, and J. Scott. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell,* **50**(6): 831–840, 1987.

[73]   S. M. Rueter, T. R. Dawson, and R. B. Emeson. Regulation of alternative splicing by RNA editing. *Nature,* **399**(6731): 75–80, 1999.

[74]   L. C. Ryner, S. F. Goodwin, D. H. Castrillon, A. Anand, A. Villella, B. S. Baker, J. C. Hall, B. J. Taylor and S. A. Wasserman. Control of male sexual behavior and sexual orientation in Drosophila by the fruitless gene. *Cell,* **87**(6): 1079–1089, 1996.

[75] R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich and N. Arnheim. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* **230**(4732): 1350–1354, 1985.

[76]   H. Sakamoto, K. Inoue, I. Higuchi, Y. Ono, and Y. Shimura. Control of Drosophila Sex-

lethal pre- mRNA splicing by its own female-specific product. *Nucleic Acids Research,* **20**(21): 5533–5540, 1992.

[77]H. K. Salz. Sex determination in insects: a binary decision based on alternative splicing. *Current Opinion in Genetics and Development,* **21**(4): 395–400, 2011.

[78]   G. Sarkar, S. Kapelner and S. S. Sommer. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Research,* **18**(24): 7465, 1990.

[79]   S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. Leon-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink, and A. Regev. Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell,* **159**(1): 148–162, 2014.

[80]   U. Sheth and R. Parker. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science,* **300**(5620): 805–808, 2003.

[81]J. J. Skehel and D. C. Wiley. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual Review of Biochemistry,* **69**: 531–569, 2000.

[82]   B. A. Sosnowski, J. M. Belote, and M. McKeown. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell,* **58**(3): 449–459, 1989.

[83]   Y. Sun, G. Hegamyer, and N. H. Colburn. Nasopharyngeal carcinoma shows no detectable retinoblastoma susceptibility gene alterations. *Oncogene,* **8**(3): 791–795, 1993.

[84]   M. Tian and T. Maniatis. A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell,* **74**(1): 105–114, 1993.

[85]   M. Tian and T. Maniatis. A splicing enhancer exhibits both constitutive and regulated activities. *Genes Development,* **8**(14): 1703–1712, 1994.

[86]   K. Usui-Aoki, H. Ito, K. Ui-Tei, K. Takahashi, T. Lukacsovich, W. Awano, H. Nakata, Z. F. Piao, E. E. Nilsson, J. Tomida, and D. Yamamoto. Formation of the male-specific muscle in female Drosophila by ectopic fruitless expression. *Nature Cell Biology,* **2**(8): 500–506, 2000.

[87]   J. Valcarcel, R. Singh, P. D. Zamore, and M. R. Green. The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature,* **362**(6416): 171–175, 1993.

[88]   G. Varani and K. Nagai. RNA recognition by RNP proteins during RNA processing. *Annual Review of Biophysics and Biomolecular Structure,* **27**: 407–445, 1998.

[89]   J. Wang and L. R. Bell. The Sex-lethal amino terminus mediates cooperative interactions in RNA binding and is essential for splicing regulation. *Genes Development,* **8**(17): 2072–2085, 1994.

[90]   X. Wang and C. He. Reading RNA methylation codes through methyl-specificbinding

proteins. *RNA Biology,* **11**(6): 669–672, 2014.

[91] M. E. Werner, Purta, K. H. Kaminska, I. A. Cymerman, D. A. Campbell, B. Mittra, J. R. Zamudio, N. R. Sturm, J. Jaworski, and J. M. Bujnicki. 2 '-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. *Nucleic Acids Research,* **39**(11): 4756–4768, 2011.

[92] R. Xu, J. Teng, and T. A. Cooper. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Molecular and Cellular Biology,* **13**(6): 3660–3674, 1993.

[93] Y. T. Yu and U. T. Meier. RNA-guided isomerization of uridine to pseudouridine-pseudouridylation. *RNA Biology,* **11**(12): 1483–1494, 2014.

[94] B. Zhang, A. zur Hausen, M. Orlowska-Volk, M. Jager, H. Bettendorf, S. Stamm, M. Hirschfeld, O. Yiqin, X. Tong, G. Gitsch, and E. Stickeler. Alternative splicing-related factor YT521: an independent prognostic factor in endometrial cancer. *International Journal of Gynecological Cancer,* **20**(4): 492–499, 2010.

[95] G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C. M. Huang, C. J. Li, C. B. Vagbo, Y. Shi, W. L. Wang, S. H. Song, Z. Lu, R. P. Bosmans, Q. Dai, Y. J. Hao, X. Yang, W. M. Zhao, W. M. Tong, X. J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y. G. Yang, and C. He. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular Cell,* **49**(1): 18–29, 2013.