IN VITRO AND *IN VIVO* CHARACTERIZATION OF DMHEXDC FUNCTION

by

Emir Erkol

B.S., Genetics and Bioengineering, Istanbul Bilgi University, 2015

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfilment of
the requirements for the degrees of
Master of Science

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2018

For those who forget…

ACKNOWLEDGEMENTS

This study was supported by the funds of Boğaziçi University (BAP No.7740, Proje Kodu: 15B01P7).

ABSTRACT

In vitro and *in vivo* Characterization of dmHexDC Function

Glycosylation of molecules is crucial for a wide variety of cellular processes ranging from cell-cell communication to cell differentiation and is altered in many neurodevelopmental disorders. β-hexosaminidases process glycoconjugates by hydrolyzing the catalysis of terminal N-acetyl-β-D-hexosamine residues found at the non-reducing end of glycans. Previously, three hexosaminidases were characterized in *Drosophila melanogaster* that cleave N-acetyl-β-D-glucosamine residues. Our group identified a new putative hexosaminidase, dmHexDC, which is expressed in R7 photoreceptors and pre-neuronal cells in the eye imaginal disc of *Drosophila*. In order to gain functional insight into the role of dmHexDC in neurodevelopment, I characterized the biochemical function of dmHexDC and showed that it removes both terminal N-acetyl-β-D-glucosamine and N-acetyl-β-D-galactosamine residues *in vitro*. Even though dmHexDC is able to remove both substrates, its substrate specificity is towards N-acetyl-β-D-galactosamine residues. Furthermore, a homology model of dmHexDC was generated and the output structure resembled a $(\beta/\alpha)_8$-barrel topology, which is a common structure among GH20 family proteins. The key residues that partake in the catalytic activity of dmHexDC was investigated from the homology model and glutamic acid at amino acid position 347 was identified as the general acid/base residue. Furthermore, E347A mutants displayed diminished catalytic activity towards the N-acetyl-β-D-hexosamine substrates. This suggested that integrating E347A mutation *in vivo* would create a hypomorph for dmHexDC. In order to generate fast and efficient dmHexDC mutants for functional studies *in vivo*, I generated a recombinase mediated cassette exchange (RMCE) template mutant by excising *dmHexDC* out and integrating *white* gene flanked by inverted attP sites using the CRISPR/Cas9 gene editing method. Any DNA of interest can be integrated to the dmHexDC gene region of the mutant by cassette exchange with high efficiency. Flies homozygous for the mutant allele were observed to viable and I showed that homozygous mutants lacking dmHexDC has mild neuronal loss at the ventral side of the eye imaginal discs.

ÖZET

dmHexDC fonksiyonunun *in vitro* ve *in vivo* karakterizasyonu

Moleküllerin glikozilasyonu, hücre-hücre iletişiminden hücre farklılaşmasına kadar uzanan çok çeşitli hücresel faaliyetler için önemlidir ve birçok nörogelişimsel bozuklukta değişmiş olduğu gözlemlenir. β-Hekzosaminidazlar glikanların indirgeyici olmayan ucunda bulunan N-asetil-β-D-hekzosamin kalıntıların katalizini hidrolize ederek glikokonjugatları işlemektedir. Önceki çalışmalarda, *Drosophila melanogaster*'de N-asetil-β-D-glukozamin kalıntılarını kesen üç hekzosaminidaz karakterize edilmiştir. Grubumuz genetik taramalarla *Drosophila*'nın göz imge diskinde R7 fotoreseptörlerinde ve pre-nöronal hücrelerde ifade edilen yeni bir hekzosaminidaz olan dmHexDC'yi tespit etmiştir. dmHexDC'nin nörogelişimdeki rolü hakkında fonksiyonel bir bakış açısı elde etmek için, dmHexDC'nin biyokimyasal fonksiyonunu tanımladım ve her iki terminal N-asetil-β-D-glukozamin ve N-asetil-β-D-galaktozamin kalıntılarını *in vitro* olarak hidrolize ettiğini gösterdim. Her ne kadar dmHexDC her iki substratı hidrolize edebilse de, substrat özgüllüğü N-asetil- β -D-galaktozamin kalıntılarına daha yatkındır. Ayrıca, dmHexDC'nin bir homoloji modeli üretildi ve bu model GH20 ailesi arasında ortak bir yapı olan bir $(\beta / \alpha)_8$ -fıçı topolojisine benzediği gösterildi. dmHexDC'nin katalitik aktivitesine katılan anahtar kalıntılar homoloji modelinden araştırılmış ve 347. pozisyondaki glutamik asit, genel asit / baz kalıntısı olarak tanımlanmıştır. Ayrıca, E347A mutantları N-asetil- β -D-hekzosamin substratlarında azalan katalitik aktivite sergilemiştir. Bu, E347A mutasyonunun *in vivo* entegrasyonunun, dmHexDC için bir hipomorf oluşturacağı ileri sürdü. *In vivo* fonksiyonel çalışmalarına hızlı ve verimli dmHexDC mutantları üretmek için, dmHexDC'nin çıkarılması ve CRISPR / Cas9 gen düzenleme yöntemi kullanarak ters attP dizileri ile çevrili *white* genin entegre edilmesi yoluyla rekombinaz aracılı kaset değişimi (RMCE) şablon mutantı ürettim. İlgilenilen herhangi bir DNA, mutantın dmHexDC gen bölgesine yüksek verimlilikle kaset değişimi ile entegre edilebilir. Mutant alel için homozigot olan sineklerin canlı olduğu gözlendi ve dmHexDC içermeyen homozigot mutantların, göz imge disklerinin ventral tarafında hafif nöronal kayıplara sahip olduğunu gösterdim.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

g	Gram
kb	Kilobase
L	Liter
ml	Milliliter
mm	Millimeter
M	Molar
ng	Nanogram
nm	Nanometer
rpm	Revolutions per minute
v	Volume
w	Weight
µg	Microgram
µm	Micrometer
µl	Microliter

LIST OF ACRONYMS/ABBREVIATIONS

CNS	Central Nervous System
DNA	Deoxyribonucleic Acid
GFP	Green Fluorescent Protein
EGFR	Epidermal Growth Factor Receptor
FDL	Fused Lobes
FRT	FLP Recombination Target
Gal	Galactose
GalNAc	N-acetylgalactosamine
GFP	Green Fluorescent Protein
GlcNAc	N-acetylglucosamine
Glc	Glucose
Man	Mannose
MF	Morphogenetic Furrow
PBS	Phosphate Buffered Saline
PFA	Paraformaldehyde
pH	Power of Hydrogen
pWS	pWhiteStar
RMCE	Recombinase Mediated Cassette Exchange
RNA	Ribonucleic Acid
RNAi	RNA Interference
RT-PCR	Reverse transcription polymerase chain reaction

1. INTRODUCTION

Information is the resolution of any uncertainty. When observed and understood by a cognitive receiver, the phenomena is termed as knowledge yet the abstract concept of information does not have to exist for the purpose of being transmitted to a receiver. For thousands of years philosophers and scientists are in the process of discovering and understanding information and expanding the collective knowledge of humanity and it was discovered that the nervous system was responsible for this sophisticated interaction between the animal and information. When observed, the nervous system is composed of the immensely complex harmony of various types of differentiated neurons, receiving and transmitting information to one another. How those cells are differentiated from a single cell is determined by the information acquired from outside and within that cell in a complex system. Complex systems as such contain more uncertainty and entropy which requires an expanded amount of knowledge to comprehend. Thus, models with lower information entropy are observed in order to gain knowledge of a particular mechanism and insight regarding the more complex systems. *Drosophila melanogaster* is a biological system with less information entropy and more manipulability utilized to understand biological mechanisms regarding the dynamics of molecular information leading to broader perspective on specific functions. Moreover, the compound eye of the organism is a well-studied organ to explore key elements of neuro-development. Among the many pathways that govern neuro-development, glycosylation is the most complicated modification pathway in terms of structure diversity and enzymatic processes to generate molecular information that can only be readable by proteins.

1.1. Glycosylation machinery

Glycosylation is the process of covalently attaching sugars to molecules thus utilizes the most abundant organic compound found on earth and is responsible for generation of any carbohydrate containing molecule, also termed as glycans (Scott and Panin, 2015). This universal machinery which assembles and disassembles carbohydrates consists of many elements such as glycosyl hydrolases, chaperones, transporters, sugar donors and acceptors, glycosyltransferases, adapter molecules and other molecules necessary for the modification of proteins or lipids with carbohydrates and various organelles takes parts in operation this complex machinery(van Kooyk et al., 2013) (Figure 1.1).

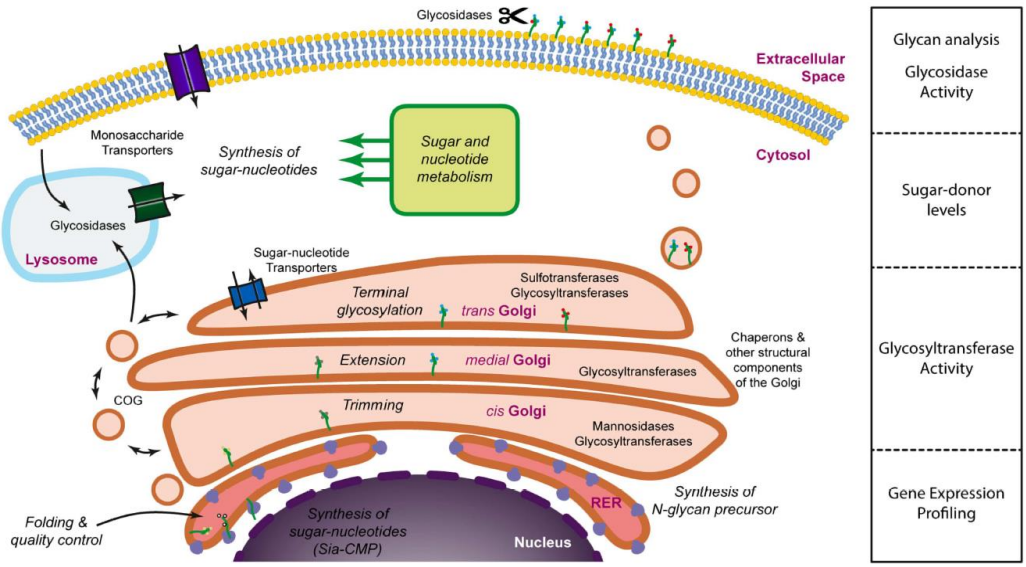

Figure 1.1 Diverse localization and function of eukaryote glycosylation machinery elements in cellular compartments. Each component has specialized activity (left) (van Kooyk et al., 2013).

The products of the glycosylation machinery are amazingly diverse in their structures. Carbohydrates consist of monosaccharides. These monosaccharides can be linked in four ways (1→2, 1→3, 1→4, 1→6 linkages) with alpha and beta configurations. The anomeric state of monosaccharides and oligosaccharide branching also contribute immensely to the spatial diversity of carbohydrates. Additional molecules on glycans may also be present as a result of phosphorylation, acetylation, sulfation, and methylation. In theory, the complexity of mammalian carbohydrate structure surpasses the primary structure of proteins (Table 1.1) (Werz et al., 2007) (Table 1).

Table 1.1 Theoretical diversity of oligosaccharides are more complex than peptides and nucleotides (adapted from Werz et al., 2007).

Oligomer	Nucleotide	Peptide	Carbohydrate
1	4	20	20
2	16	400	1360
3	64	8,000	126.080
4	256	160.000	13.495.040

5	1024	3.200.000	1.569.745.920
6	4096	64.000.000	192.780.943.360

1.1.1. Glycosylation Pathway

Addition of carbohydrate moieties covalently to proteins is the most characterized type of glycosylation. In addition, glycosylation is the most common post-translational modification (PTM) in eukaryotes (Zhu et al., 2015). Protein modification by glycosylation has shown to affect the biophysical properties of proteins thus controlling the function further than the information that is provided on the genome of the organism by the alteration of kinetic, structural and thermodynamic properties (Zhu et al., 2015). Thus, it would not come as a surprise that the role of protein glycosylation is critical in cellular activities such as protein folding, stability and sorting, protein-protein interactions, signal transduction, cell-cell communications, immunity, and cell fate specification (Ungar, 2009) The studies on synthesis of glycoprotein type glycans mostly focus on N and O-linked glycosylation. Their name refers to the glyosidic linkage to the amino acids they are attached to. Between these two processes, N glycosylation is the most characterized event.

1.1.2. Biosynthesis of N-Glycans

N-glycosylation is most often described as a process that is conserved in three domains of life. It is well known that the vast majority of N-glycan biosynthesis occurs in the endoplasmic reticulum. In prokaryotes N-glycan biosynthesis takes place at the periplasmic space of the organism and the two processes are thought to be homologous (Schwarz and Aebi, 2011).

At the plasma membrane of the ER, sugars are donated on a polyisoprenoid lipid termed dolichol. Phosphorylation of dolichol allows it to be used as a docking site for oligosaccharide assembly (Burda and Aebi, 1999). Oligosaccharide assembly occurs by the sequential addition of monosaccharides from the nucleotide sugar donors on the cytosolic surface of the ER (Andrade, 1993). Generally, three glucose (Glc), nine mannose (Man), and two *N*-acetyl glucosamine (GlcNAc) residues, which are essential for the processes, are added on the dolichol pyrophosphate carrier. The oligosaccharides that have been assembled on the plasma membrane are to be covalently attached to a protein which is located at the lumen of the ER. For this reason an enzyme called dolichol flippase causes the dolichyl pyrophosphate and oligosaccharide structure to migrate transversely from the cytosolic

surface to the luminal surface (Schwarz and Aebi, 2011). The transfer of oligosaccharides to proteins is catalyzed by the oligosaccharyltransferase (OST) enzyme. It is important to note that OST is a quality controller for proper protein folding. If OST subunits cannot recognize a glycosylation competent formation, the glycosyltransferase activity will be inhibited for that particular polypeptide (Schwarz and Aebi, 2011). Usually co-translationally, OST recognizes the Asn-X-Ser/Thr (X is any amino acid excluding proline) motif and OST activity results in a *N*-glycosidic linkage between oligosaccharide and amide group at the side chain of asparagine (Harada et al., 2013). The protein than translocates to the Golgi apparatus, where further trimming and extension of oligosaccharide branches occur in a more specialized manner. In order to act sequentially on a previously modified glycan, *cis*, *medial* and *trans* compartments of the Golgi have glycan modifying enzymes. The produced glycans are build up in trans Golgi compartment and distributed to their proper subcellular localizations. The specific repertoires of glycosidases and glycosyltransferases at the Golgi is responsible for the diversity of glycan structures among cells, tissues and developmental stages of the same organism.

1.1.3. Biosynthesis of O-glycans

O-glycosylation is a common process in all metazoans and it is much more diverse in the initiation step with respect to N-glycosylation (Mcmahan and Wigley, 2015). The mechanisms of O-glycosylation is less understood than N-glycosylation and there is no decisive report on initiation of the O-glycosylation event. The current opinion is that biosynthesis of O-glycans begins at the *cis* Golgi (Gill et al., 2010) and ER (Gill et al., 2013). Even though the O-linked glycosylation motif is not as clear as the N-glycosylation motif, mass spectrometry data analysis of O-linked glycans showed that O-glycosylation is most likely to occur in proline rich regions in contrast to N-glycosylation. Especially the presence of proline at +3 and/or -1 positions around the N-glycosylation site, Asparagine-X-Serine/Threonine. In addition, cysteine and amino acids containing bulky side chains have inhibitory affects for O-glycosylation (Hema Thanka Christlet and Veluraja, 2001).

O-glycosylation begins with the covalent attachment of Mannose, Fucose, GlcNAc, Glucose, GalNAc, or Xylose to the OH group of serine or threonine residues via O-glycosidic bond on the target protein, however, the most common monosaccharide that initiates the O-glycosylation process is GalNAc (Mcmahan and Wigley, 2015). GalNAc is transferred to serine and/or threonine residues by ppGalNAc transferase from nucleic acid

sugar carriers and this process is required for the proper development of mouse and fly embryos (Gerken et al., 2011). Proteins that are modified with an α-linked GalNAc or glycosaminoglycans (GAGs) modified with O-Xyl are termed mucins and usually O-glycans are also referred to as mucins. O-glycans are modified immensely during their maturation in the Golgi. The O-glycan matured in comportments of the Golgi are generally secreted outside of the cell or localized to the plasma membrane. In some cases nuclear and cytoplasmic proteins are O-glycosylated (Comer and Hart, 2000).

1.1.4. Biosynthesis of Glycosphingolipids

Similaremb to O- and N-glycosylated proteins glycosphingolipids (GSL) are also synthesized in the ER-Golgi complex and they have high structural variance. *Drosophila melanogaster* glycosphingolipids are structurally more different than mammalian glycosphingolipids and made up primarily of Manβ1-4Glcβ1-ceramide core. This core can be further elongated with the addition of GalNAc, GlcNAc, Galactose and phosphoethanolamine or their combinations. Two genes named *egghead (egh)* and *brainiac (brn)* were identified and their products were shown to function in the biosynthesis of GSL. Loss of function in both genes resulted in disruption of cell to cell adhesion and hypertrophy in the nervous system similar to phenotypes that occur when Notch and epidermal growth factor (EGF) receptors have disrupted interactions with their ligands. The mutants were embryonic lethal and dorsal fates could not be established (Chen et al., 2007). Further investigations showed that the gradient of EGF receptor ligand Gurken was controlled by glycosphingolipids, further confirming that glycosphingolipids are important factors in cell to cell signaling and morphogen distribution (Pizette et al., 2009).

1.2. *Drosophila* as a model for glycosylation during neurodevelopment

Drosophila melanogaster (common fruit fly) is a powerful model organism in terms of availability of sophisticated molecular and genetic tools. Moreover, its relatively cheap husbandry and its ability to produce large number of offspring has been assessed to be very desirable for a model organism.

Drosophila melanogaster has a rapid life cycle. Embryogenesis takes one day after fertilization. Development of the fly is then followed by 1st, 2nd, and 3rd larval stages termed as instars. Development of imaginal discs, which are fated to become functioning organs in adult stages, are progressed during the 3 instars of larval development (Figure 1.2).

Therefore a functional disruption in earlier stages is likely to be observed as phenotypes in the adult flies. The larval stage is completed after 5 days and puparium formation occurs. The event of metamorphosis happens during the pupal stage by shifts in gene expression and the adult fly hatches 4-5 days after the initiation of the pupal stage. Thus it takes approximately 10 days to obtain an adult fly (Hales et al., 2015).

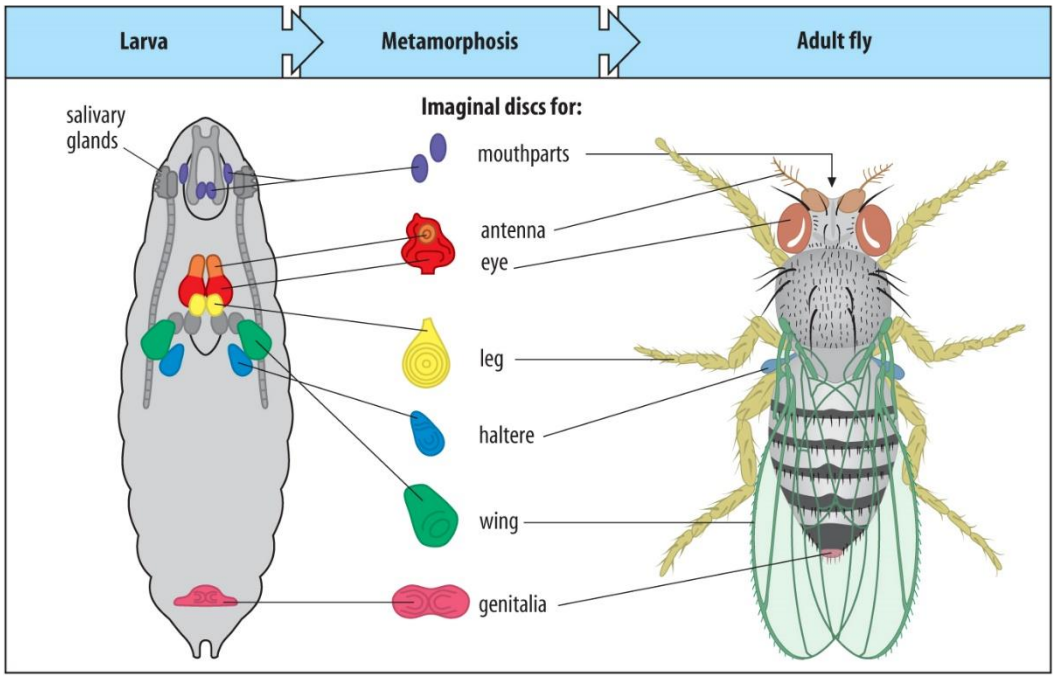

Figure 1.2. Overall scheme of developing imaginal organs of larvae to functioning organs in adult. Imaginal organs gain their function after metamorphosis. Thus, developmental defects in these organs in earlier stages of development can manifest themselves in adult phenotypes (taken from Nienhaus, 2013).

Advancements of sequencing technology has unraveled that 75% of human disease causing genes have homologues in fruit flies and half of *Drosophila* proteins have mammalian orthologues with decreased redundancy of function. This increased the interest in characterization of complex molecular pathways in *Drosophila* (Pandey and Nichols, 2011). Furthermore, the *Drosophila* nervous system is less complex compared to mammals yet operates on the same principles. Therefore, dissecting the components that make up a functional nervous system in *Drosophila* is deemed much more efficient and the sophisticated features of *Drosophila* have allowed the study of complex biochemical mechanisms of glycosylation in the nervous system.

One of the pioneering works on glycosylation in neurodevelopment was on the regulation of Notch signaling during eye development by a glycosyltransferase unique to *Drosophila* called Fringe, which was identified from a genetic screen (Irvine, 1999). Notch is a heavily glycosylated cell surface protein and its proper function is vital to neurogenesis and cell fate specification (Zhou et al., 2010). It was shown that addition of O-linked GlcNAc to *O*-linked fucose present on EGF-like repeats of Notch inhibits its interaction with its ligand Serrate and promotes binding to Delta (Stanley and Okajima, 2010). Mirror, a member of the Iroquois family proteins represses the expression of fringe in the dorsal part of the eye imaginal disc (Mazzoni et al., 2008). Expression of *Mirr* is blocked at the ventral side of the eye imaginal disc and this results in a distinction of Notch in terms of ligand selection. This results in proper orientation established by a 45° rotation of photoreceptors and an angle of 90° is observed between the dorsal and ventral photoreceptor cells (Cho and Choi, 1998). Loss of fringe activity disrupts this planar cell polarity and results in diminished function of the adult eye. Later on, the inhibition of O-fucose addition to Notch was shown to completely disrupt Notch signaling (Okajima and Irvine, 2002).

The *Drosophila melanogaster* glycome is mostly abundant of paucimannosidic glycans unlike mammals, which exhibit abundance of complex and paucimannosidic glycans. Fused Lobes (Fdl) is responsible for the trimming of N-glycans and generates paucimannose structures. It was shown that fdl is a hexosaminidase that cleaves N-linked GlcNAc from glycoconjugates. Mutations in fdl resulted in fusion of mushroom body β lobes, disrupting the proper function of the *Drosophila* central nervous system (CNS) (Léonard et al., 2006).

Abundance of paucimannose glycans suggests that glycoconjugates are trimmed before leaving the ER-Golgi network. GlcNAc is a building block for the elongation of N-glycan oligosaccharides and therefore to obtain a paucimannosidic glycome, GlcNAc residues are trimmed (Hagen et al., 2009).

Hexosaminidases are not only required for the proper function of the CNS. It was shown that they are also required for photoreceptor function. In order for the *Drosophila* eye to function properly, maturation of Rhodopsin 1 (Rh1), a G-protein coupled protein that absorbs photons and initiates visual transduction, is necessary. Rh1 is present in 6 photoreceptor cells (R1-R6) except R7 and R8 (Kumar and Ready, 1995). The proper maturation of Rh1 requires its deglycosylation and besides other glycosidases,

hexosaminidase 1 (Hexo1) and fdl is involved in the maturation and secretion of Rhodopsin 1 by cleaving its terminal GlcNAc residues (Rosenbaum et al., 2014).

Recent studies gave rise to an increased interest in hexosaminidases in development due to the fact that almost all glycoproteins and the majority of glycoconjugates carry hexosamine residues.

1.3. Hexosaminidases

Hexosaminidases catalyze the hydrolysis of terminal N-acetyl-hexosaminyl (2-acetamido-2-deoxy-β-d-glycosides) residues in glycoconjugates such as glycosphingolipids, glycoproteins and GAGs. The Carbohydrate-Active Enzymes database (CAZy) classifies hexosaminidases in GH3, GH18, GH20, GH22, GH56, and GH84 families according to amino acid similarity (Davies, Gideon J & Sinnot, 2008). Although diverse in amino acid structures, hexosaminidases share similar mechanistic principles. Generally, hexosaminidases utilize a configuration retaining double-displacement mechanism (Maier et al., 2003) meaning that they catalyze the hydrolysis reaction by a substituted enzyme intermediate, which occurs by group exchange between the substrate and the enzyme, although slight variation is observed in hydrolysis reactions. Members of GH3, GH22 families utilize the Koshland double-displacement mechanism where members of GH18, GH20, GH56, and GH84 families cleaves glyosidic bonds with a substrate-assisted mechanism (R. Zhang et al., 2010). The difference between both mechanisms is that the substrate plays a vital part during substrate-assisted mechanism and it is more common among hexosaminidases.

In the substrate-assisted mechanism used by hexosaminidases, the acetimido group of the substrate is polarized by a carboxyl residue. This results in the increased nucleophilicity that causes an attack on the acetimido group of the substrate. Consequently, an oxazolinium ion intermediate is formed. A general acid/base residue of the hexosaminidase catalyzes the protonation of a glycone group and deprotonation of a water molecule. Deprotonation of the water molecule causes a nucleophilic attack at the anomeric center of the substrate, resulting in the cleavage of the glycosidic bond (Mark et al., 2001).

The most characterized hexosaminidases reside in the GH20 group. The main reason for the interest for the GH20 group came from two human glycoside hydrolases named hexosaminidase A (HexA) and hexosaminidase B (HexB). HexA and HexB were given

special attention due to their link with lysosomal storage disorders (LSD) (Wenger et al., 2003). Both proteins reside in lysosomes where they catalyze the hydrolysis of terminal GalNAc and GlcNAc residues, glycoconjugates involving hydrophobic GM2 gangliosides with the aid of GM2 activator protein (GM2A), which binds and presents gangliosides to hexosaminidases in soluble form for their proper degradation (Wenger et al., 2003). HexA protein is a heterodimer composed of an α subunit expressed from *HEXA* gene and a β subunit expressed from *HEXB* gene. HexB protein is a homodimer composed β subunits expressed from the *HEXB* gene (Fernandes et al., 1997). Mutations in the *HEXA* gene can cause Tay-Sachs disease and mutations in the *HEXB* gene can cause Sandhoff disease (Dastsooz et al., 2018). Both diseases are the best known LSD due their clinical features, which are described as cherry-red spot on the retina, blindness, mental retardation and death before the age of four. Loss of function of GM2 activator protein and/or α and β subunits of hexosaminidases causes the generation of GM2 gangliosides, which are most abundant in neurons of the CNS, accumulate inside the lysosomes and cause neuronal death by apoptosis (Sandhoff and Harzer, 2013).

In 2010, another member of the GH20 family, hexosaminidase D was characterized in humans and mouse. It was shown that its substrate specificity is towards GalNAc with a pH optimum of 5.5 at 37°C. Moreover, the subcellular localization of HEXDC was nucleocytoplasmic instead of lysozymes, which is generally the case for human hexosaminidases (Gutternigg, 2010). Mechanism of human hexosaminidase D action was also uncovered as configuration retaining double-displacement substrate-assisted mechanism, where the aspartic acid at 148th position acts as the polarizing residue and glutamic acid at 149th position is the general acid/base residue (Alteen et al., 2016). It was shown that hexosaminidase D activity was elevated in rheumatoid arthritis and the authors hypothesize that the elevated exoglycosidase activity may be a threat to homeostasis by excessive matrix degradation (Pásztói et al., 2013).

1.3.1. A novel *Drosophila* Hexosaminidase: dmHexDC

For the purpose of isolating genes that take part in cell fate specification during eye development of *Drosophila*, Dr. Çelik constructed an enhancer-trap library using a PiggyBac transposable element during her post-doc study. From the screening of those lines, dmHexDC was isolated and analysis of dmHexDC in third instar larva eye imaginal discs

showed its expression in R7 photoreceptors (Öztürk, 2010). Later, dmHexDC was observed at the preneuronal cells of the eye imaginal disc (Kıral, 2015) by GFP-tagged overexpression line generated by BAC recombineering (Kaçmaz, 2013).

The *dmHexDC* gene is located at the right arm of the 3rd chromosome and resides between 3R:18,199,857-18,215,001 positions (has a length of 15,144 bp). It has five exons and four introns, in which intron 2 is much larger (≈ 12 kb) than the rest of the annotated sequences. A small and uncharacterized gene with one exon, *CG44158*, resides in intron 2. According to Eukaryotic Promoter Database putative promoter elements are present just upstream of exon 1, exon 2 and CG44158 (Dreos et al., 2017) (Figure 1.3). Two transcription products of *dmhexdc* were annotated. Exon 2 is not present in the first transcript product and exon 1 is not present in the second transcript product.

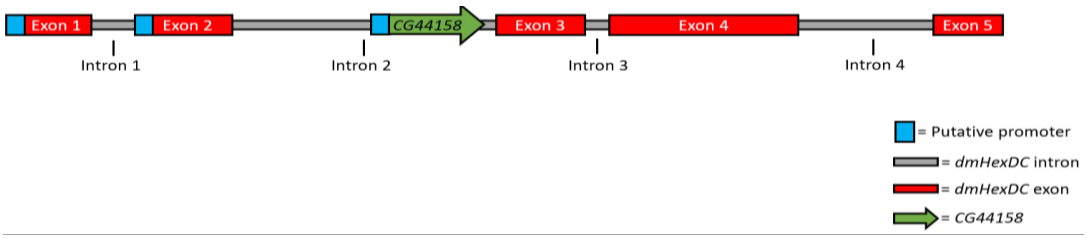

Figure 1.3. Schematic representation of *dmHexDC* gene region. Turquoise boxes are the putative RNA polymerase II binding sites. *dmHexDC* exons are represented in red rectangles. Gray region is *dmHexDC* introns. Green arrow indicates the single exon of *CG44158* and its orientation.

The coding sequence (CDS) of dmHexDC starts at exon 3 and ends in exon 4. *dmHexDC* encodes a 708 amino acid long protein that has a putative transmembrane domain at the N-terminus and a conserved putative catalytic domain between 188th and 517th positions. In the putative catalytic domain His/Asn-Xaa-Gly-Ala/Cys/Gly/Met-Asp- Glu-Ala/Ile/Leu/Val motif is present, which is highly conserved among GH20 hexosaminidases (Gutternigg, 2010).

By using the overexpression line tagged with GFP, it was shown that dmHexDC is localized to the endoplasmic reticulum (ER), lysosome, Golgi, early and late endosome by using antigen specific antibodies (Basdag, 2018) in 3rd instar eye imaginal disc.

By analyzing a protein trap line that disrupts the transcription of dmHexDC in 3rd instar eye imaginal discs with immunohistochemistry (IHC), our group showed that loss of function results in lipid accumulation inside the lysosomes of pre-neuronal cells, which resembles LSD. Lipid accumulation consequently causes the enlargement of lysosomes and

this results in the activation of the apoptosis pathway. Interestingly, the dying cells secrete wingless (wg). Wg serves as compensatory proliferation signal to the neighboring cells and promotes mitosis. Proliferating cells do not undergo cell fate specification, thus tumor-like growth and neuronal loss is observed in the eye disc (Kıral, 2015).

1.4. CRISPR system

The CRISPR/Cas9 method, now considered as an irreplaceable tool for genome engineering, was first discovered as a mechanism for bacterial adaptive immunity (Rath et al., 2015). Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated proteins (Cas) are a part of a RNA-guided bacterial defence mechanism that encodes the exogenous DNA in host DNA as a genetic memory and uses the genetic element against the invader (Rath et al., 2015). Thus similar to eukaryotic RNA interference (RNAi), another genetic interference system, CRISPR, provides the host with a defensive mechanism and is therefore considered analogous to but not homologous to RNAi (Sontheimer, 2011).

CRISPR is a part of the bacterial and archaeal genomes. Approximately 50% of the sequenced bacteria and 90% of the sequenced archaea contain the CRISPR locus. The locus that contains CRISPR elements harbors the Cas proteins and the CRISPR array (Hille and Charpentier, 2016). The CRISPR array is usually preceded by an AT-rich sequence and is known as leader sequence. The leader sequence contains the transcription signal for the CRISPR locus. The CRISPR locus contains nearly identical repeats of DNA sequence. The length of repeats ranges between 23-55 base pairs (bp). Regions called “spacers” reside between the repeats. Spacer regions usually have a range of 21-72 bp and the DNA sequence of this region is the same with an exogenous DNA, which is usually (almost 90%) a virus that has invaded the same or a very similar bacterial species in the past (Shmakov, Sergey; Sitnik, Vassilii; Makarova, Kira; Wolf, Yuri; Severinov, Konstantin; Koonin, 2017). The integration of DNA sequences as spacer regions is accomplished by the products of the *cas* genes. *Cas* genes generally flank the leader sequence region and integrate the spacer sequence from foreign DNA to the leader sequence. Cas1 and Cas2 complexes target the foreign DNA, termed as protospacer, and mediates its excision and integration to the leader sequence. The protospacer sequence is just upstream of a protospacer adjacent motif sequence (PAM). Integration of the protospacer to the genome without the PAM is the first step of the CRISPR defence mechanism (Figure 2) and this step ultimately results in immunization to foreign DNA. The second step consist of the transcription and maturation

of the CRISPR array (Bakr Shabbir et al., 2016). Maturation of the pre-CRISPR RNA is accomplished by different means. In the most well-known and utilized class II system of CRISPR-mediated defense, the repeat sequences on pre-CRISPR RNA hybridizes *trans*-acting CRISPR (tracr) RNA. *trans*-acting CRISPR RNA has a secondary structure that causes direct binding to Cas9 endonuclease (Huai et al., 2017). The binding tracr RNA to pre-crRNA results in the formation of double stranded RNA. Consequently, the double stranded RNA is cleaved by the host RNase III enzyme, releasing the tracr RNA-crRNA-Cas9 complex. Mature crRNA then acts as a "guide" for the Cas9 endonuclease against a foreign DNA (Deltcheva et al., 2011). Cas9 endonuclease consists of 6 domains. 2 of them are nuclease domains called HNH and RuvC. Near the C-terminus of the protein, a PAM recognition region resides. The other 3 domains that Cas9 contains an α-helical lobe, an arginine-rich region, a topo-homology domain. When crRNA interacts with the complementary region of foreign DNA, the PAM recognition domain interacts with the PAM site found on the dsDNA. The arginine-rich region is suggested to help mediate interaction of Cas9 with DNA, probably by electrostatic interaction. When Cas9 finds the target DNA sequence with the appropriate PAM site, it will melt the double-stranded DNA to induce hybridization with the complementary region of crRNA. In the event of a successful hybridization the HNH nuclease domain cleaves the complementary or target DNA strand and RuvC domain catalyzes the cleavage of the non-target DNA strand resulting in the excision of the foreign DNA 3 nucleotides upstream of the PAM region (Huai et al., 2017). It is important to note that the host genome does not contain PAM downstream of the protospacer sequence thus Cas9 will not be able to target the host genome.

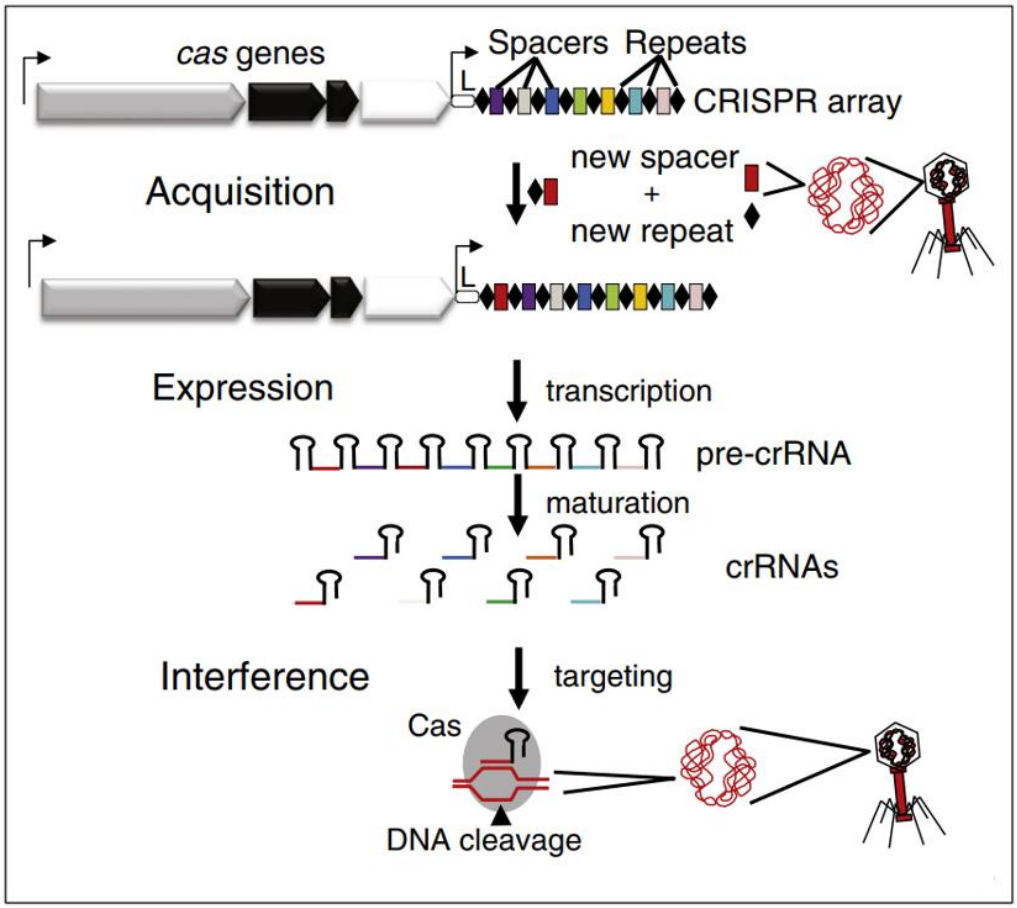

Figure 1.4. Representation of CRISPR mediated adaptive immune system. Acquisition of immunity is performed by excision and integration of foreign spacer sequence upstream of PAM site. The spacer sequence is integrated to the CRISPR array and expressed in the event of an invasion. Transcribed pre-crRNA is maturated and the foreign DNA is interfered by crRNA guided Cas9 endonuclease activity (taken from Barrangou, 2015).

1.4.1. Utilization of CRISPR/Cas9 in Genome Engineering

The research on bacterial immunity has resulted in a breakthrough in genetic engineering. Previously zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), were used to generate a double stranded break (DSB) to induce DNA repair mechanisms and making genome manipulation possible by using the

cell's intrinsic pathways (Gaj, 2014). The CRISPR/Cas9 system was shown to generate a DSB response in a cheaper and more efficient way (Song et al., 2016). Since then, the type II system was utilized to induce alterations to the genome. In theory, any DNA sequence that resides upstream of a PAM can be targeted by the CRISPR/Cas9 system. However, the sequence of a PAM site differs for Cas9 proteins from different species. *Streptococcus pyogenes* Cas9 (spCas9) has a NGG DNA sequence as a PAM motif. This property of spCas9 is advantageous compared to other Cas9 since shorter PAM means increased range of possible target sites in the genome. However, besides NGG sequence, spCas9 has a tendency to interact with NAG and NGA sites as PAM motifs with very low frequency (Cui et al., 2018). For this reason, spCas9 is optimized for gene editing. Cas9 is a RNA-guided nuclease and needs a specific RNA to generate DSB to the intended target site. For this reason, the crRNA and tracrRNA region of the CRISPR locus is fused together and termed as guide RNA (gRNA) or single guide RNA (sgRNA). Yet the design of each gRNA is extremely important to target the intended site and to ensure high mutation rates. Especially if the presence of the mutant is going to be detected by polymerase chain reaction (PCR) screening.

gRNA consists of a scaffold region and a spacer region. The sequence of the scaffold region is universal for Cas9 proteins as it is a RNA sequence that can bind to the Cas9 and activate its DNA binding properties. It is important to note that the spacer sequence guides Cas9 to the DNA. Therefore, the design of this region determines where Cas9 targets in the DNA. The design of this spacer region must contain sufficient homology with the intended region for a strong RNA-DNA interaction (X. Liu et al., 2016). Spacer gRNA consists of two regions. The proximal region of the 20-nucleotide spacer region consists of the first 12 nt (nucleotides) right after the PAM site. It has been observed that mismatches in this region strongly impair the double-stranded break event. Mismatches in the 8 nt distal region were deemed tolerable (Song et al., 2016). In addition, the GC content, position and preference of specific nucleotides and stability of the gRNA are known to affect CRISPR experiments (Chu et al., 2018).

Since it was shown that mismatches in gRNAs may be tolerable, a major concern in using the CRISPR/Cas9 system is the occurrence of off-target effects. There are many controversies on this particular subject. A recommended method in selecting gRNAs is the use of an *in silico* prediction tool that selects the gRNAs from a specific region and compares homology of the selected gRNA to the rest of the genome in order for researchers to select

the gRNA sequence that gives the least off-targets. Moreover *in silico* prediction tools reference the published experimental results as a way to assess the quality of the gRNA by its sequence (Chuai et al., 2016).

Expression of a gRNA that complements a region upstream of PAM site in a genome and Cas9 together in any cell can create DSBs. Breaches such as these must be repaired in order to maintain genomic stability. In order to mend the instability caused by the DSB, organisms have evolved a repair mechanism as a response. Non-homologous end-joining (NHEJ) and homology-directed repair (HDR) pathways are the two most extensively studied DSB responses. The main difference in the outcome between these pathways is the presence of mutations (Yao et al., 2017). The NHEJ pathway seals the bridge by possibly adding or deleting nucleotides (Zhu et al., 2015), while the HDR pathway uses a nearby donor DNA homologous to the DNA near the DSB site and integrates the DNA via homologous recombination (Bassett and Liu, 2014). NHEJ pathway is present in every stage of the cell cycle whereas HDR is restricted to late S phase and G2 phase due to presence of a sister chromatin in very close proximity after DNA replication (Huang et al., 2017). In genetic engineering, the utilization of the HDR pathway can be accomplished by providing a homology template. Therefore, genome editing tools along with CRISPR/Cas9 can provide random frameshift mutations by NHEJ or integrate a DNA sequence to the genome by HDR. Moreover, utilization of CRISPR/Cas9 has opened the way for facilitating advanced genetic manipulations such as recombinase-mediated cassette exchange (RMCE) (Huang et al., 2017).

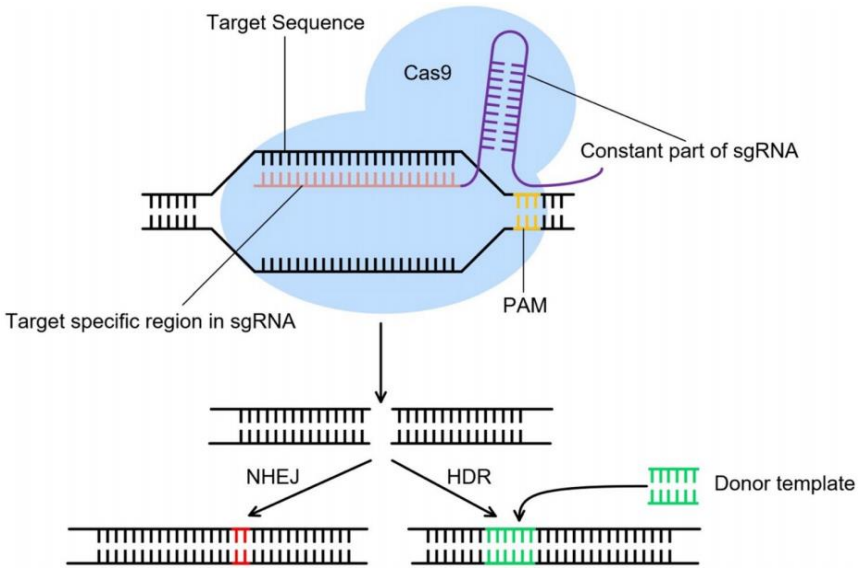

Figure 1.5. Overview of Cas9 endonuclease activity in CRISPR/Cas9 system. 20 nucleotide sgRNA complements the target region upstream of PAM site in genomic DNA. DSB response can result in activation of NHEJ or if a homology donor DNA is present, HDR pathway (adapted from Cui et al., 2018)

1.5. RMCE

Drosophila melanogaster is one of the most advanced organisms in terms of genetic manipulability. In *Drosophila*, cutting-edge genetic manipulation techniques involve the utilization of site specific recombinases. Site specific recombination occurs between two pieces of specialized genetic elements and requires an enzyme, a recombinase, which recognizes these special DNA sequences and generates breakage in DNA (J. Lee et al., 1997). After the breakage, recombinase joins the DNA at by transesterification reaction and consequently creates recombination between precise sites (J. Lee et al., 1997). Site-specific recombination does not use the intrinsic mechanism of homology-directed repair of the organism. This is because site-specific recombinases are of viral origin. The recombinases and their target sites utilized in various organisms for genetic manipulations (Nern et al., 2011). Currently two classes of recombinases were identified, namely serine and tyrosine recombinases (Raymond and Soriano, 2007). These recombinase classes are named after the amino acid residue that initially attacks the phosphodiester bond to generate a breach in DNA. In genetic engineering, most commonly used recombinases are flippase (FLP), cre,

bxb1, and φC31, of which the first two are from the tyrosine recombinase family and the latter are from the serine recombinase family (Smith and Thorpe, 2002).

Recombinases target specific DNA sequences for recombination. Flippase enzyme recognizes flippase recognition target and cre recognizes locus of X-over P1 (loxP). These two tyrosine recombinases are commonly used for mosaic analysis and both recombination reactions are reversible, meaning that the target DNA site is still intact after recombination and is targetable by recombinase (T. Lee, 2014). Even though cre/loxP system was shown to be more efficient in mammalian cells, FRT/FLP system was efficiently established earlier in *Drosophila* with success (Kent G. Golic and Susan Lindquist, 1989). Moreover, cre/loxP system was shown to have toxic effects in proliferating *Drosophila* cells (Heidmann and Lehner, 2001). Even though both systems are sufficient to generate mosaic analysis, they are not as efficient as irreversible φC31 *mediated attP-attB* recombination. The high efficiency of φC31 mediated *attP-attB* recombination drove researchers to use it as a tool for genetic engineering. For years integration of *attP* sites was mediated by transposable elements such as *P*-element, *piggyBac* and *Minos*; this allowed cassette exchange through recombination with an *attB* construct (Venken et al., 2011). A downside to this approach is that integration of attachment sites was not targeted to a specific site in the genome. For this reason, this integration is now mediated by the CRISPR/Cas9 technique.

Advancements in genetic engineering has been facilitated by CRISPR/Cas9. Integration of DNA to a desired genomic position using the CRISPR/Cas9 system requires a target in the genome to generate double stranded break. Thus, successful integration of any given sequence is at least limited by the gRNA efficiency. If the integrated DNA does not contain any visible marker, the screening has to be based on PCR. For this reason, the chosen gRNA should have a high mutagenic rate for efficient screening, which may not be the case in every region of the genome (X. Zhang et al., 2014). When studying the effects of different regions of a given gene, this variability in mutagenesis rate may cause an experimental setback. Constructs with site specific recombinases flanked with a visible marker are used to address this problem. Once these constructs are integrated to the desired location, they can be used as a template to integrate any DNA of interest into the genome by RMCE (Vilain et al., 2014). This two-step method allows efficient engineering of a targeted region in the genome.

2. AIM OF THE STUDY

The glycosylation machinery is composed of many elements and its involvement in neurodevelopment is well established. In order to fully understand how this complicated system works the elements that partake in glycosylation machineries are dissected and characterized by reverse genetic approach for a broader view. Characterization of these elements requires fine tools in order to shed light on their overall function in neurodevelopment. Differences in substrate specificity defines the functional role of hexosaminidases. Therefore, in this study, I characterized the biochemical function of a hexosaminidase by the investigation of its enzymatic activity and substrate specificity by identifying amino acids that are critical for its catalytic function. Furthermore, I generated a dmHexDC mutant by RMCE in which dmHexDC is deleted and replaced by the white gene and is flanked by attP sites. This mutant will enable the facilitated manipulation of *dmHexDC* at its endogenous locus.

3. MATERIALS AND METHODS

3.1. Biological Materials

Drosophila melanogaster lines were kept in air-permeable transparent vials with fly food at constant temperature (25ºC or 18ºC) and humidity (70%) with 12h:12h light and dark cycle. Genesee Scientific Nutri-Fly™ Bloomington Formula was used as nutrition source for flies and cooked according to company's instruction for the *Drosophila* lines.

Table 3.1. Transgenic constructs in *Drosophila melanogaster* lines used in this study.

Transgene	Inserted Chr. No.	Description
Gal4 Drivers		
AC887-Gal4	3	*Gal4* sequence is inserted in the first intron of *CG7985* gene. Gal4 activity depends on the enhancer elements of *CG7985* gene.
UAS Constructs		
UAS-GFPnls	3	Encodes GFP with nuclear localization signal under the control of *UAS*. Expressed GFP labels the nucleus.
UAS-CG7985^{RNAi}	2	Encodes double stranded RNAi of *CG7985* gene under the control of *UAS*.
General Stocks		
dmHexDC::eGFP (BAC)	2	GFP fused with transgenic *dmHexDC* construct inserted in the genome by BAC recombineering.
CG7985^{CPTI100032} (Protein Trap)	3	Protein trap line of dmHexDC. YFP sequence with splice acceptor and splice donor site is inserted in the first intron of *dmHexDC* gene. Null mutant allele of *dmHexDC* (*dmHexDC^{null}*).
Nos-Cas9	1	Expresses Cas9 protein under the control of *nanos* promoter. Cas9 expressed in germ line cells.
RMCE Template	3	*dmHexDC* gene region was excised out and *white* gene flanked by inverted attP sites was inserted to dmHexDC region by CRISPR/Cas9.
Markers & Balancers		

w^{1118} (w)	1	Marker mutation in *white* gene which cause white eye phenotype.
yw	1	Marker mutation which cause yellow body and white eye phenotype.
Sp	2	Sternopleural marker mutation which cause increased bristle number.
CyO	2	Balancer chromosome with curled wing marker.
TM2	3	Balancer chromosome with ubx marker which causes larger halteres and extra bristle(s) on halteres.
TM6B	3	Balancer chromosome with *tb* and *Hum* marker.
tb	2/3	Marker mutation which cause tubby size of larvae and pupae.
sb'	3	Stubble marker mutation which cause short bristles.

3.2. Chemicals and Equipment

3.3. Chemical Supplies

Chemicals used in this study are given in Table 3.2.

Table 3.2 Chemical list used in this study

Chemical	Producing Company
1 kb Marker	NEB, USA (N3232L)
100 bp Marker	NEB, USA (B7025)
Acetic Acid	Sigma-Aldrich, USA (71251)
Acrylamide/Bisacrylamide solution, 30%	BioRad, U.S.A (1610156)
Agarose	SeaKem Cambrex, U.S.A (50004)
Ampicilin	Sigma-Aldrich, USA (59349)
Bovine Serum Albumin	Sigma-Aldrich, USA (A9647)
Citric Acid	Sigma-Aldrich, USA (251275)
DMSO	Sigma-Aldrich, USA (D8418)
DTT	Sigma-Aldrich, USA (D9779)
EDTA	Sigma-Aldrich, USA (59417C)
Ethidium Bromide	Sigma Aldrich, USA (E1510)
Glycerol	Sigma-Aldrich, USA G5516
HCl	Sigma-Aldrich, USA (H1758)
Isopropanol	VWR Chemicals, Germany (20842.330)
Methanol	VWR Chemicals, Germany (20847.320)
$MgCl_2$	Riedel-de Haen, Germany (13152)

N-Acetyl-D-galactosamine	Sigma-Aldrich, USA (A2795)
N-Acetyl-D-glucoseamine	Sigma-Aldrich, USA (A8625)
NaCl	Sigma-Aldrich, USA (S7653)
Paraformaldehyde	Sigma-Aldrich, USA (P6148)
Proteinase K	Roche, Germany (139963000)
Precision Plus Protein™ Dual Color Standards	BioRad (610374)
Sodium Deoxycholate	Sigma-Aldrich, USA (30970)
Trichloroacetic Acid	Sigma-Aldrich, USA (T6399)
Tris	Sigma-Aldrich, USA (T6066)
Triton X-100	AppliChem, USA (A4975)
Tween 20	BioRad (1706531)

3.3.1. Buffers and Solutions

Buffers and solutions in this study are listed in 3.3 with their contents.

Table 3.3 Buffers and solutions used in this study.

Buffer/Solution	Content
Formaldehyde Solution (16%)	8 g paraformaldehyde in 50 ml dH_2O 1M NaOH until solution becomes transparent
LB Agar	5 g/L NaCl 10 g/L Tryptone 5 g/L Yeast extract 14 g/L Agar
LB Broth	5 g/L NaCl 10 g/L Tryptone 5 g/L Yeast extract
McIlvaine Buffer*	X mL 0.1M Citric Acid Y mL 0.2M Na_2HPO_4
Non-ionic Lysis Buffer	20 mM Tris-Cl pH 8.0 50 mM NaCl 1% NP-40 2 mM EDTA 1X Roche Protease Inhibitor Cocktail
PaxD	10 g BSA 3 g Sodium Deoxycholate 3 ml Triton X-100 100 ml 10X PBS dH_2O to 1 L
PBS (1X)	137 mM NaCl

	2.7 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4
PBX3	0.3% Triton X-100 in 1X PBS
Running Buffer (1X)	25 mM Tris 192 mM glycine 0.1% SDS
Sonication Buffer	50 mM Tris HCl 150 mM NaCl 10 mM $MgCl_2$
Squishing Buffer	10 mM Tris, pH 8.0 1 mM EDTA 25 mM NaCl
TAE buffer (1X)	40 mM Tris-Cl 1 mM EDTA 0.1% Acetic acid
TBS (1X)	50 mM Tris-Cl, pH 7.5 150 mM NaCl
TBS-T	0.05% Tween 20 in TBS
Transfer Buffer (1X)	25 mM Tris 192 mM glycine 20 % methanol

* McIlvaine Buffer is prepared using varying volumes of 0.2M Na_2HPO_4 and 0.1M Citric Acid.

3.3.1.1.McIlvaine Buffers

Table 3.4. McIlvaine Buffers with pH 3-8 and their contents.

McIlvaine Buffer pH	**Content**
3.0	17 mL 0.1M Citric Acid 3 mL 0.2M Na_2HPO_4
3.5	15 mL 0.1M Citric Acid 5 mL 0.2M Na_2HPO_4
4.0	12.5 mL 0.1M Citric Acid 7.5 mL 0.2M Na_2HPO_4
4.5	11.25 mL 0.1M Citric Acid 8.13 mL 0.2M Na_2HPO_4
5.0	10 mL 0.1M Citric Acid 10 mL 0.2M Na_2HPO_4
5.5	9.0 mL 0.1M Citric Acid 11.0 mL 0.2M Na_2HPO_4
6.0	8.7 mL 0.1M Citric Acid

	12.3 mL 0.2M Na_2HPO_4
6.5	6.5 mL 0.1M Citric Acid 13.5 mL 0.2M Na_2HPO_4
7.0	6 mL 0.1M Citric Acid 14 mL 0.2M Na_2HPO_4
7.5	5 mL 0.1M Citric Acid 15 mL 0.2M Na_2HPO_4
8.0	3.5 mL 0.1M Citric Acid 16.5 mL 0.2M Na_2HPO_4

3.3.2. Antibodies

Primary and secondary antibodies used in the immunohistochemistry experiments are listed with their dilution ratios in Table 3.4. Storage temperature for primary antibodies was 4°C and the secondary antibodies were kept at -20°C. For the dilution of primary and secondary antibodies, the blocking solution PAXD was used. Dilution of Alexa secondary antibodies was 1:800 except when they are used after the incubation of tissues with anti-GFP antibody (Abcam). In that case, secondary antibodies were diluted to 1:200 with the blocking solution PAXD. High noise-to-signal ratio of the anti-GFP antibody was also eliminated by increased washing time with PBX3.

Table 0.5. Antibodies used in the course of this study.

Name	Antigen	Host Species	Dilution	Source
Primary Antibodies				
Anti-Elav	Elav	Mouse	1:20	DSHB (9F8A9)
Anti-Elav	Elav	Rat	1:20	DSHB (7E8A10)
Anti-GFP	GFP	Rabbit	1:500	Torrey Pines (TP401)
Anti-Pros	Prospero	Mouse	1:20	DSHB (MR1A)
Anti-Repo	Repo	Mouse	1:20	DSHB (8D12)
Secondary Antibodies				
Alexa 488	Rabbit	Goat	1:800	Invitrogen
Alexa 488	Mouse	Goat	1:800	Invitrogen
Alexa 555	Rat	Goat	1:800	Invitrogen
Alexa 555	Mouse	Goat	1:800	Invitrogen
Alexa 647	Mouse	Goat	1:800	Invitrogen
Alexa 647	Rat	Goat	1:800	Invitrogen

3.3.3. Embedding Media

Tissues were mounted in Vectashield Embedding Medium (Vector Laboratories, Inc). Embedded tissues were kept in dark at 4°C until visualization by confocal microscopy.

3.3.4. Disposable Labware

Disposable labware used during the course of this study are given in 3.6

Table 0.6. Disposable labware used in this study.

Material	**Manufacturer**
Culture tubes (14 ml)	Greiner Bio-One, Belgium
Filter Tips	Greiner Bio-One, Belgium
Microscope cover glass	Fisher Scientific, UK
Microscope slides	Fisher Scientific, UK
PCR tubes (200 µl)	Bio-Rad, USA
Petri Dishes, 60 x 15 mm	TPP Techno Plastic Products AG, Switzerland
Pipette Tips	VWR, USA
Plastic Pasteur pipettes	TPP Techno Plastic Products AG, Switzerland
PVDF membrane	Roche Life Science
Syringe (1cc)	Becton, Dickinson and Company, USA
Syringe (2cc)	Becton, Dickinson and Company, USA
Syringe (5cc)	Becton, Dickinson and Company, USA
Test Tubes, 0.5 ml	Citotest Labware Manufacturing, China
Test Tubes, 1.5 ml	Citotest Labware Manufacturing, China
Test Tubes, 2 ml	Citotest Labware Manufacturing, China
Test Tubes, 15 ml	Becton, Dickinson and Company, USA
Test Tubes, 50 ml	Becton, Dickinson and Company, USA

3.3.5. Equipment

Equipment used during the course of this study is listed in 3.7.

Table 0.7. Equipment used in this study.

Equipment	Manufacturer
Autoclave	Astell Scientific Ltd., UK
Centrifuges	Eppendorf, Germany (Centrifuge 5424, 5417R)
Confocal Microscope	Leica Microsystems, USA (TCS SP5)
Electrophoresis Equipment	Bio-Rad Labs, USA
Fluorescence Stereomicroscope	Leica Microsystems, USA (MZ16FA)
Freezers	Arçelik, Turkey
Gel Documentation System	Bio-Rad Labs, USA (Gel Doc XR)
Heating Block	Fisher Scientific, France
Heating Magnetic Stirrer	IKA, China (RCT Basic)
Incubator	Weiss Gallenkamp, USA (Incubator Plus Series)
Laboratory Bottles	Isolab, Germany
Micropipettes	Eppendorf, Germany
Microwave oven	Vestel, Turkey
Mini-PROTEAN Tetra Cell	Bio-Rad Labs, USA
pH meter	WTW, Germany (Ph330i)
Refrigerators	Arçelik, Turkey
Stereo Microscope	Olympus, USA (SZ61)
Thermal Cycler	Bio-Rad Labs, USA (C1000 Thermal Cycler)

3.3.6. Oligonucleotide Primers

Primers were diluted with dH_2O to obtain a final concentration of 100 pmol/µl. Diluted primers were stored at -20 °C. E347A fw, E347A reverse, D346A fw, D346A rv, Double Mutant fw, Double Mutant rv, I343V_fw, I343V_rv, Pam_mutation_Catalytic_fwd, Pam_mutation_Catalytic_rv RMCE_PAM_SDM_fw, RMCE_PAM_SDM_rv primers are site directed mutagenesis primers. The procedure is described at 3.3.1.2. RMCE_gibson_PAM_rv, RMCE_Gibson_PAM_fw. RMCE-PCR1fwd, RMCE-PCR2rev, HA_Up_Left_RMCE_fw, HA_Up_Right_RMCE_rv, HA_Down_Right_RMCE_fw, HA_Down_Left_RMCE_rev primers contain homology arms with other DNA and are used for Gibson assembly. The Gibson assembly procedure is described at 3.3.7.

Table 3.8. Oligonucleotide primers used during the course of this study.

Primer Name	Primer Sequence (5' to 3')	T_m °C
HexDC-NcoI Forward	TATCCATGGCAGGTTCCACTCCATTTC	68.3
HexDC-KpnI Reverse (1)	ATTGGTACCGCATGAGCTCTCCCCTCA	71.3
HexDC-KpnI Reverse (2)	ATTGGTACCTCACAAAGCGGGTCTCC	69.5
dmHexDC-CDS-NcoI Fw	TTTCCATGGACTCAAATCTTTGGATATTTATTT	65.3
dmHexDC-CDS-KpnI rv	ATTGGTACCTCATTGTTGCTCCTGCTGATG	70.7
E347A fw	GTTGTGACGCGGTGCAGCGCATGGGCGA	77.5
E347A reverse	GCGCTGCACCGCGTCACAACCGATGTGAATATG	77.7
D346A fw	TCGGTTGTGCCGAGGTGCAGCGCATGGG	77.5
D346A rv	GCGCTGCACCTCGGCACAACCGATGTGAATATGGGT	80.1
Double Mutant fw	TCGGTTGTGCCGCGGTGCAGCGCATGGG	78.9
Double Mutant rv	GCGCTGCACCGCGGCACAACCGATGTGAATATGGGT	81.2
I343V_fw	CATATTCACGTCGGTTGTGACGAGGTGC	71.6
I343V_rv	TCACAACCGACGTGAATATGGGTGAACCTGATT	72.7
Hex347A_seqfwd	GCGGAGCGACAGACTCAGTACGAG	70.4
HDR2_Cat_ Fw	TCGATACGACAGCCTCAGCAC	63.3
HDR2_Cat_ rv	CATGTCCTCCTGCACTTTCAAGG	64.7
Top_Oligo_CatMut_gRNA	TGCACGTCACAACCGATGTGAATA	63.5
Bottom_Oligo_CatMut_gRNA	AAACTATTCACATCGGTTGTGACG	61.8
Pam_mutation_Catalytic_fwd	CAGGTTCACGCATATTCACATCGGTTGTGACG	73.8
Pam_mutation_Catalytic_rv	GATGTGAATATGCGTGAACCTGATTCGCGCCAC	75.2
Doube_grna_catmut_fw	GCGGCCCGGGTTCGATTCCCGGCCGATGCACGTCACAACCGATGTGAATAGTTTTAGAGCTAGAAATAGCAAG	89.6
Doube_grna_catmut_rv	ATTTTAACTTGCTATTTCTAGCTCTAAAACTATTCACATCGGTTGTGACGTGCACCAGCCGGGAATCGAACCC	85.7

Catalytic_site_3prime	CCCATATTCACATCGGTTGTGACGA	65.8
CatmutSurveyor_rv	CGTCATCAGACTGGGAATGCCTA	64.7
CatmutSurveyor_Fwd	AGCGTCTCGTTCATCTGGACTTG	64.7
Catalytic_mutant_3prime	CCCATATTCACATCGGTTGTGACGC	67.4
WT_PAM_HDR	TGGCGCGAATCAGGTTCACC	62.5
Mutant_PAM_HDR	TGGCGCGAATCAGGTTCACG	62.5
Catmut_Surveyor2_rv	CGACAGCGACGTCATCAGACTGG	68.2
Surveyor_ctrl_fw	GAGTGTCGTGTGCTGCTGATTGC	66.4
Surveyor_ctrl_rv	CGGTCCTCCTTGGTGAATTGCCA	66.4
Catmut_sequencing_fw	GGGCACATGGAATACGTCCTC	63.3
Catmut_sequencing_rev	GGTATATGTCACTGGCATACACC	62.9
RMCE-PCR1fwd	GCGGCCCGGGTTCGATTCCCGGCCGATGCAGTG CGTTTCTCAGCTCGCAGGTTTTAGAGCTAGAAA TAGCAAG	91.3
RMCE-PCR2rev	ATTTTAACTTGCTATTTCTAGCTCTAAAACACGC GCTTTACTAAACGTTTTGCACCAGCCGGGAATC GAACCC	85.1
HA_Up_Left_RMCE_fw	GGGATAACAGGGTAATGGTACTTTGTTGTTTAT TGTTGGCTTTGGC	78.0
HA_Up_Right_RMCE_rv	GCGGCCGCATATCCTAGGTAGTTAACCTTTACC ACACCTAAATAATAA	85.2
HA_Down_Right_RMCE_fw	GGCACTACGCGGCCGCTATAAGCGCGTAGGTGT CTAGAATCT	82.8
HA_Down_Left_RMCE_rev	CCTGTTATCCCTACTCGAGTTATTCGTGCATTGC GTGCCTTAAAG	79.8
pWS_HA_up_forward_screen	TTGCGAGTACGCAAAGCTAGAG	62.1
pWS_HA_up_reverse_screen	ACTAGAATTCCTACGCCCCC	60.5
RMCE_downHA_screen_Fwd	GTGACCTGTTCGGAGTGATTAG	62.1
RMCE_downHA_screen_Rev	CGAAATGCGTCGTTTAGAGCAGC	64.7
gRNA(U6:1)RMCE_fwd	TATATAGGAAAGATATCCGGGTGAACTTCGAAC GTTTAGTAAAGCGCGTGTTTAGAGCTAGAAATA GCAAG	83.5
gRNA(U6:3)RMCE_rev	ATTTTAACTTGCTATTTCTAGCTCTAAAACCTGC GAGCTGAGAAACGCACGACGTTAAATTGAAAA TAGGTC	82.9
RMCE_PAM_SDM_fw	TAAATAATCGGCTGCGAGCTGAGAAACG	68.7
RMCE_PAM_SDM_rv	CAGCTCGCAGCCGATTATTTAAATGCTCGGCAA AC	75.4
RMCE_Pam_mut_screen2	CACAGAGAGCTATTGCGTTG	58.4
RMCE_Gibson_PAM_fw	CTATTAGTTTGGTTTAATCGTTTAAACACGGCAA AATCGTATAA	72.4
RMCE_gibson_PAM_rv	CTACGCGGCCGCATATCCTAGGTAGTTAACCTT TACCACACCTAAAT	80.7
RMCE_Screen1_fwd	CGGACAGGTGCATCAGTTACTAG	64.7

RMCE_Screen1_rv	GCTGTCACCCTGCAATGATAGGG	66.4
RMCE_Screen2_fwd	TCTTCTTGAACTCGGGCTCGG	63.3
RMCE_Screen2_rv	GTTTGCTCTCCACGGCGTTTG	63.3

3.4. Molecular Biological Techniques

3.4.1. Polymerase Chain Reaction

3.4.1.1. Conventional PCR

Q5 high fidelity polymerase was the DNA polymerase that was most used in this project thus the preparation of master mix and adjustment of Thermocycler conditions were applied as the manufacturer's suggestions with consideration of the primer's melting temperatures. A conventional PCR with Q5 High Fidelity polymerase includes 1X Q5 Reaction Buffer, 10 mM dNTPs, 10 µM forward and reverse primers, 0.02U/µl Q5 High Fidelity polymerase, variable concentrations of template DNA and ddH_2O up to a final volume, which can vary depending on the downstream applications. Initial denaturation is 30 seconds at 98°C. After initial denaturation a total of 25-35 cycle of 98°C for 10 seconds, annealing temperature depending on the primers Tm (melting temperature) for 30 seconds and extension by Q5 polymerase is done at 72°C for 2 minutes, respectively. After the cycles, a final extension time is given for 5 minutes at 72°C.

3.4.1.2. Site Directed Mutagenesis

Site directed mutagenesis is used to mutate and amplify a selected region of the plasmid template. Forward and reverse primers that were partially overlapping were designed to have mismatches in the desired region of template plasmid. Polymerase Chain Reaction was made with Q5 high fidelity polymerase. Since the whole plasmid is amplified, the extension setting at the thermocycler was set to 5 minutes per cycle.

3.4.1.3. Two-Step PCR

Two step PCR is made for addition of nucleotides that are not present in the template. The selected primers are extended with a matching homology sequence and overhang sequence for the addition of desired sequence. For the initial 5-6 cycles, annealing temperature of the PCR is set for the annealing temperature of matching homology region of primer only. After the initial cycles, the annealing temperature is increased according to the overhang sequence and the homology sequence to match the template primer.

3.4.2. *Dpn*I Digestion

The methylated template DNA was eliminated with *Dpn*I restriction enzyme. The enzyme was diluted in 10X Cutsmart buffer and distributed to PCR tubes after Site Directed Mutagenesis PCR and was incubated at 37°C for 3 hours.

3.4.3. Agarose Gel Preparation

Agarose was added to 1X TAE buffer to a final concentration of 0.8 or 1%. Microwave was used to boil the agarose-1X TAE mixture in order to dissolve the agarose. The mixture was allowed to cool down and then 30ng/mL ethidium bromide was added.

3.4.4. Agarose Gel Electrophoresis

DNA samples were prepared with loading dye to a final concentration of 1X. The samples were then loaded into wells. Moreover, a 1kb or 100bp DNA ladder was added to one of the wells for the comparison of DNA product sizes. The electrophoresis tank is filled with 1XTAE buffer and agarose gel that was previously prepared was placed into the tank. Gel was run at 100V for 45 minutes and the Gel Doc System was used to visualize the results.

3.4.5. DNA Purification from Agarose Gel

After agarose gel electrophoresis, Agarose gel that contained DNA of interest was sliced out on a UV box using a sterilized scalpel and was isolated according to the instructions of High Pure PCR purification Kit (Roche). The piece of agarose was placed into a 1.5 mL Eppendorf tube and 300 μL Binding buffer was added to every 100mg of agarose. DNA was released from the agarose by incubating with Binding buffer at 56°C and vortexed every 3 minutes for 10 minutes. After incubation, isopropanol was added half the volume of Binding Buffer and the solution was pipetted into a DNA purification column. The column was centrifuged at 13.2 rpm for 1 minute and the flow-through was discarded. DNA bound to the column was washed with 500 μL and 200 μL Washing buffer sequentially and one more centrifuge was made to get rid of ethanol residues in the column. 30-50 μL Elution Buffer was added to the column, the column was placed into a new 1.5 mL Eppendorf tube and centrifuged once again to collect the DNA of interest.

3.4.6. Restriction Digestion

Restriction enzymes and their appropriate buffers were purchased from NEB and were used according to the manufacturer's protocols. 1 μg of DNA was digested with 1 μL of enzyme in 10X appropriate buffer at a total reaction volume of 50 μL for 1h at 37°C.

3.4.7. Gibson Assembly

0.025pmoles of template plasmid and 0.075pmol insert with homology arms were mixed and dH_2O was added until a total volume of 10 µL. 10 µL of 2X Gibson Assembly Reaction Mix was added to the DNA mix and the reaction was incubated at 50°C for 1 hour. After Gibson assembly, the reaction was transformed into competent *E.coli* cells.

3.4.8. Ligation

T4 DNA ligase (NEB) was used for ligation of DNA. The molar ratios of insert and vector backbone was 3:1. Sticky ends were ligated at 16°C O.N. After ligation, T4 DNA Ligase was inactivated at 65°C for 10 minutes. 5 µL of the reaction was chilled on ice and transformed into competent *E.coli* cells

3.4.9. Total RNA Isolation from HEK293T cells

$8x10^6$ number of cells were scraped from the cell culture plates using 1mL of TRIzol reagent and collected into an Eppendorf tube where they are incubated for 5 minutes in room temperature. 0.2 mL of chloroform was added to the incubated mixture and was incubated for 3 minutes. The sample was then centrifuged at 12.000 rcf at 4°C. The centrifugation allows for phase separation and upper aqueous phase was pipetted to a new Eppendorf tube. 0.5mL of isopropanol was added to the mixture to precipitate it. After addition of isopropanol the sample was incubated for 10 minutes at room temperature and centrifuged at 12.000 rcf at 4°C. RNA is now precipitated in a gel-like form at the bottom of the tube. The supernatants were discarded. 1 mL of 75% ethanol was used to wash the precipitated RNA. Ethanol+RNA mixture was vortexed and centrifuged for 5 minutes at 7500 rcf at 4°C. The pellet was air dried for 10 minutes but it was ensured that pellet was not too dry. 50µL of RNase-free water that contains 0.1M EDTA was used to resuspend the RNA pellet. RNA was incubated for 15 minutes at 60°C for its solubilization. The solubilized RNA was stored at -80°C.

3.4.10. cDNA Synthesis

Total RNA was mixed with10mMM dNTP, 5µg oligo$(DT)_{12-18}$ primers and DEPC treated water with a final volume of 10 µl. This mixture was incubated at 65°C for 5 minutes for the purpose of annealing the primers to poly(A) tail containing RNA. During the incubation a reaction mix was prepared. The reaction mix was composed of 5mM $MgCl_2$ 10mM DTT and 40 Units of RNaseOUT and 10X Reaction Buffer. 9 µL of the reaction mix was added to the RNA primer mixture that was incubated and the new mix was incubated at

42°C for 2 minutes. 40 units of SuperScript™ II RT (reverse transcriptase enzyme) was added. The reaction was incubated for 50 minutes at 42°C in a thermocycler. The reverse transcription reaction was terminated by increasing the temperature to 70°C for 15 minutes and was chilled on ice. cDNA was synthesized after the reaction and the RNAs were digested by RNAase H for 20 minutes at 37°C. The cDNA library was stored at -20°C.

3.4.11. Chemical Transformation

DH5α, TOP10F, JM109 or Rosetta pLysS competent *E. coli* cells were transformed with plasmid of interest by heat-shock method. 1ng DNA was mixed with 50 µl of competent cells on ice. The cells were rested on ice for 30 minutes and heat shocked at 42°C for 1 minute. The cells + DNA mixture was then incubated 5 minutes on ice. After that, 500 µl of LB was added to the tube and the cells were incubated at 37°C for one hour. 100 µl of the culture was plated on LB + antibiotic plates. The plates were incubated at 37°C O.N. The colonies were selected for downstream applications on the next day.

3.4.12. Plasmid DNA Isolation

3.4.12.1. MiniPrep

Small scale of plasmids were isolated by using GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Bacteria were selected from agar plates and inoculated in 3mL LB broth + Antibiotics. After 14-16 hours, grown bacteria was centrifuged for 3 minutes at 8000 rpm in a tabletop centrifuge. The supernatant was discarded and bacterial pellet was resuspended in 250 µL Resuspension buffer in a 1.5 mL Eppendorf tube. After resuspending the bacteria, 250 µL of Lysis solution was added and was incubated on RT for 4 minutes. The mix was inverted 5-6 times until the solution became clear and the basic solution was neutralized by 350 µL Neutralization buffer and the solution was mixed thoroughly. The cloudy solution was centrifuged at 13.000 rpm for 10 minutes at a standard tabletop centrifuge and the supernatant was pipetted into GeneJET Spin Columns. The columns that contain bacterial extract were centrifuged for 1 minute at 13.000 rpm and the flow-through was discarded. The column was washed with 500 µl wash solution and centrifuged at 13.000 rpm for 1 minute. This process has been repeated twice and the flow-through was discarded in both. Column was centrifuged without any addition of solution to get rid of unwanted residues. At the last step, column was transferred into a clean 1.5 mL Eppendorf tube. 50 µl of elution buffer was added to the column and the column was incubated with the elution buffer for 5 minutes to increase the DNA yield. Finally, the column was centrifuged at

13.000 rpm for 1 minute. The flow-through was collected and DNA concentration was measured at NanoDrop.

3.4.12.2. MidiPrep

ZymoPure™ Plasmid Midiprep Kit from Zymo Research is used in order to obtain endotoxin free medium scale plasmid isolation according to the manufacturer's protocol. bacteria containing the plasmid of interest were grown in 50 mL LB medium O.N at 37°C. The bacteria were the centrifuged ≥ 3,400 x g for 10 minutes. The pellet is the resuspended with 8 mL red colored P1 buffer. Then 8 mL P2 buffer is added. P2 buffer is a green solution and it is the lysis solution. Thus, the tube is inverted 5-6 times and not subjected to any harsh actions such as vortexing. The mixture is then incubated at room temperature for 3 minutes. At this stage, solution should be purple colored, clear and viscous. Yellow colored P3 buffer is added. P3 is the neutralization buffer and will allow the adjustment of pH to neutral. After neutralization, the sample is yellow-colored and a precipitate is formed. The plug is attached to the Luer Lock below of the ZymoPURE™ Syringe Filter. The sample is then poured into the syringe filter and sat at room temperature for 8 minutes. After the 8 minutes the yellowish precipitate was floating at the top of the mixture. The Luer Lock is unlocked and the sample was pushed through the syringe filter to a clean 50 mL test tube and thus, the precipitate was separated from the sample. 8 ml ZymoPURE™ Binding Buffer was added to the cleared lysate and mixed well by inverting the tube up and down. The mixture is poured into Zymo-Spin™ III-P Column Assembly which is previously integrated into a vacuum tank. The vacuum is then turned on and all liquid passed from the column. At this point DNA binds to the column and is washed with Wash 1 solution and two times with Wash 2 solution. The column is then transferred into a clean 1.5 mL Eppendorf tube and 200 µL DEPC-treated ddH_2O was added into the column. The column was centrifuged at 12.000 x g for 1 minute and the DNA concentration of the eluted plasmids were measured using a NanoDrop.

3.5. Biochemical Methods

3.5.1. Large Scale Protein Expression

A single transformed JM109 *E.coli* colony was selected and grown in LB broth medium with 150 µg/ml ampicillin. The volume of this medium is usually 100 fold smaller than the protein expression medium and is labeled as the "starter culture". The starter culture is incubated at 37°C O.N. at 200-250 rpm. The starter culture is then poured into a LB + ampicillin that has a 100 fold higher volume. This culture is incubated at 25°C. Optical

density of the culture is measured as it is an indicator for bacterial growth. IPTG induction was made with 1mM IPTG when OD_{600}=0.6 as it is the exponential phase of bacterial growth thus the most suitable time for standard protein expression. The culture was collected and centrifuged at 4°C for 20 minutes at 4500 rpm with JA-14 rotor 4 hours after induction. Only the dry bacterial pellet was stored at -80°C.

3.5.2. Extraction of Bacterial Lysate

3 grams of bacterial pellet is resuspended in 9 mL sonication buffer. The resuspended sample is sonicated at 40% power, 5 cycles for 30 seconds with 20 second intervals. The sonicate is centrifuged at 4°C, 10.000 rpm for 20 minutes with JA-14 rotor. The supernatant is aliquoted in small volumes to in order to prevent freeze thawing and the samples were stored in -80°C.

3.5.3. Immobilized Metal Affinity Chromatography

5 mL HisTrap (GE Healthcare Life Sciences) that contains Nickel beads in ethanol were washed with cold ddH_2O and equilibrated with 50 mL binding buffer that contains imidazole. Introduction of imidazole to Nickel beads in this step is required to prevent unspecific binding. Sonicated lysates were centrifuged at 10.000 rpm for 20 minutes with JA-14 rotor and were introduced to the column by a controlled flow rate of 0.5 mL/min. The flow through was loaded twice to ensure maximum efficiency. The column was washed with binding buffer to get rid of unspecific bindings. Elutions with a range of imidazole concentration from 20mM to 250mM was applied in order to obtain the purified protein while connected to the Akta Prime chromatography system that measures absorbance at 280nm to distinguish the imidazole concentration which proteins are leaving Nickel beads. Purification steps were performed at 4°C.

3.5.4. Enzyme Assay

2.5 µl Crude JM109 cell extracts were incubated on wells of a 96 well plate with 21.25 µl dH_2O, 25 µl of McIlvaine citrate-phosphate buffer and glycosides as substrate. Substrate concentrations of *p*-nitrophenyl-GlcNAc or *p*-nitrophenyl-GalNAc were 1.25 mM. Extracts were incubated for one hour at 37 °C. The reaction was stopped with 250 µl 0.4 M glycine-

NaOH which has a pH of 10.4. Absorbance at 405 nm was measured after the addition of glycine-NaOH.

3.5.5. Statistical Analysis of Enzyme Assays

Every Enzyme test contains a blank sample (- lysate), and a negative control (- dmHexDC) and experimental groups (+ dmHexDC, E347A) for every pH value. Statistical analyses are made on the raw absorbance data by extracting the blank value from the negative control and experimental group, which is done separately for every pH value. Background corrected groups are then compared by Students T-test. The data is then normalized as 0=0% and the highest background corrected absorbance value is set at 100% for the sake of convenience. All data are represented by mean averages and +/- standard deviations. The graphs are made by GraphPad Prism 7.00 Software.

3.5.6. Protein Extraction from Third Instar Wandering Larvae

10-14 larvae were collected, washed with PBS and frozen for 5 minutes at -20°C in an Eppendorf tube. 100-140 µl lysis buffer with 1X Protease inhibitor cocktail were added to the Eppendorf tube and the larvae were then homogenized with a pestle. The homogenized extracts were then incubated on ice for 30 minutes. Incubated extracts were centrifuged at 4°C for 10 min at 13200 rpm. The clear part containing the proteins was collected while the lipid-rich upper layer was avoided and transferred to a new Eppendorf tube. Laemmli's sample Buffer was added to the clear extract to a final concentration of 1X. The sample was then boiled for 10 minutes at 98°C to disrupt the secondary and tertiary structures of proteins and DTT or β-Mercaptoethanol found in Laemmli's sample buffer allowed breakage of disulphide bonds. The boiled sample was then stored at -20°C prior to SDS-PAGE/Western blotting.

3.5.7. Polyacrylamide Gel Preparation

Glass plates with appropriate separation (usually 1mm) was held together by a frame and placed on the gasket on a stand. The resolving gel was then poured between the glass plates and isopropanol was added on top of the resolving gel solution. Isopropanol cuts the interaction of air and resolving gel solution and smoothens the upper edge of the resolving gel solution before polymerization. After the resolving gel polymerizes the isopropanol is discarded. The stacking gel was poured on top of the polymerized resolving gel and an appropriate comb was placed between the glass plates before the stacking gel polymerized.

The acrylamide gel was either stored with wet tissue paper at 4°C or used for SDS-PAGE/Western blot after polymerization.

3.5.8. SDS-PAGE

The polyacrylamide gel is clamped and placed at the buffer tank for electrophoresis. 10X Running buffer was diluted to 1X with distilled water and poured into the tank. Samples were loaded and prestained protein ladder was also added to determine the size of the proteins after the experiment. The gel was run at 80V until the samples passed the stacking gel and then the voltage was increased to 120V. The protein gel was either stained with Coomassie blue or used in Western blotting.

3.5.9. Coomassie Blue Staining

The gel was collected with the help of a gel releaser and washed with water. After washing, the gel was incubated for 30 minutes with Coomassie blue dye, which stains every protein on the protein gel. Coomassie blue dye was saved for later experiments and the gel was washed with water and incubated with destain solution for two hours. Destain solution was then discarded and the gel was analyzed by naked eye.

3.5.10. Western Blotting

10X Transfer buffer were diluted to 1 X using distilled water and 20% methanol. After SDS-PAGE, the acrylamide gel was put in water to wash away SDS from Running Buffer. Polyvinylidene difluoride (PVDF) membrane was activated by methanol and transferred to water to wash away the methanol. The membrane was then taken from the water and incubated in 1X transfer buffer. The gel sandwich for the transfer of proteins from acrylamide to the membrane was made in following order: fiber pad Whatman paper-gel-methanol activated PVDF membrane-Whatman paper-fiber pad, in order from the black side to the white site of cassette. Proteins were transferred to the membrane with 200-220mA current at 4°C for 1.5 hours. The transfer was validated by Ponceau staining. Ponceau's Red binds to proteins in the membrane, giving bands at visual spectrum. Ponceau's Red dye is then washed with TBS-T three times for 5 minutes and the membrane was blocked with 5% non-fat milk powder in TBS-T for 2 hours. The blocking solution was washed with TBS-T three times for 10 minutes and primary antibody with appropriate dilution was prepared in 5% BSA + TBS-T. The membrane was incubated with primary antibody containing solution over-night at 4°C. Primary antibody was then collected to be used in other experiments and the membrane was washed with TBS-T three times for 10 minutes. The secondary antibody

conjugated with HRP was prepared in 5% non-fat milk powder+TBS-T solution and incubated with membrane at room temperature for 2 hours. The secondary antibody was then washed with TBS-T three times for 10 minutes. Western Bright ECL was used as a substrate for HRP and the bands were visualized by Syngene visualization systems. The settings were used specifically for Western blot visualization.

3.6. Histological Methods

3.6.1. Grape Agar Preparation

Grape Agar was prepared by adding 500 mL of dH_2O to Grape Agar Powder Premix and boiling it in a microwave. The mixture was then poured to 10mm or 60mm culture plates.

3.6.2. Embryo Collection

Mutant and wild-type *Drosophila* with a 1:1 male to female ratio were placed in population cages on grape agar plates with yeast paste. After desired time, the adult flies were removed with the cages and the yeast paste was scraped as much as possible. Embryos are covered with chorion which disrupts the imaging process. Therefore, embryos on the plate were dechorionated by 100% bleach for 1 minute. Dechorionated embryos were washed with distilled water. Embryos were then fixed with n-Heptane and 4% PFA at a 1:1 ratio for 30 minutes taped onto a shaker at 300 rpm. After the fixation step the vitelline membrane was removed as follows. The lower layer (Formaldehyde) was pipetted out and 100% Methanol was added at a 1:1 ratio. The embryos were then vortexed for 1 minute for the removal of the vitelline membrane. The fixed embryos were washed and stored in absolute Ethanol at -20°C until further experiments.

3.6.3. Immunohistochemistry of Drosophila Brain and Eye imaginal Discs

Drosophila tissues that were dissected in cold PBS were fixed with 4% PFA for 15 minutes on a shaker at 70 rpm. Until the mounting step brains were stored at 0.5ml test tubes and eye imaginal discs will be stored in three-well glass dishes. Tissues were washed three times with PBX3 for 15 minutes and blocked with PAXD for two hours at room temperature. Primary antibody solution was prepared by diluting primary antibodies in PAXD solution on ice and tissues were incubated in primary Ab solution O.N. at 4°C. After the incubation time primary Ab solution was removed and stored at 4°C for later experiments and tissues were washed three times with PBX3 for 15 minutes. Secondary antibody solution was also prepared on ice by appropriately diluting fluorescence dye-conjugated secondary antibodies

in PAXD. During this step, the tissues were incubated with secondary antibody solution for 2 hours at RT in the dark. Tissues were then washed three times with PBX3. The washed tissues were then dissected once again to get rid of unwanted tissue and were mounted on a slide with Vectashield and the edges of the cover was sealed with nail polish. The images of the mounted tissues were then taken with a Leica TCS SP5 confocal microscope. The images were further processed with Image J or Leica LAS EX programs.

3.6.4. Immunohistochemistry of *Drosophila* Embryos

Embryos stored at -20°C were washed with ethanol one time and three times with PBX3 for 15 minutes on RT. The embryos were then blocked with PAXD for 2h at RT on a nutator. The embryos were then incubated with primary antibody solution O.N. at 4°C. Embryos were then washed with PAXD four times for 20 minutes and then the washed embryos were incubated in secondary antibody solution for 2 hours at room temperature in the dark. After incubation with the secondary antibodies, the embryos were washed three times with PBX3 for 15 minutes once again and were mounted with Vectashield and the edges of the cover was sealed with nail polish. The images of the mounted tissues were then taken with a Leica TCS SP5 confocal microscope. The images were further processed with Image J or Leica LAS EX programs.

3.7. Computational Biology Methods

3.7.1. Homology modeling of dmHexDC using MODELLER

The MODELLER tool was used as the program to create a homology model and UCSF Chimera was used as an interface. PSI-BLAST algorithm was selected as it is a sensitive iterative search for evolutionary conserved domains and BLOSUM 80 matrix were chosen and a search was made to identify the amino acid sequences similar to dmHexDC in the pdb database, which contains the information of coordinates that are experimentally obtained from the spatial distribution of atoms in proteins. The structures with the highest scores for different regions were fetched and 5 output models were created.

3.7.2. Homology modeling of dmHexDC using I-TASSER

dmHexDC amino acid sequence was submitted to the I-TASSER server without any restraining parameters and one week later the output .pdb file was obtained.

4. RESULTS

4.1. Bioinformatics of dmHexDC

4.1.1. dmHexDC has a putative N-glycosylation domain

In order to characterize dmHexDC function in *Drosophila* we aimed to investigate its domain structure and made use of bioinformatic analysis tools. Insect hexosaminidases are known to be glycosylated. This glycosylation was shown to have no effect on their catalytic activity (Dragosits et al., 2015) and did not provide information on their proper localization. Thus, to investigate whether dmHexDC has a glycosylation motif I used the NetNGlyo 1.0 server that predicts putative glycosylation sites by searching for glycosylation motifs (Asn-Xaa-Ser/Thr) in a given sequence. The analysis showed that dmHexDC has a potential N-glycosylation site near the C-terminus at amino acid 611 that codes for Asparagine. The analysis results are shown in Figure 4.1.

It is known that β-hexosaminidases are located in lysosomes and trafficked into lysosomes via the M6P pathway (Dersh et al., 2016). To investigate whether dmHexDC is also localized to the lysosome when looking at its sequence we used the Protein Subcellular localization prediction server WoLF PSORT and as expected lysosome was the highest scoring organelle with a score of 10 followed by a score of 9 for extracellular and 5, 4, 3, and 1 for ER, Golgi, peroxisome, and mitochondria, respectively (Table 4.1).

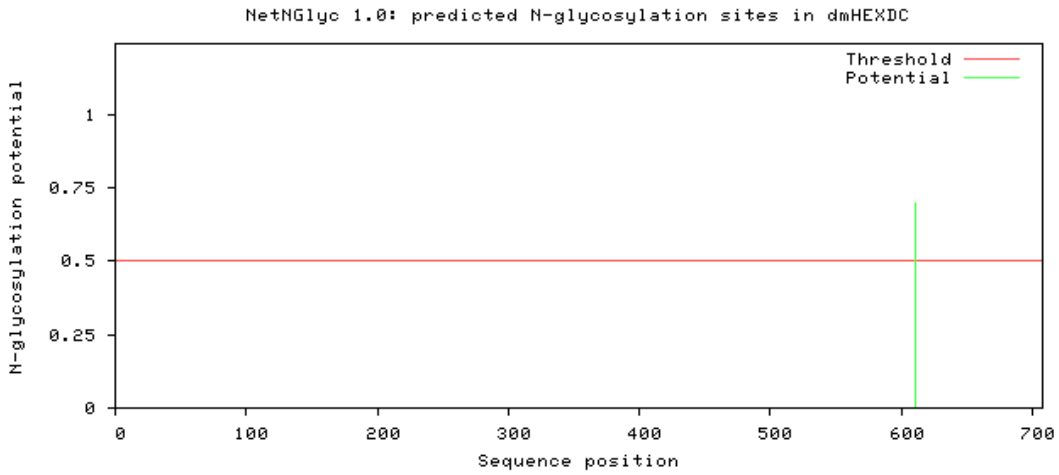

Figure 4.1. Analysis of putative glycosylation sites on dmHexDC using NetNGlyo software predicts a putative glycosylation site at amino acid position 611.

Table 4.1. WoLF PSORT Protein Subcellular localization software prediction suggests that dmHexDC is most likely localized to the lysosome.

Organelle	Score
Lysosome	10
Extracellular	9
ER	5
Golgi	4
Peroxisome	3
Mitochondria	1

4.2. Elucidation of the biochemical function of dmHexDC

Previously, three *Drosophila* β-hexosaminidases were characterized *in vitro*. These hexosaminidases, Hexo1, Hexo2, and Fused Lobes were shown to hydrolyze terminal N-acetyl-β-D-glucosamine residues and Hexo1 and Hexo2 function in N-glycan trimming and chitin metabolism (Dragosits et al., 2015). Our goal here is to characterize the substrate specificity of dmHexDC.

4.2.1. Expression of dmHexDC in *E. coli*

Knowledge about the substrate specificity of dmHexDC will shed light on its function and will reveal information about its physiological substrate. Thus, to characterize its biochemical properties I expressed it in the JM109 *E. coli* strain after assessing if dmHexDC can be expressed in *E. coli*.

The dmHexDC sequence was analyzed against rare codons in *E. coli* and this analysis showed that it can be produced in *E. coli* and does not need codon optimization (see Appendix). However, a eukaryotic protein can undergo critical post translational modifications (PTMs) that affect its function. These modifications may not happen when the protein is heterologously expressed in *E. coli* due to its relatively limited PTM capacity (Cain et al., 2014). Since the murine and human HEXDC were previously successfully expressed

in *E. coli* (Gutternigg et al., 2010), we presumed that dmHexDC may also be functional when expressed in *E. coli* cells.

Another important variable in protein expression is the solubility of the recombinant protein. High levels of expression of a hydrophobic protein may cause the hydrophobic regions to interact with other expressed hydrophobic host proteins. Such interactions of hydrophobic stretches can form insoluble aggregates known as inclusion bodies. The toxic effect of inclusion bodies can hamper the growth of the host organism greatly and therefore the expression and isolation of the protein of interest becomes a more difficult task. I have analyzed the hydrophobic regions of dmHexDC using multiple prediction tools (CCTOP, HMMTOP, MemBrain, Memsat, Octopus, Pro, Prodiv, Scampi, ScampiMsa, TMHMM). The results suggest that dmHexDC has a putative transmembrane domain between amino acids 13 and 35 (Figure 4.2). Therefore, I concluded that expression of dmHexDC could be hampered because of the hydrophobic transmembrane domain.

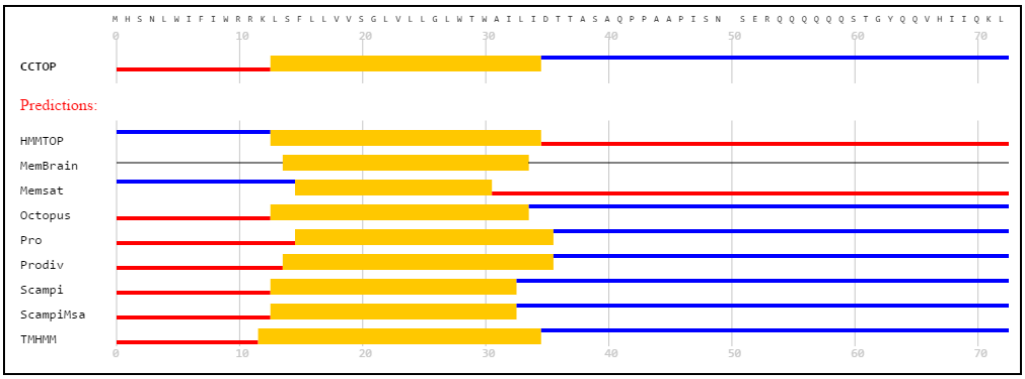

Figure 4.2. Predictions done by multiple servers (left) suggest that dmHexDC contains an N-terminal transmembrane domain between amino acids 13 and 35 (amino acids in the yellow region).

4.2.2. Cloning of two dmHexDC constructs into pETM-20 vector

Keeping in mind that hydrophobic regions can affect the solubility of the protein product, I decided to clone dmHexDC into an expression vector that contains a protein tag to help solubilize the protein during bacterial protein expression. The pETM-20 vector (see Appendix for vector map) chosen here contains an N-terminal Thioredoxin A (TrxA) tag. The TrxA tag originates from *E. coli* and is used to solubilize heterologously expressed proteins (Young et al., 2012). However, even with the solubility tag, there was still a possibility that dmHexDC might not be soluble. Therefore, two versions of dmHexDC were

cloned for bacterial expression, a full-length and a truncated version without the transmembrane domain. The constructs are shown in Figure 4.3.

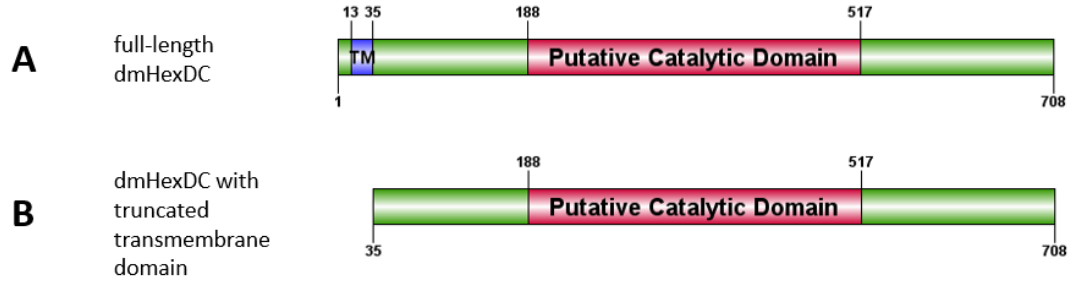

Figure 4.3 Constructs for dmHexDC expression. (A) Full-length dmHexDC construct that consists of transmembrane domain and putative catalytic domain. (B) Truncated dmHexDC construct that does not contain a transmembrane domain.

Besides an N terminus TrxA, pETM-20 contains a 6xHis tag, which can be utilized for immobilized metal affinity purification (IMAC). Histidine has an imidazole ring as a side chain. Electron donor groups of imidazole cause strong interaction with transition state metals. IMAC takes advantage of the interaction between histidine and immobilized transition state metals such as nickel, thus making the purification of proteins that contains 6-9 histidine repeats possible.

The full-length and transmembrane region truncated versions of dmHexDC were amplified (Figure 4.4A) with primers that contain 5' *Nco*I and 3' *Kpn*I restriction sites and cloned into pETM-20 by restriction ligation**.**

Six colonies for each cloning experiment were selected and subjected to restriction digestion with *Nco*I and *Kpn*I enzymes for (Figure 4.4B). Positive colonies were sequenced and confirmed the correct cloning of the desired fragments. Colonies 2 and 8 were chosen for further experiments.

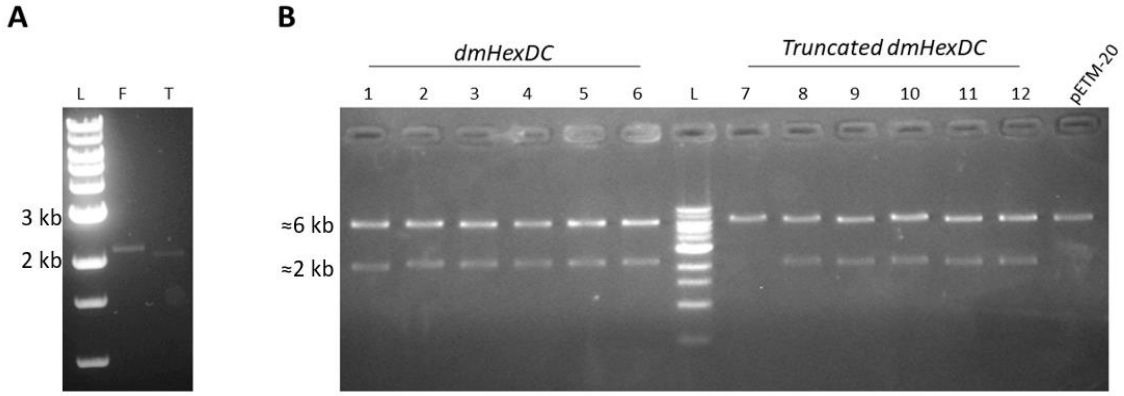

Figure 4.4. Amplification and cloning of dmHexDC. (A) Full-length and truncated versions of dmHexDC were amplified with primers that contain *Nco*I and *Kpn*I restriction sites. The PCR products were then gel purified. (B) Both truncated and full-length versions of dmHexDC CDS were cloned into the pETM20 vector.

Isolated plasmids were transformed into JM109 *E.coli* competent cells that carry T7 RNA polymerase under the lacUV5 promoter. LacUV5 promoter is repressed by LacI. The presence of IPTG represses the binding activity of LacI to Lac promoters in both vectors and chromosome, allowing expression of T7 polymerase and as a consequence, expression of truncated and full-length dmHexDC.

4.2.3. dmHexDC is expressed for 4 hours after IPTG induction

Bacteria were grown at standard conditions and different expression parameters including 1, 2, 4, and 16 hours of IPTG induction and growth at 37°C (Figure 4.5A, 4.5B) and 25°C (Figure 4.5C, 4.5D) were tested. Expression of truncated dmHexDC was high at 4h and 16h after IPTG induction at 25°C. Similar results were obtained for the same experiment at 37°C. 1 mL of culture was collected at the given time points after IPTG induction and proteins were precipitated by trichloroacetic acid (TCA) precipitation. At low concentrations TCA disrupts the intra-molecular electrostatic interactions (Rajalingam et al., 2009). This disruption of the protein structure results in exposure of non-polar surfaces and the structural intermediates reversibly precipitate due to intermolecular interaction of such surfaces. The collected samples were analyzed by SDS-PAGE. Successful protein extraction resulted in a thick band between 100-75 kDa. Further protein expression experiments were performed at 25°C since it is closer to the physiological temperature of *Drosophila melanogaster* and protein extraction was performed 4 hours after IPTG induction.

The total number of amino acids that encode dmHexDC is 708. Thus, the predicted mass of wild-type dmHexDC is 81.2 kDa. Since I expressed dmHexDC heterologously with TrxA tag and 6xHis tags, the approximate size of the full-length TrxA::6xHis::dmHexDC (will be referred to as full-length dmHexDC) is composed of 823 amino acids and has a size of 93.84 kDa. TrxA::6xHis::truncated dmHexDC (will be referred to as truncated or ΔdmHexDC) is composed of 789 amino acids and its size is 89.79 kDa.

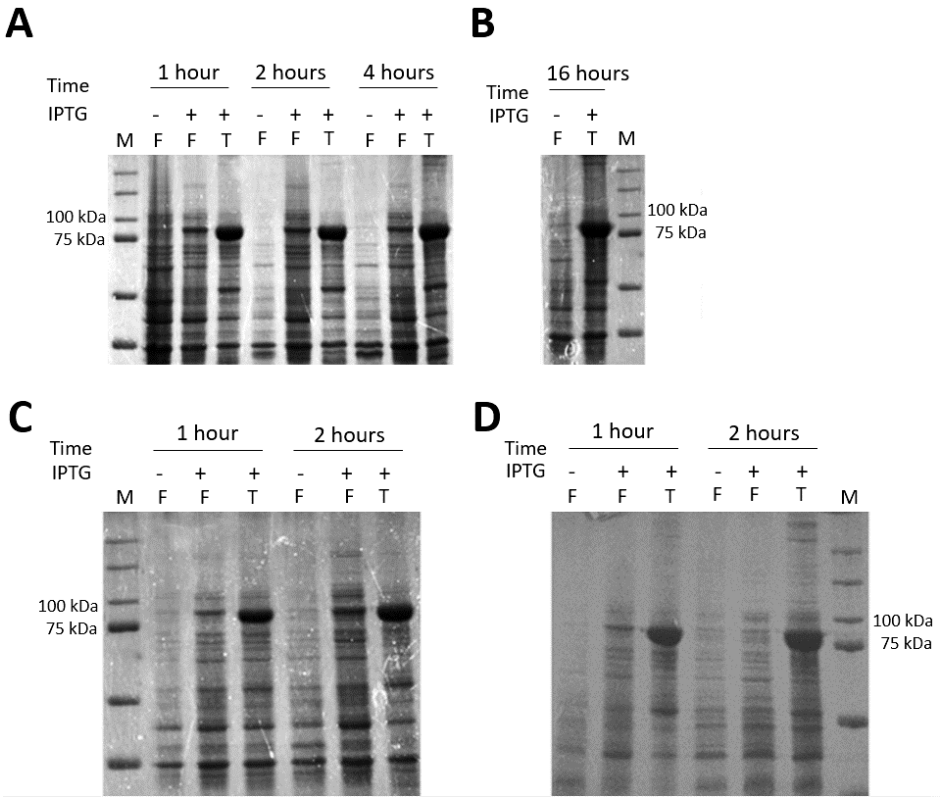

Figure 4.5. Expression of dmHexDC constructs. (A) Expression of dmHexDC construct in 1.2 and 4 hours after IPTG induction at 37°C. (B) Expression of dmHexDC constructs after 16 hours after IPTG induction at 37°C. (C) Expression of dmHexDC construct 1 hour and 2 hours after IPTG induction at 25°C. (D) Expression of dmHexDC constructs 4 hours and 16 hours after IPTG induction 25°C. L: Protein Ladder, (-): No IPTG induction, F: Full-length dmHexDC construct, T: Truncated dmHexDC construct.

The result of the initial protein expression experiments showed that truncated dmHexDC is expressed at much higher rates than full length dmHexDC.

4.2.4. Western Blot confirms dmHexDC expression

In order to confirm that the expressed protein is dmHexDC and not a frameshifted version of it, western blot analysis using a dmHexDC antibody generated in our lab (Kıral, 2015) and an antibody against the His-tag were performed (Figure 4.6).

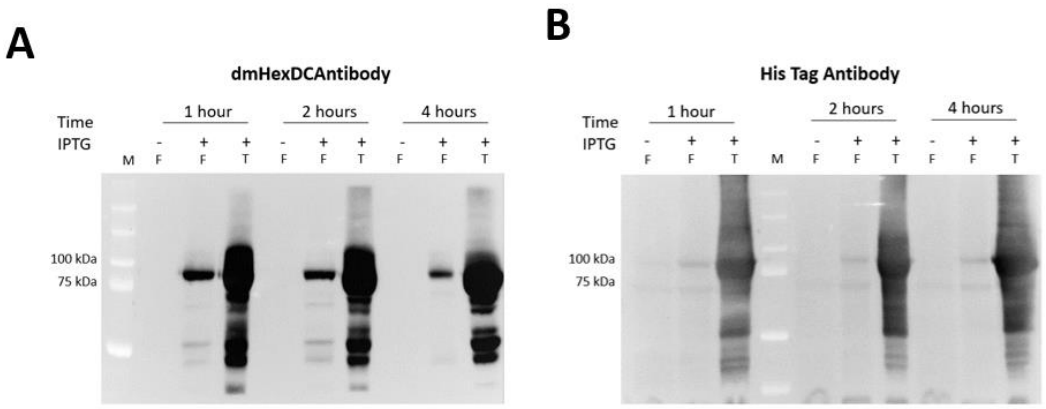

Figure 4.6. Western blot of dmHexDC constructs at different time points after IPTG induction. (A) Western Blot with dmHexDC antibody. (B) Western Blot with His-tag antibody.

The Western blot analysis confirmed that the high intensity bands observed in the Coomassie staining corresponded to heterologous dmHexDC.

4.2.5. Purification of dmHexDC

Initial experiments showed that the truncated version of dmHexDC is more efficiently expressed than the full-length enzyme, therefore, I aimed to purify the truncated protein. Purification was performed with the AKTA prime liquid chromatography system using columns containing Nickel beads. The AKTA prime system allows the purification of proteins in an automated manner, while measuring UV and conductivity, generating gradients and collecting fractions. Thus, the measurements performed by the system inform the user if the protein is eluted. However, despite several attempts, no protein could be purified.

In order to investigate at which step purification of truncated dmHexDC was hindered, I analyzed the samples obtained at each step of the purification process. Crude extract containing truncated dmHexDC after IPTG induction (T), the flow through samples that are collected after loading the crude extract to the nickel bead columns (Ft1-2), fractions obtained by washing the nickel bead column with imidazole containing binding buffer (Fr1-

5) and elution samples obtained by a gradient of high concentrations of imidazole containing buffers (E1-E11) were analyzed by SDS-PAGE. The results showed that dmHexDC is lost after being loaded onto the purification columns (Figure 4.7). A possible reason could be that truncated dmHexDC is still being partly insoluble and hydrophobic interactions could still be occurring. This may result in the generation of an insoluble aggregate of dmHexDC and other proteins on nickel beads.

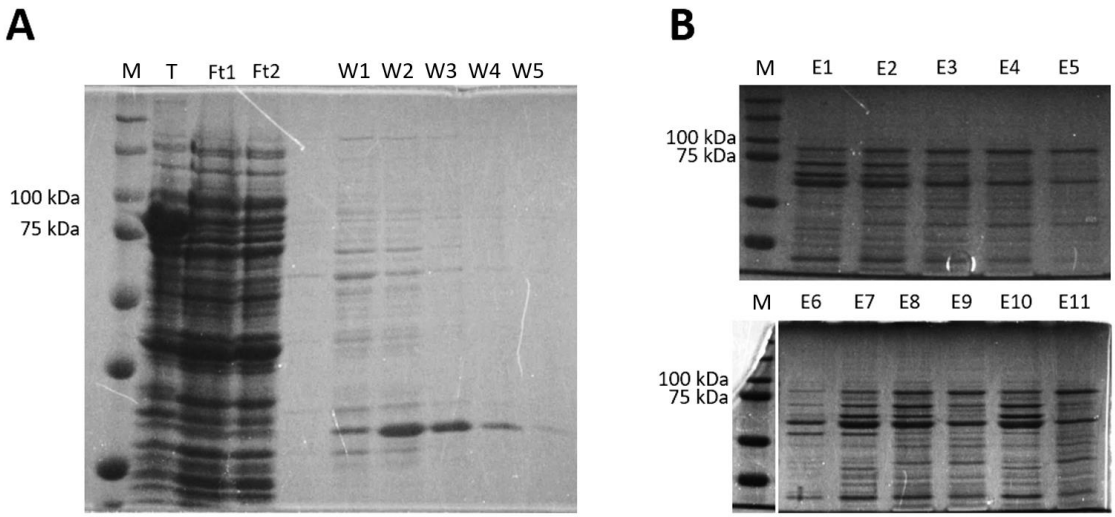

Figure 4.7 Purification of trucated dmHexDC (A) Expressed Truncated dmHexDC was loaded on nickel bead columns and the majority of dmHexDC was bound to nickel beads (lane T). (B) The column was washed with imidazole containing binding buffer. Washed proteins were eluted from columns by elution buffer. Different concentration of elution buffer have been used.

L: Protein Ladder, T: Truncated dmHexDC sonicated 4 hours after IPTG induction, Ft: Flow throughs after T(4h) has been loaded into nickel bead column, Fr: Flow through after nickel beads were washed with binding buffer, E: Flow through after elution buffer.

4.2.6. N-terminus TrxA and His-tagged, truncated dmHexDC cleaves terminal N-acetyl beta-D-galactosamine and has a pH optimum at pH 6.0

The purification of dmHexDC is not essential to show its substrate specificity (Gutternigg, 2010), thus I used the whole lysate to test the heterologously expressed protein. *In vitro* experiments were performed using crude bacterial extracts for the detection of hexosaminidase activity.

Soluble proteins were extracted from JM109 cells by sonication. The experiments measured the catalytic activity of crude *E. coli* extracts against p-nitrophenyl N-acetyl beta-

D-galactosaminide hydrolysis of terminal N-acetyl-beta-D-galactosamine residues frees para-nitrophenol (pNP) and at alkaline conditions para-nitrophenolate gains chromogenic properties and its absorption of light at 405 nm can be measured. Experiments were performed at different pH conditions to determine the optimal pH. A pH range of pH 3-8 was chosen, which has previously been described to be optimal for hexosaminidases. No activity was observed at pH3.0 and 3.5 while at pH4.0 significant activity was observed as compared to controls and rose steadily until pH6.0 (Figure 4.8). Raw absorbance (405 nm) data were subjected to Student's T-test. We concluded that truncated dmHexDC cleaves p-nitrophenyl N-acetyl beta-D-galactosamine at pH 4.5-8.0.

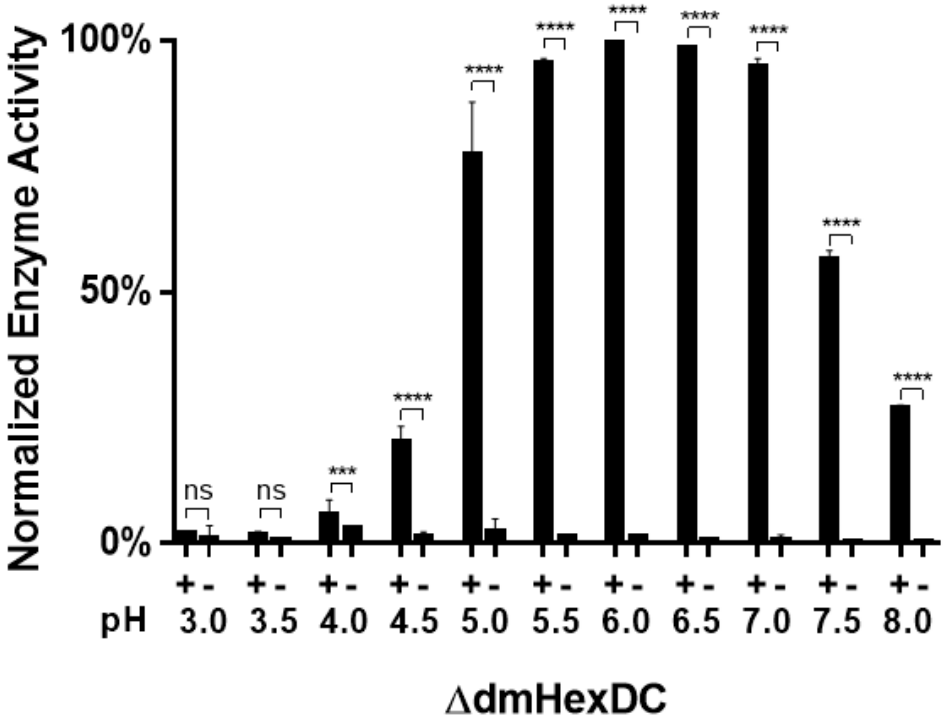

Figure 4.8. Hexosaminidase Assay Against 4-Nitrophenyl N-acetyl-ß-D-galactosaminide between pH 3.0-pH 8.0. 3 biological replicates were tested for each experimental group. The result of the enzyme assay shows that there is a significant difference between truncated dmHexDC containing lysates (+) and lysates that do not contain truncated dmHexDC (-) between pH 4.5-8.0 (pH=4.5, *** $p<0.001$, pH>4.5 **** $p<0.0001$) but there is no difference in terms of catalytic activity below pH 4.5 (ns

p>0.05). Absorbance (405 nm) data from the hexosaminidase assay were normalized with 100% enzyme activity from the highest data value.

4.2.7. *In silico* modelling of dmHexDC to identify catalytically important residues

I next aimed to identify catalytically important residues of dmHexDC. Therefore, I did comparative modelling (homology modeling) to estimate the location of alpha carbons of dmHexDC *in silico*. The MODELLER program is used to generate 3D structures of proteins by comparative modeling by the spatial restraints for a given amino acid sequence (Eswar et al., 2006). For visualization and analysis of in silico modeling, I used UCSF Chimera. The PSI-BLAST algorithm, a sensitive iterative search for evolutionary conserved domains, was chosen and the matrix BLOSUM 80 was used to identify amino acid sequences similar to dmHexDC in the PDB database. The PDB database contains the information of coordinates that are experimentally obtained from the spatial distribution of atoms in proteins. Structures with the highest score for different regions were selected and used to generate 5 output models (see Appendix). Only the putative catalytic site of the enzyme had a 3D structure probably due to the catalytic site being the only conserved domain in the template structures.

Lastly, I-TASSER, the top ranking protein prediction server, was used (Yamasaki et al., 2018). No restraints were assigned for the modeling job. The structure obtained from I-TASSER server was the only structure that contained every region of dmHexDC (Figure 4.9A). The transmembrane domain of the dmHexDC was modeled as a random coil and therefore I concluded that the structure of the transmembrane domain was not modeled properly. The structure of the dmHexDC homology model resembles a common tertiary structure called TIM barrel which usually contains a common $(\beta\alpha)_8$-barrel (See Appendix). The distinctive property of this structure is that it contains 8 α helixes and 8 β strands closing in the structure to form a doughnut-like shape. The 8 β strands are found in the most inward part of the doughnut-like topology and α helices surround the β barrels (Wang et al., 2017). This structure is enzyme specific and interestingly the proteins that contain these structures can have very little sequence similarity (Yamasaki et al., 2018). The dmHexDC structure appears to be similar but not identical to this topology. It is also suggested that catalytic amino acids of $(\beta\alpha)_8$-barrels reside in the C terminal part of the β strands or at the loop regions between the inner β and α strands. I-TASSER results also reported a hexosaminidase function prediction based on structure and sequence alignments.

The "low-resolution" structure was generated for more structural approximation towards the catalytic site in the protein, which guided the design of new experiments such as site directed mutagenesis to generate a dmHexDC mutant that is unable to hydrolyze its substrate.

In order to gain more insight on where the catalytic amino acids of dmHexDC reside, I did a structural alignment between dmHexDC and a structurally similar *Streptomyces plicatus* hexosaminidase captured during interaction with NAG thiozoline (Wang et al., 2017), a competitive inhibitor for hexosaminidases (T. Liu et al., 2015). The amino acids that are spatially closest to the catalytic amino acids of *Streptomyces plicatus* hexosaminidase were aspartic acid and glutamic acid at the positions 346 and 347 and were identified as the two amino acids that are utilized directly in the hydrolysis of terminal N-acetyl-beta-D-hexosamine containing residues (Figure 4.9B).

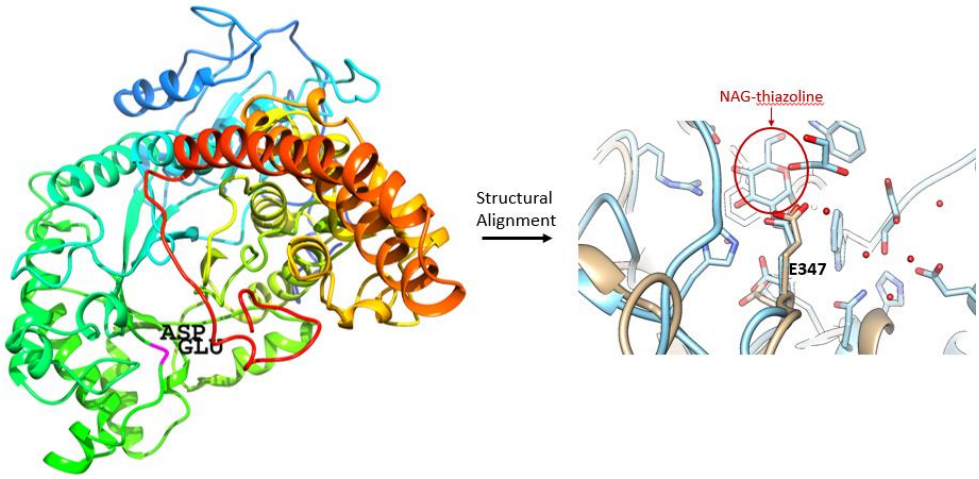

Figure 4.9 Homology model of dmHexDC was generated by comparative modeling. (A) Structure obtained using I-TASSER server. Catalytic region is shown in purple. (B) The generated homology model of dmHexDC and structural alignment with *Streptomyces plicatus* beta-N-acetylhexosaminidase complexed with intermediate analogue NAG-Thiazoline shows D346 and E347 is closest to the competitive inhibitor NAG-Thiazoline.

4.2.8. Site directed mutagenesis to identify the catalytic amino acid

It was shown in human hexosaminidases that Glutamic Acid acts as the general acid-base residue whereas Aspartic Acid helps in the orientation of C2 acetamido group to prepare it for nucleophilic attack and stabilizes the positive charge on the oxazolinium ion intermediate. Thus, the model verified that E347 is the general acid-base residue for the catalytic activity. Therefore, to disrupt the nucleophilic attack and mimic the loss of catalytic activity a site-directed mutagenesis experiment was performed to introduce the E347A mutation in the truncated dmHexDC containing pETM-20. Primers containing the mutation were generated and used in a PCR reaction to amplify the truncated plasmid (Figure 4.10). Colonies obtained after transformation with the PCR products were isolated and the presence of the E347A mutation was validated by sequencing.

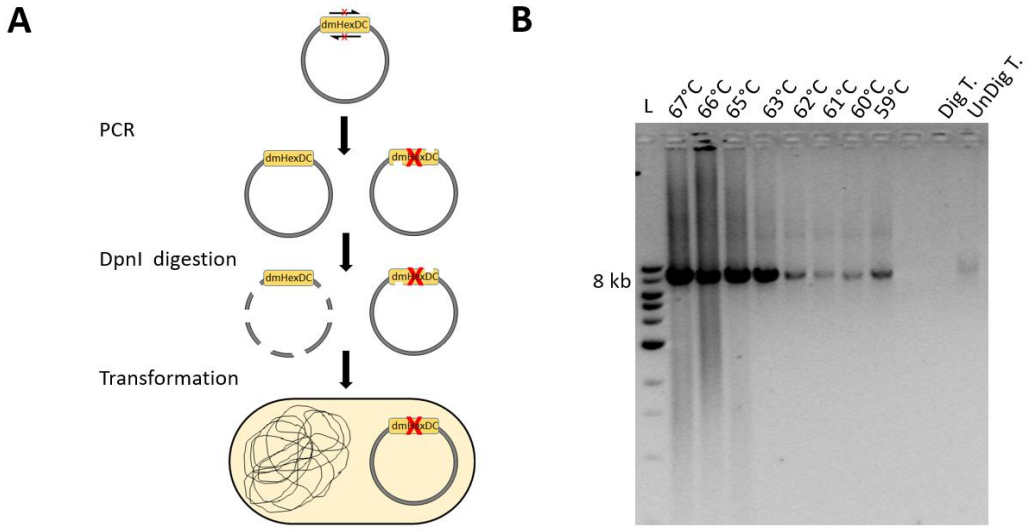

Figure 4.10 Site directed mutagenesis was made to disrupt the catalytic activity. (A) Schematic representation of the site directed mutagenesis. (B) Gradient of temperatures were used for annealing during site directed mutagenesis PCR. L: Ladder, 67-59°C: Annealing temperatures, DigT: DpnI digested template, UnDig T.: Undigested Template.

4.2.9. Glutamic acid at position 347 is the general-acid base residue for catalytic activity

The next step to evaluate the E347A mutant as the general acid-base residue needed for the catalytic activity of dmHexDC was to express the protein and test the hexosaminidase activity. The E347A mutant was expressed exactly like truncated dmHexDC and the

bacterial lysates containing the expressed proteins were collected using the same method (Figure 4.11).

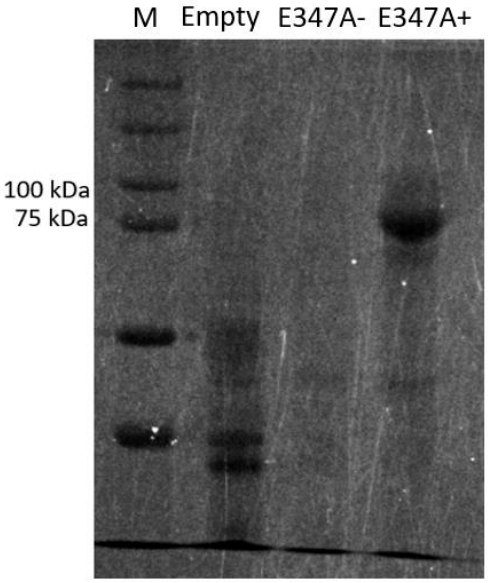

Figure 4.11. Expression of truncated dmHexDC after site-directed mutagenesis. Truncated dmHexDC is expressed for 4 hours after IPTG induction. L: Protein ladder, M: Mock Plasmid, (-): IPTG uninduced for 4 hours after OD_{600} =0.6, E347A: IPTG induced expression of E347A mutant.

After the extraction of E347A mutant containing extracts, the hexosaminidase assay was performed against p-nitrophenyl GalNAc as previously described (Figure 4.12).

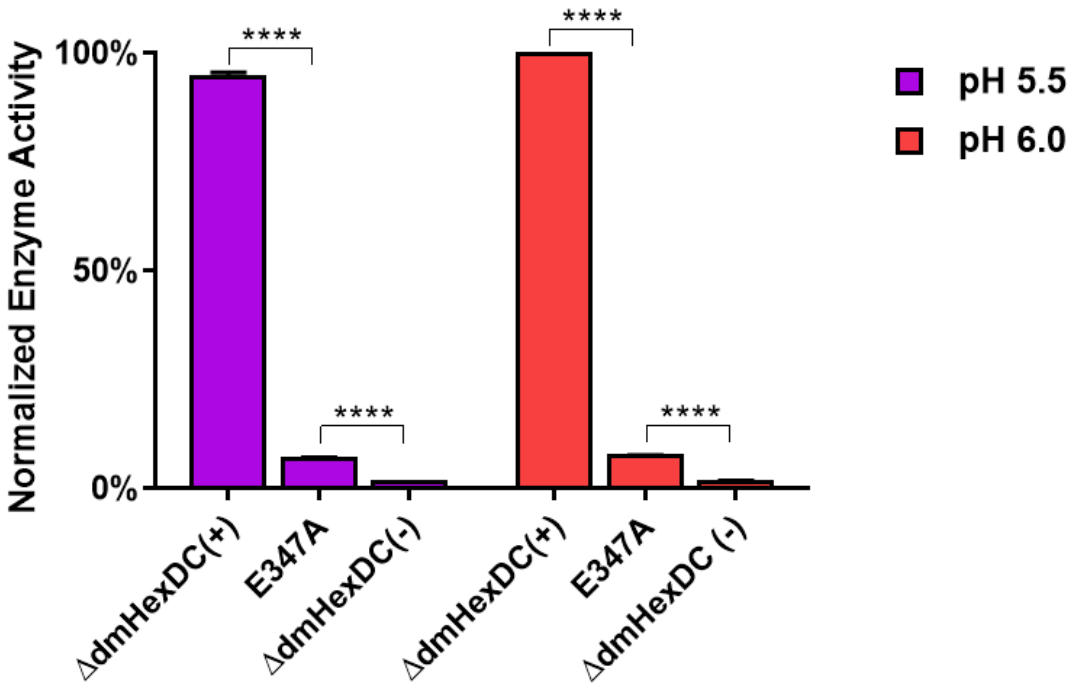

Figure 4.12. Hexosaminidase assay shows high catalytic activity against p-nitrophenyl GalNAc. E347A containing lysates shows diminished catalytic activity compared to ΔdmHexDC (+) lysates between pH 5.5 and pH 6.0 (**** $p<0.0001$). Also there is a significance difference between E347A and ΔdmHexDC (-) lysates.(**** $p<0.0001$).

The observed activity of E347A is approximately 7% and 7.5% in pH 5.5 and 6.0 respectively. This data suggests the catalytic amino acid is indeed E347. Since the aspartic acid and glutamic acid are next to each other according to the hexosaminidase catalytic motif, this also suggests that possibly D346 also takes part in electrostatic stabilization of the substrate for nucleophilic attack by E347.

4.2.10. dmHexDC cleaves GalNAc and GlcNAc residues, but the substrate preference is GalNAc

Homologues of dmHexDC include human and murine HEXDC and *C. elegans* Hex-2, Hex-3 and Hex-5 (Kaçmaz, 2013). I showed that dmHexDC cleaves GalNAc residues. However, homologues of dmHexDC also cleave GlcNAc, possibly increasing the range of their glycosylated physiological substrates. Even though the homologues of dmHexDC are active against GlcNAc, they prefer GalNAc in terms of catalytic activity. Human HexA, HexB and HexS are also known to cleave both substrates. To clarify the substrate specificity of dmHexDC I tested p-nitrophenyl N-acetyl beta-D-glucosamine substrate in a hexosaminidase assay. Both substrates were tested between 3.0-8.0. The assay showed

dmHexDC cleaves both GalNAc and GlcNAc residues, but it is clear that the substrate specificity of dmHexDC is towards GalNAc (Figure 4.13). This suggests that glycoconjugates that contains a GlcNAc at their terminals may also be physiological substrates for dmHexDC.

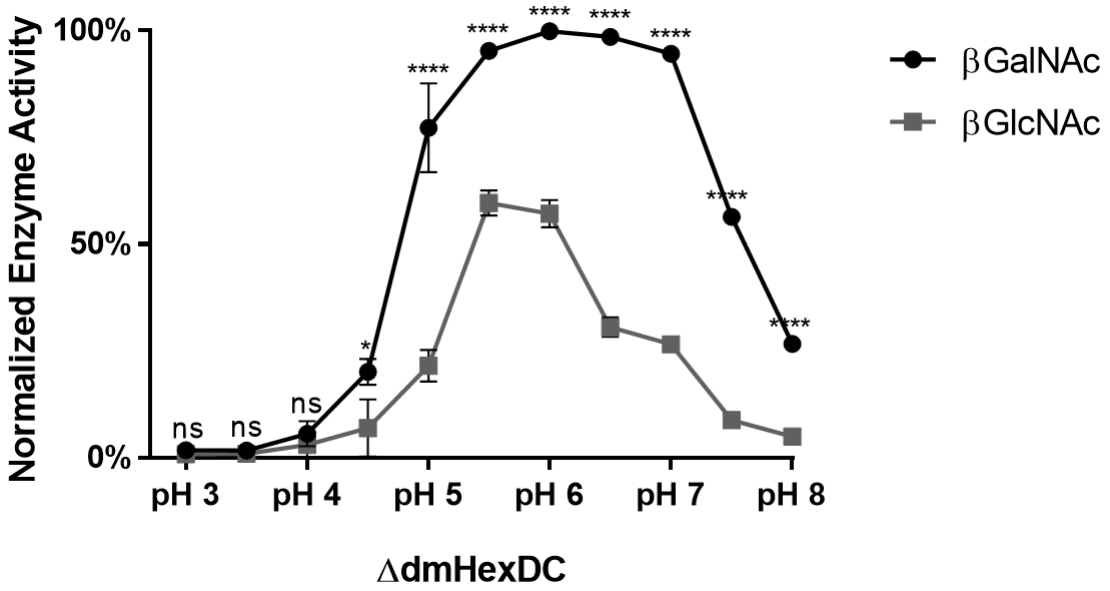

Figure 4.13. Hexosaminidase assay was performed with bacterial lysates between pH 3.0-8.0 using GalNAc and GlcNAc as substrates. There is significance difference between GalNAc and GlcNAc in terms of catalytic activity of truncated dmHexDC at pH 4.5 (* $p<0.05$) and between pH 5.0-8.0 (**** $p<0.0001$).

4.3. Optimization of dmHexDC antibody to study cellular localization of dmHexDC

In order to detect localization of dmHexDC, I used three transgenic lines, two of which were generated in our lab. These lines consisted of an enhancer trap line with a GAL4 integrated into the first intron of dmHexDC, an overexpression line that was generated by BAC recombineering, dmHexDC::eGFP (BAC), and a commercially available protein trap (P.T) line that interferes with the transcription of *dmHexDC* and expresses YFP from the first intron of *dmHexDC*. However, all of these lines have their drawbacks and it is unclear if any of them represents the endogenous expression and/or localization of dmHexDC. Enhancer-trap line expresses the GAL4 transcription factor and can be used to express of any DNA sequence under UAS, the GAL4 target sequence, control. This does efficiently mark the cells but does not show the subcellular localization of dmHexDC and might not recapitulate the whole endogenous expression pattern. dmHexDC::eGFP line overexpresses

dmHexDC and thus might not show the endogenous localization. It is also possible that dmHexDC::eGFP is not expressed under the same regulatory mechanisms compared to endogenous dmHexDC since the construct might be lacking some distant regulatory regions. Furthermore, the GFP fused to the protein might hinder its proper localization. The protein trap line does express YFP instead of dmHexDC therefore it is not possible to obtain the subcellular localization from this line. The expression patterns from these different lines have been analyzed previously (Kıral, 2015). In addition to these constructs, a polyclonal peptide antibody was generated and initially tested in the eye imaginal discs with poor results (Kıral, 2015). Thus, in this study, I aimed to assess the usefulness of this antibody further and tried to optimize it for immunostainings.

First, I pre-adsorbed the dmHexDC antibody with tissues taken from the protein trap line, generated protein extracts from wild-type (w^{1118}), BAC and P.T. larvae and tested the antibody on western blots. In wild-type, a band near 82 kDa band was expected, while two bands were expected in the BAC sample since BAC contains endogenous dmHexDC and dmHexDC::eGFP, which have been calculated to have a mass of 82 kDa and 120 kDa, respectively. Moreover, no band in the P.T. sample was expected as we determined this line to be a null mutant for dmHexDC (Kıral, 2015). Western blot results showed an unexpected but clear band around 180 kDa in all protein extract samples (Figure 4.14A). This band was significantly weaker in the P.T. sample. This result indicated the probability of non-specific binding. Thus, the truncated dmHexDC extract generated from heterologous expression in *E. coli* was used as control for dmHexDC specificity and showed a clear band at the expected dmHexDC size (Figure 4.14B) and thus I concluded that the antibody recognizes heterologously expressed dmHexDC, but not dmHexDC isolated from tissues. A second band in the Western blot could indicate cleavage of truncated dmHexDC from the N-terminus by *E.coli* proteases.

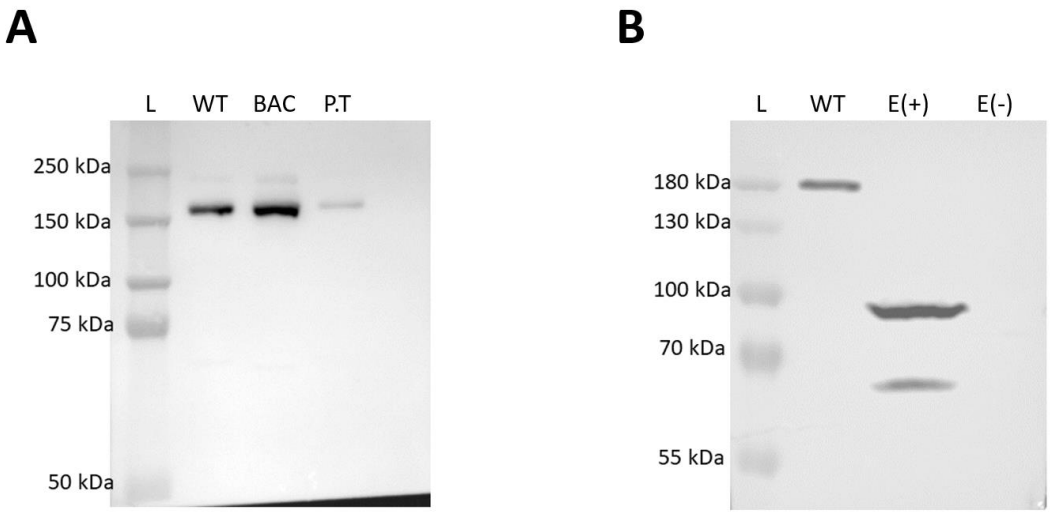

Figure 4.14 Western blot analysis of dmHexDC antibody specificity. Western blot on larval tissue protein extracts from WT, BAC and P.T. lines. (A) Band with a mass of 180 kDa for WT BAC and P.T tissue extracts was observed. (B) Extracts that contained truncated dmHexDC (E(+)) were tested. The lower band at is possibly due to a N-terminal cleavage event during protein expression.

4.3.1. Peptide blocking to identify dmHexDC bands

To conclude that the observed band contains the peptide that the antibody was generated against, a peptide blocking experiment was performed. Since this antibody was synthesized by a peptide sequence from dmHexDC, preincubating the dmHexDC antibody with the peptide is expected to make the specific bands disappear. I incubated the antibody against its peptide at 1µg/mL and 10µg/mL O.N at 4°C before Western blot analysis.

The result of the Western blot showed that the band I was observing in larval tissues does not contain the peptide sequence, because those bands did not disappear when blocked with different concentrations of peptide like the extract that contains truncated dmHexDC (Figure 4.15). Considering the possibility that the hydrophobic regions of dmHexDC may have impaired its extraction from larvae, I continued with IHC experiments which gave inconclusive results (see Appendix).

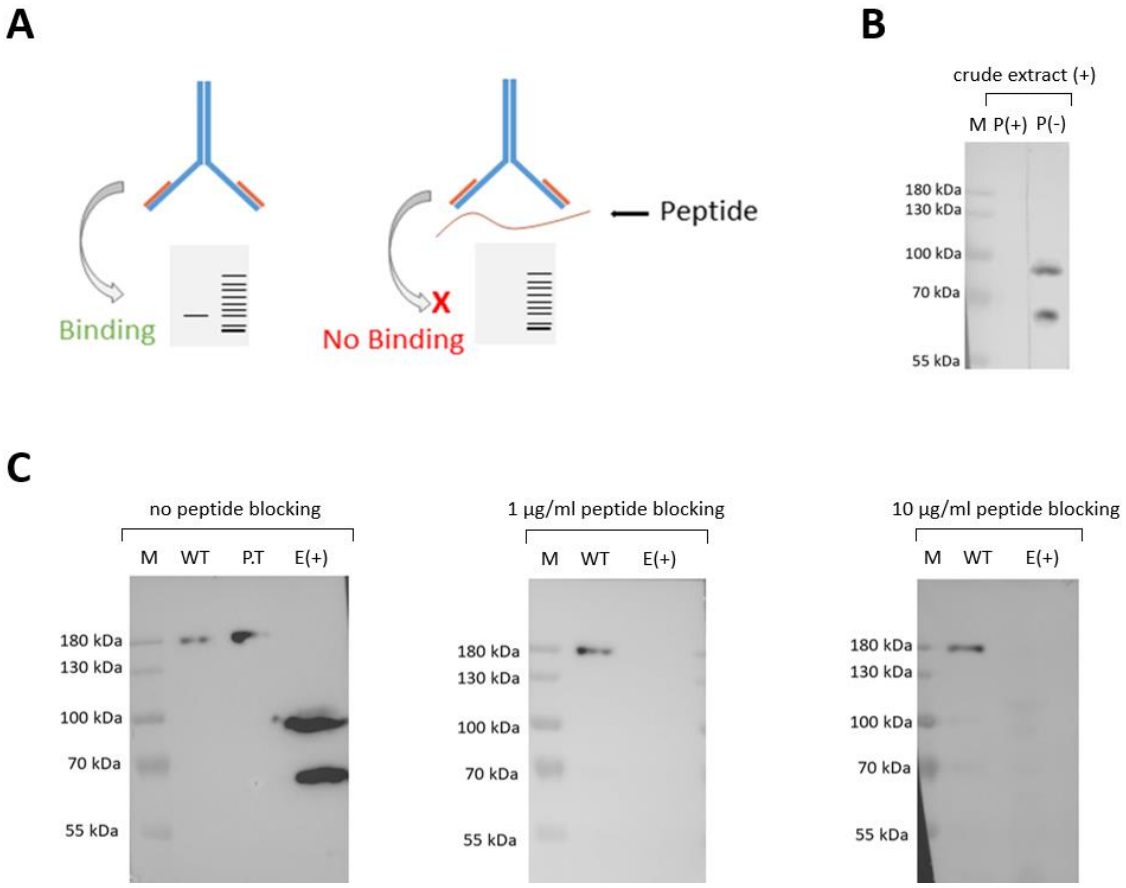

Figure 4.15. Peptide blocking experiment to assess the specificity of the 180 kDa protein observed in Western blots from larval tissues. (A) Schematic representation of the experiment. (B) The lane of truncated dmHexDC containing crude extract was cut in half and blocked to show the peptide blocking. (C) 180 kDa did not disappear when blocked.

4.4. Generation of a catalytic mutant with CRISPR/Cas9

4.4.1. Strategy overview

Previous analysis made with the NHEJ and *CG7985*^{CPTI100032} knock-out mutants showed that dmHexDC mutants were embryonic lethal with some exceptions. In this study we aimed to generate a hypomorph for dmHexDC and for this purpose we decided to utilize CRISPR/Cas9 technique to make a point mutation to the nucleophilic residue of dmHexDC. We hypothesize that dmHexDC mutants with diminished catalytic activity will show milder phenotypes in the context of lysosomal storage disorders, similar to human HexA (Leinekugel et al., 1992). Higher survival rates will allow us to investigate and analyze the role of dmHexDC in cell fate determination in later stages of the *Drosophila* life cycle.

CRISPR/Cas9 can utilize two pathways to generate mutations, non-homologus end joining (NHEJ) or homology directed repair pathways (HDR). In order to introduce a point

mutation to the catalytic amino acid of dmHexDC, we aimed to use the HDR system (Figure 4.16).

Activation of the HDR pathway relies on a double strand break, and a donor DNA with homology to the region of interest. A double-stranded break is obtained by a small guide RNA molecule guiding the endonuclease Cas9 to the target site. When endonuclease activity occurs the cell will respond to repair this damage. We planned to introduce a donor DNA that the cell can use to repair the DSB damage thereby introducing a mutation that will result in a decrease of dmHexDC catalytic activity. For this purpose, I generated a gRNA construct and a homology template DNA. These constructs were injected to embryos expressing Nos-Cas9. These flies express Cas9 under the *nanos* promoter and 3'UTR elements allowing Cas9 to be expressed specifically in germ cells and increasing the probability of germ-line transformation. This system also allows for the study of lethal mutations.

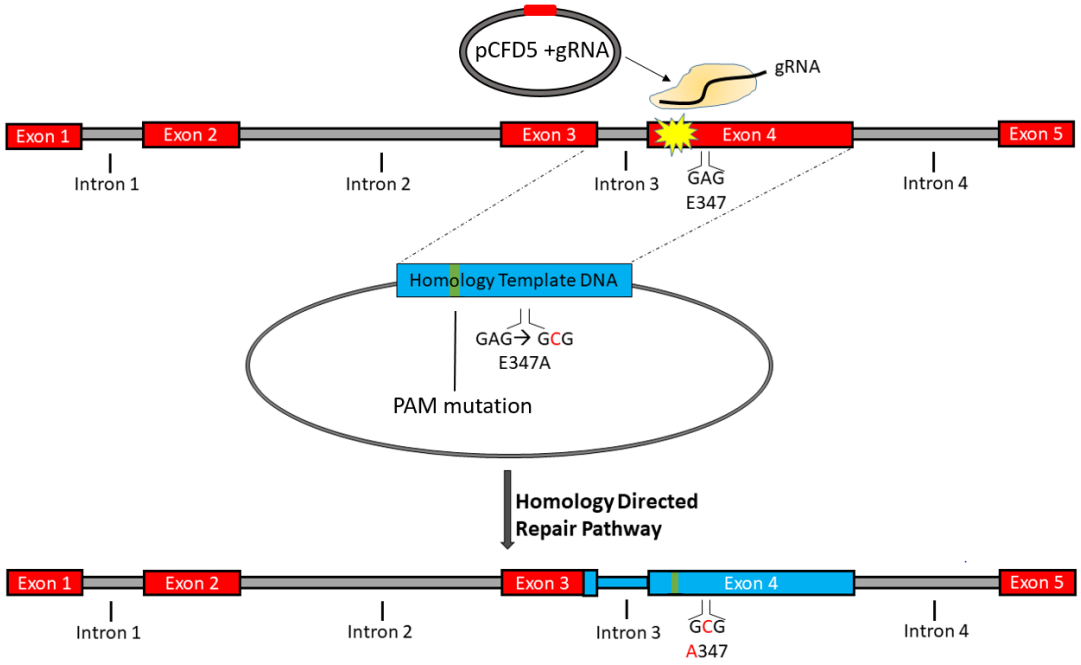

Figure 4.16 Strategy to generate a hypomorph of dmHexDC using CRISPR/Cas9. gRNA will be expressed from pCFD5 vector and will guide Cas9 near the codon of dmHexDC and Cas9 will cleave DNA, causing a DSB 3 nucleotides near the PAM site. Cell will repair this damage by incorporating the homology template DNA. Homology template DNA will contain the mutation to mutate nucleophilic residue to alanine, E347A. The donor DNA also contains a silent mutation for the PAM site to protect homology template DNA from Cas9 (Green region).

4.4.2. Generation of a Homology template

Our strategy was to generate a homology template and a gRNA and inject the two DNA constructs to Nos-Cas9 embryos, thus, the genomic region around dmHexDC in Nos-Cas9 flies was analyzed. In particular, genomic DNA from Nos-Cas9 adult flies was amplified and sequenced and it was verified that these flies do not contain any single nucleotide polymorphisms (SNPs) that could disrupt the CRISPR experiment. In order to generate a homology template 1000 bp upstream and downstream homology arms flanking the double stranded break site were chosen and amplified using genomic DNA from Nos-Cas9 flies as template. The commonly used pGEM®-T -Easy vector was used as the vector backbone. pGEM®-T–Easy is a linearized vector with 5' thymine overhangs. This allows circularization with a construct that has 3' adenine overhangs. These non-template adenine overhangs are added to the DNA by Taq polymerase. I amplified the 2 kb homology region using Q5® High-Fidelity DNA polymerase. Since Q5® High-Fidelity DNA polymerase has proof reading activity it is less likely to introduce any unwanted mutations and it leaves the amplified DNA blunt-ended. I purified the amplified 2 kb DNA from agarose gels and incubated it with Taq polymerase for the introduction of 3' adenine overhangs.

The homology DNA was cloned into the pGEM®-T–Easy vector by TA cloning. Colonies were tested by analytical digestion with *Eco*RI. Plasmids that contained an additional 2kb band were sequenced for verification (Figure 4.17).

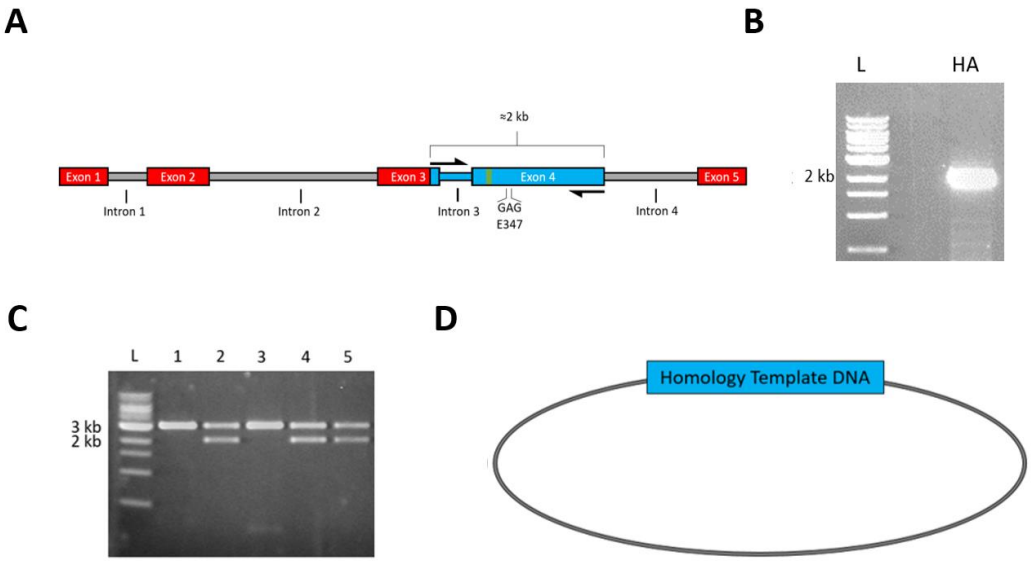

Figure 4.17.Cloning of homology template. (A) Schematic representation of the 2044 bp amplified region (turquoise). (B) Donor DNA was amplified. (C) 2 kb homology arm was

cloned into to pGEM®-T –Easy vector. Samples with 2kb bands were considered as positive and validated by sequencing. (D) Representation of homology template.

In the next step the plasmid containing the homology arms was subjected to site-directed mutagenesis to introduce the E347A mutation (Figure 4.18). The PCR product was digested with the *Dpn*I enzyme to digest the original template DNA before transformation into bacteria. Plasmids were isolated and the presence of the point mutation was validated by sequencing.

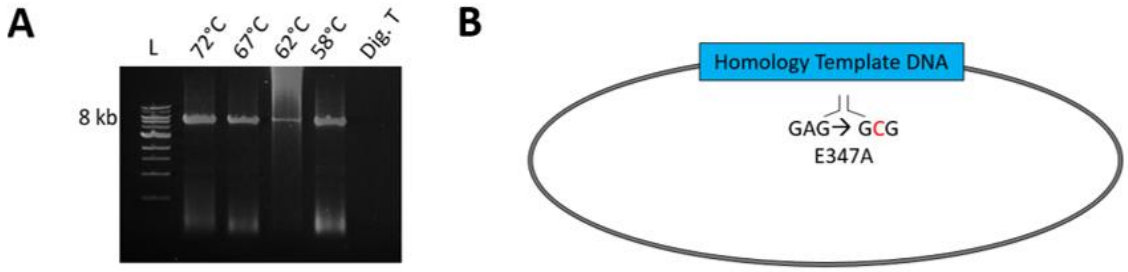

Figure 4.18 Generation of catalytic mutant of dmHexDC on homology template DNA. (A) A gradient of temperatures were applied for the annealing step of site directed mutagenesis PCR and setup at 72°C annealing was selected for transformation. (B) Representation of homology template DNA plasmid containing E347A mutation.

In the last step, to increase the efficiency of the recombination rate the PAM site close to the selected gRNA was mutated by site-directed mutagenesis. This was necessary as this sequence can be targeted by Cas9 as well and thus decrease targeting of Cas9 to the genomic site. A silent mutation was introduced to the threonine at position 339 and as a result GGN, the PAM site on the negative strand, was converted to CGN. This mutation will protect the homology template DNA plasmid from being cleaved by Cas9 and does not disrupt the threonine at position 339 as it is a silent mutation. The site-directed mutagenesis experiment was performed and the presence of the mutation was verified by sequencing (Figure 4.19). The plasmid containing the E347A and the PAM site mutation was prepared in large scale for injection into *Drosophila* embryos.

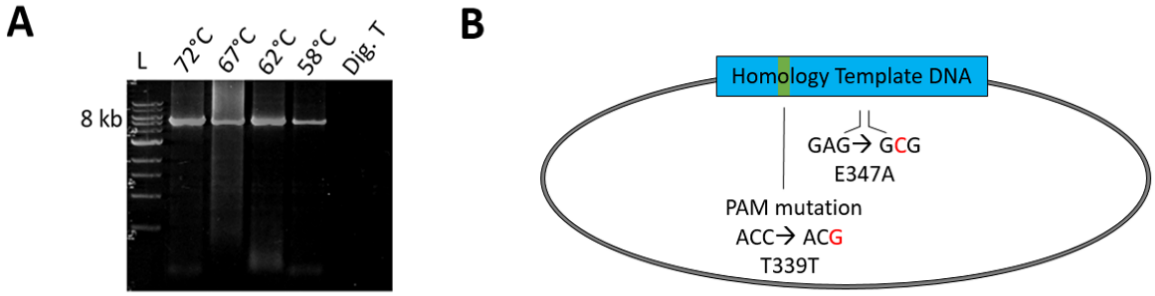

Figure 4.19. Generation of PAM mutant on homology template DNA. (A) A gradient of temperatures was applied for the annealing step of site directed mutagenesis PCR and setup at 72°C annealing was selected for transformation. (B) Representation of homology template DNA containing PAM and E347A mutations. Green shows the PAM mutation.

4.4.3. Cloning of the gRNA

In order to generate a DSB at the desired location, a gRNA was designed using the flyCRISPR Target Finder tool. Since the glutamic acid at position 347 was going to be replaced with alanine the genomic sequence 100bp upstream and downstream of this nucleotide was used for screening as a target site to generate the double stranded break. A gRNA with no predicted off-targets that causes a putative DSB 17 bp upstream of position 347 was identified and selected for cloning (Figure 4.20).

The chosen spacer gRNA sequence was cloned into the expression vector pCFD5. pCFD5 contains a strong ubiquitous promoter, U6:3. Elements cloned under this promoter are expressed by RNA polymerase III; the expressed RNAs do not undergo post-transcriptional modifications like those expressed by RNA polymerase II. In the vector, the gRNA is flanked by tRNAs that are cleaved by the RNase P and RNase Z enzymes. These enzymes are expressed endogenously, thus it is possible to clone one or multiple gRNAs to this vector and all of them will be liberated by these enzymes.

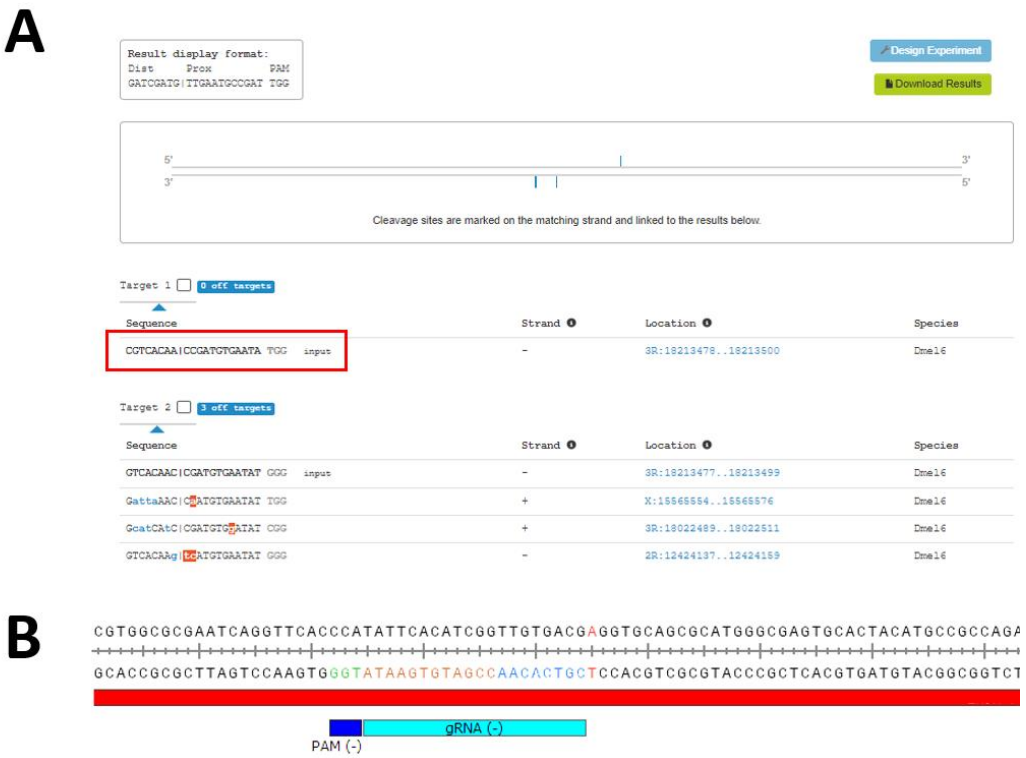

Figure 4.20 Selection of a gRNA to generate DSB near the catalytic site. (A) Top scored gRNA with no off-targets was chosen from DNA close to catalytic site. Red box shows the chosen sequence. (B) gRNA sequence is close to the point mutation target. Red: Point mutation target, Blue: Distal gRNA, Brown: Proximal gRNA Green: PAM site.

In order to clone the chosen spacer gRNA, the pCFD5 plasmid was digested with the type IIS restriction enzyme *Bbs*I. Sense and antisense oligos encoding the guide RNA and overhangs complementing the *Bbs*I restriction site were designed and commercially obtained. The oligonucleotides were phosphorylated using polynucleotide kinase (PNK), incubated at 95°C and annealed by ramping the temperature down to 20°C at a rate of 5°C/min. The phosphorylated and annealed oligonucleotides were then ligated to the *Bbs*I digested pCFD5 vector using T4 DNA ligase (Figure 4.21A). Plasmids from individual colonies were isolated and digested with *Kpn*I enzyme for analysis (Figure 4.21B). Positive colonies were sent to sequencing for verification. One of the verified plasmids was prepared in large scale for injection into fly embryos.

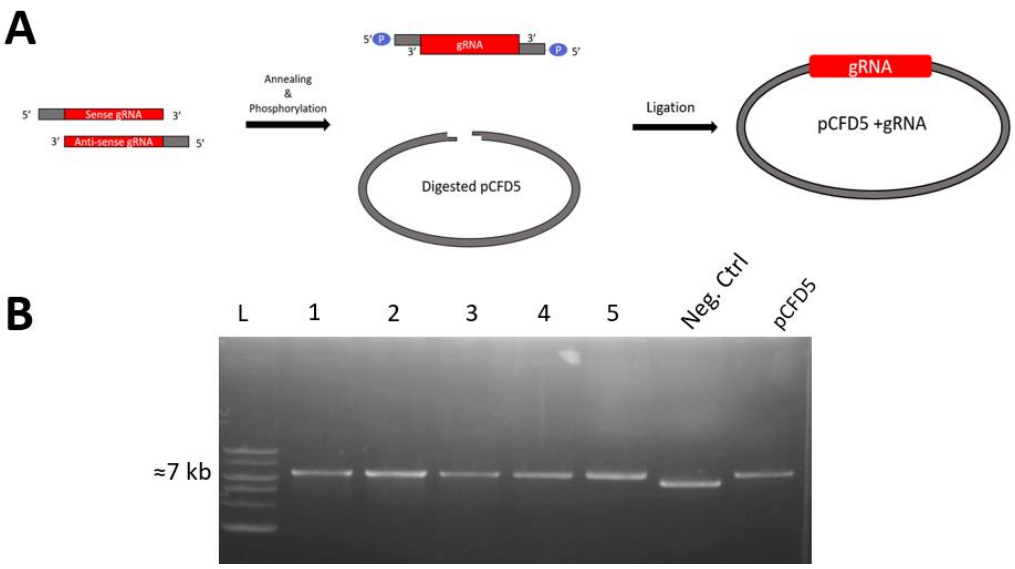

Figure 4.21 Cloning of gRNA DNA to pCFD5 vector. (A) Schematic overview of gRNA cloning strategy. (B) Analytical digestion of the selected colonies show that all of the samples were positive.

4.4.4. Injection and screening for the catalytic mutant

The homology template and the gRNA vector were injected into 480 preblastoderm Nos-Cas9 embryos. Individual G0 flies were crossed with balancer flies and the 3rd chromosome of the putative mutants were balanced with a balancer chromosome (Figure 4.22). 450 single crosses were established, however only about 180 of these crosses gave viable progeny.

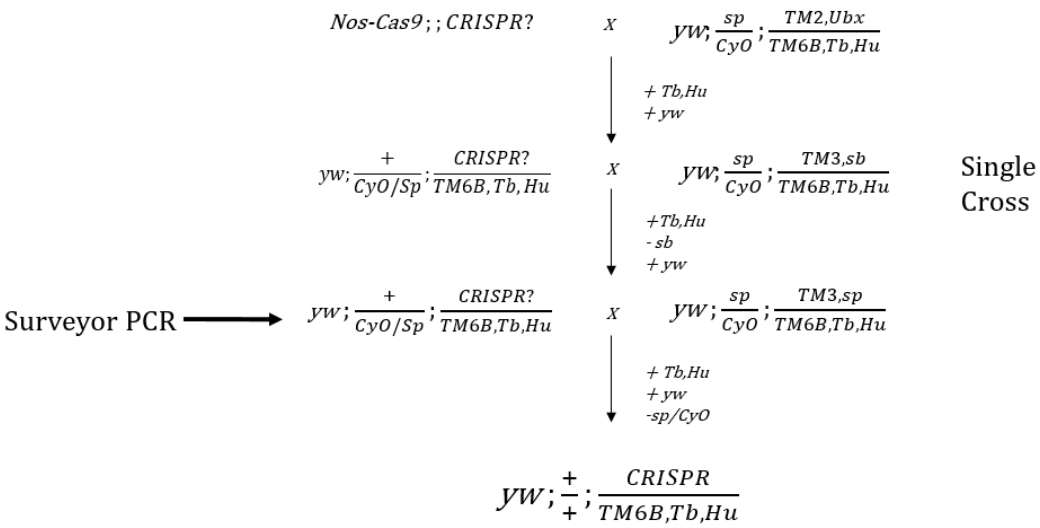

Figure 4.22. Crossing scheme after the hatching of the injected nos-Cas9 embryos. F1 generation is single crossed to isolate possible mutant allele at the 3rd chromosome and then balanced with TM6B for subsequent crosses.

To identify the flies carrying the point mutation, putative mutants were screened by surveyor PCR. Surveyor PCR is usually used to screen point mutations. One of the primers is designed to complement the intended mutation at its 3' end. It is expected that the band obtained from the PCR product of a mutant allele is much more intense than that of a wild type allele due to the fact that the 3' end is the most important part of the primer for the elongation reaction. Surveyor PCR primers were designed for the PAM site as well as the catalytic mutant site, since successful integration would result in the integration of point mutation at these sites. The PCR was optimized (See Appendix) to detect the PAM mutation and 152 putative mutant lines were screened (Figure 4.23B).

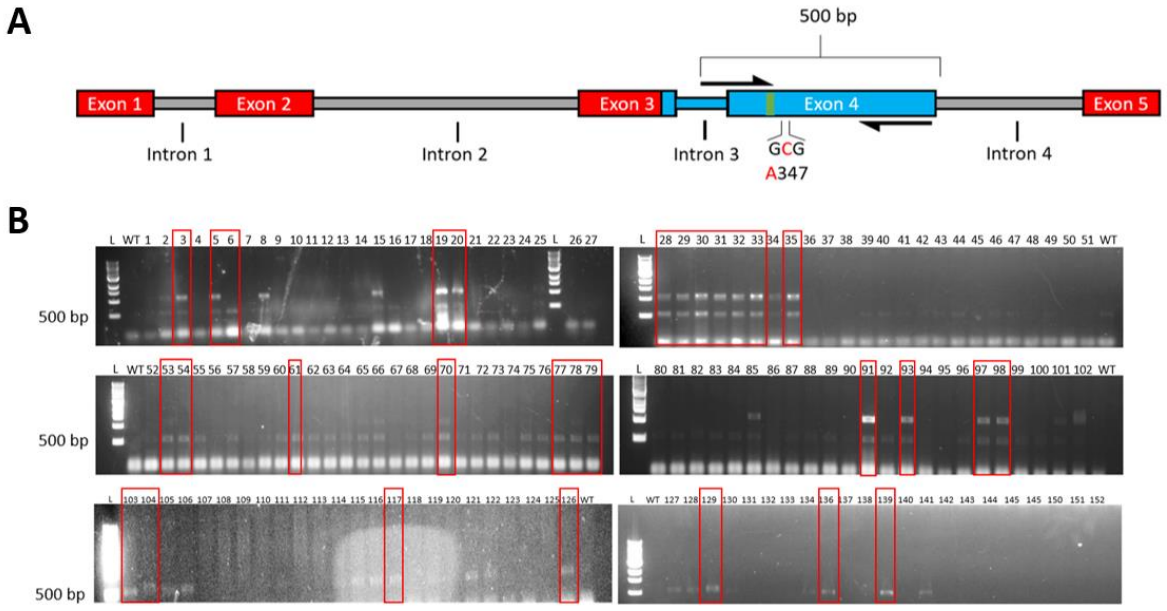

Figure 4.23 Surveyor PCR screen for the identification of PAM mutation. (A) Overview of the strategy. The forward primer is designed to complement the mutation of PAM site on its 3'end (Green) (B) Screening PCR was made and the most intense bands with 500 bp were selected for sequencing (red boxes). Upper bands around 1 kb are control reactions.

I collected the bands with higher intensity than the others and selected the lines that are homozygous lethal for the 3rd chromosome by selecting for the *tubby* marker on the TM6B chromosome. A total of 46 lines were selected and sequenced. 16 of these lines were homozygous lethal lines that did not give an intense band in surveyor PCR screening and unexpectedly no mutation was detected in these lines. More surprisingly, none of the homozygous lethal lines contained any mutation 3 nucleotides after the PAM site meaning that the lethality was not caused by disruption of *dmHexDC*.

4.5. Generation of a RMCE template for dmHexDC by using CRISPR/Cas9

4.5.1. Strategy to utilize RMCE for dmHexDC

Elucidation of the function of dmHexDC still requires experiments that show endogenous and subcellular localization, interaction partners, experiments and its *in vivo* substrates and more loss of function experiments. In this study, we aimed to generate a dmHexDC mutant that can be used to easily integrate any kind of sequence between two attachment sites that have been engineered to flank the dmHexDC gene. Thus, I designed

another CRISPR experiment to introduce site-specific recombinases into the genomic locus of *dmHexDC*, while knocking it out and replacing it with the *white* gene, which is responsible for the red eye color of flies and will serve as a marker for a successfully generated mutant line. It was shown that the most efficient site-specific recombinase was the attP-attB recombination for RMCE and thus I chose to integrate attP sites. The downside of attP-attB recombination is that the recombination event leaves a "scar DNA", consisting of attR and attB sites.

My strategy to generate the described mutant is to excise out the *dmHexDC* gene and integrate *white* flanked by inverse attP sites using CRISPR/Cas9. Once again, I planned to create DSB in *dmHexDC* locus and integrate DNA from a donor plasmid. To excise *dmHexDC* two gRNAs were chosen to guide Cas9 to perform its endonuclease activity in the upstream and downstream of the *dmHexDC* gene. Since the insertion of attP sites will result in upstream and a downstream attR scars I chose to integrate the attP sites in unconserved regions of the intergenic DNA between *dmHexDC* and its neighboring genes on the 3rd chromosome (Figure 4.24).

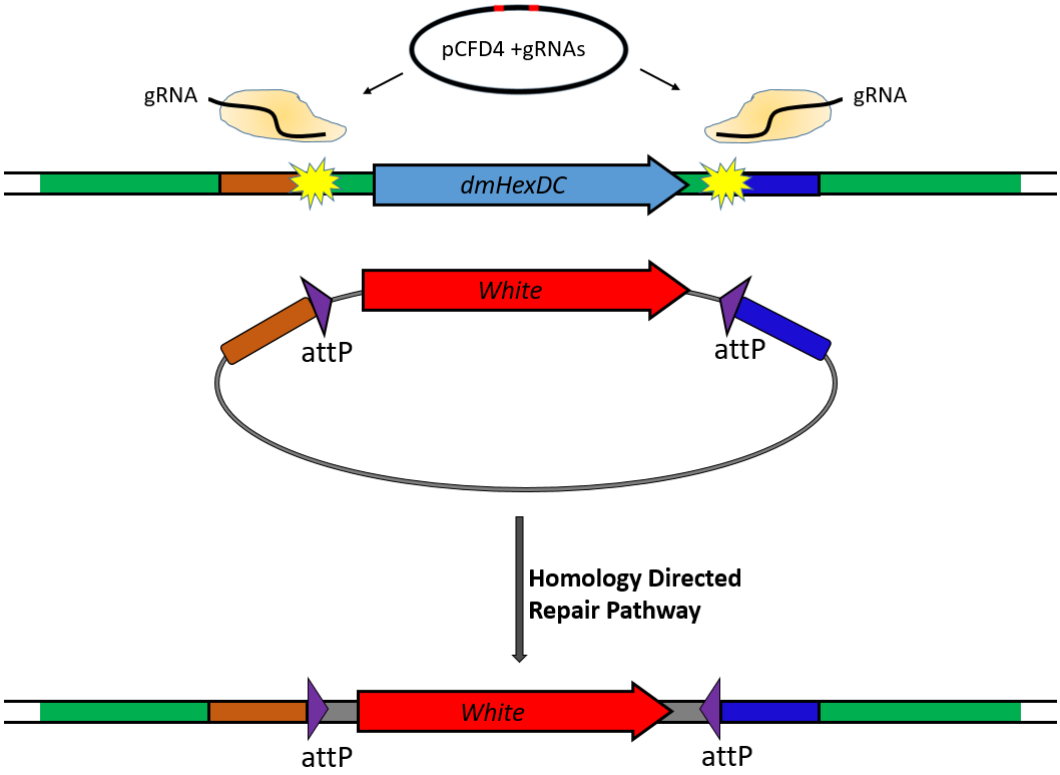

Figure 4.24 The schematic representation of knocking out *dmHexDC* while knocking in *white gene* flanked by attP sites by HDR. Two gRNAs expressed from pCFD4 vector guide the heterologously expressed Cas9 to upstream and downstream of *dmHexDC* leading to the excision of dmHexDC. HDR pathway integrates the region between

homology arms in donor DNA. Successful integration results in w+ phenotype that can be screened.

After the generation of the *dmHexDC*^{null} CRISPR mutant as a RMCE template mutant, the next step will be RMCE. The RMCE experiment will allow the exchange between any DNA of interest that is flanked by attB sites with the *white* gene (Figure 4.25). Since the product of the *white* is a transporter of red pigment. The resulting phenotype will be observed as red eyes in a white negative background. Thus, RMCE will result in replacement of the *white* gene with any DNA and thus the successful exchange will be monitored by loss of red eye color in the offspring.

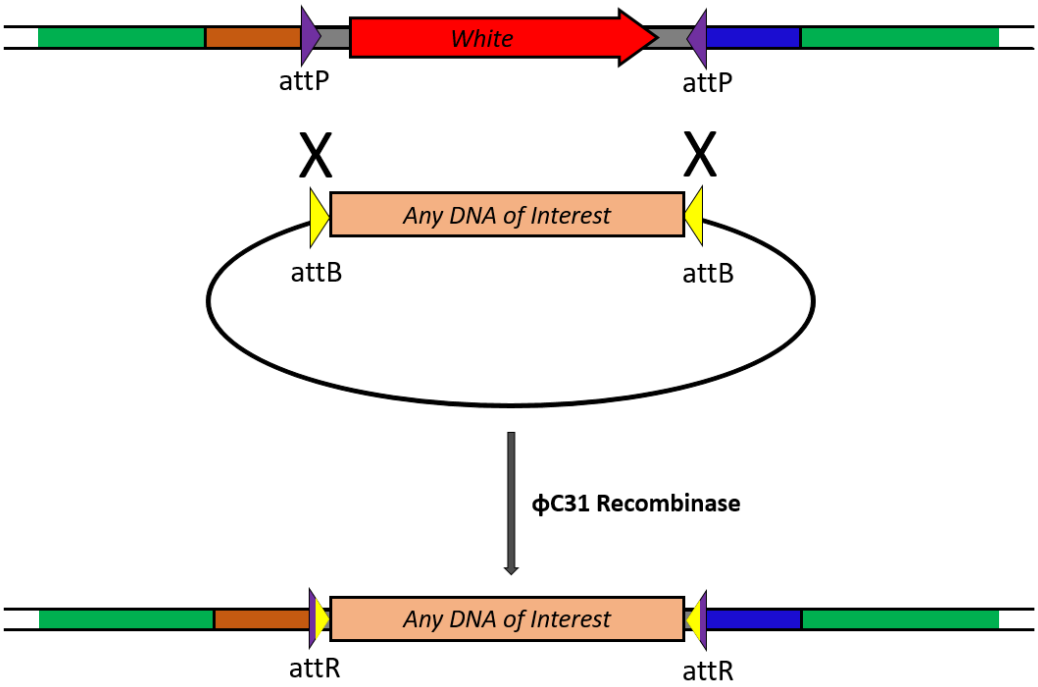

Figure 4.25 Schematic overview of RMCE. Irreversible non-random integration mediated by ϕ31 recombinase occurs between attP and attB sites. Result of this recombination generates hybrid sequences inverted attR and sites. During this recombination, the *white* gene between the inverse attP sites are excised out and any DNA of interest between attB sites are integrated.

Our objective was to utilize the powerful RMCE system for *dmHexDC*. The first step is to generate the CRISPR mutant. This mutant is also a reliable loss of function mutant for analysis of the functional role of dmHexDC in *Drosophila* development.

4.5.2. Donor DNA vector was constructed to generate a RMCE template mutant with CRISPR/Cas9

In order to generate this mutant, we were to choose gRNA sequences from unconserved intergenic regions then clone the spacer gRNA DNA into a gRNA expression vector and to clone the homology arms from intergenic regions to a donor DNA vector that contains the inverse attP sites flanking *white*. pWhiteStar was used as the donor DNA vector because it contains the inverse attP sites flanking *white* and it contains restriction sites near these elements. These restriction sites make it possible for us to clone the homology arms upstream of the first attP site and downstream of the second inverted attP site. In this way the flanked attP-white-attP construct flanked by homology arms can be integrated to the desired position in the genome via the HDR pathway. pWhiteStar has a size of approximately 7.2 kb and was digested with *Kpn*I enzyme, making the vector linear. Homology arms were amplified with an extra 21 bp sequence for 5' and 3' ends. A 21 bp sequence was selected to overlap the terminal 5' and 3' ends of the linearized pWhiteStar, except in the case that there is a 5'overhang in the linearized vector. We strategized sequential Gibson assembly for the cloning of the homology arms. 1 kb homology arm from the intergenic region upstream of *dmHexDC* was amplified from Nos-Cas9 flies. Linearized pWhiteStar and the amplified upstream homology arm were mixed and cloned in a scarless manner using Gibson assembly. The *Kpn*I restriction site is disrupted after Gibson assembly. Positive colonies were selected and plasmids were isolated. Analytical digestion was made with *Kpn*I and *Spe*I plasmids. *Kpn*I site was disrupted in the case of correct assembly and only *Spe*I linearized the correct construct which has a size of approximately 8.2 kb (Figure 4.26).

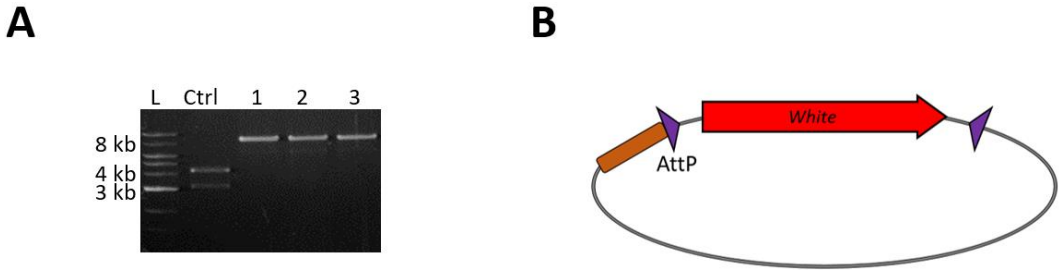

Figure 4.26 The analysis of the plasmids for upstream homology arm cloning to pWhiteStar vector. (A) SpeI and KpnI digested constructs was analyzed. Control (Ctrl) vector was digested into two bands with ≈3 kb and 4.2 kb. However positive colonies were ≈8.2 kb, implying correct assembly. (B) Schematic representation of donor DNA with upstream homology arm.

The 1 kb downstream homology arm was amplified with 21 bp sequence identical to pWhiteStar and was cloned into pWhiteStar+Homology Arm Up (pWS+HAup) plasmid by Gibson assembly. Ligated DNA was transformed to Dh5α competent *E. coli* cells. Plasmids that are isolated from single colonies were subjected to analytical restriction digestion with *Spe*I endonuclease. The plasmids with an increased size of digestion product were sequenced for validation.

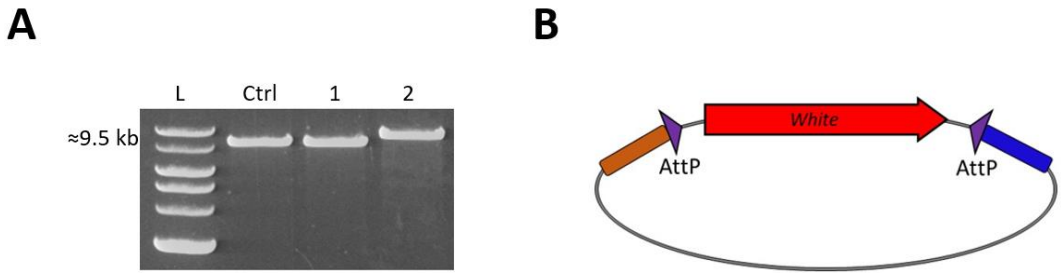

Figure 4.27 The analysis of the plasmids for downstream homology arm cloning to pWhiteStar vector. (A) Analytical digestion results showed the plasmid with increased size (B). Schematic representation of pWhiteStar + Homology Arm Up + Homology Arm Down.

pWhiteStar+Homology Arm Up+Homology Arm Down (pWS+HAup+HAdwn) contained the homology arms necessary for integration by HDR using CRISPR. The PAM motif was present for only one gRNA in the upstream homology arm sequence. To protect the donor DNA we introduced the point mutation to disrupt this PAM site by site-directed mutagenesis. 6 isolated colonies were subjected to analytical digestion. Interestingly, even though site-directed mutagenesis was successful large deletions occurred in all of the plasmids. However as the plasmids still contained the sequence of mutagenized PAM site I subcloned the intended mutation from one of the plasmids to the pWS+HAup+HAdwn vector. After transformation into *E. coli* the isolated colonies were selected and their plasmids were subjected to analytical digestion with *Pme*I and *Afl*II and sequencing (Figure 4.28). According to the results of the analysis plasmid #2 was chosen as a donor DNA for the CRISPR experiment.

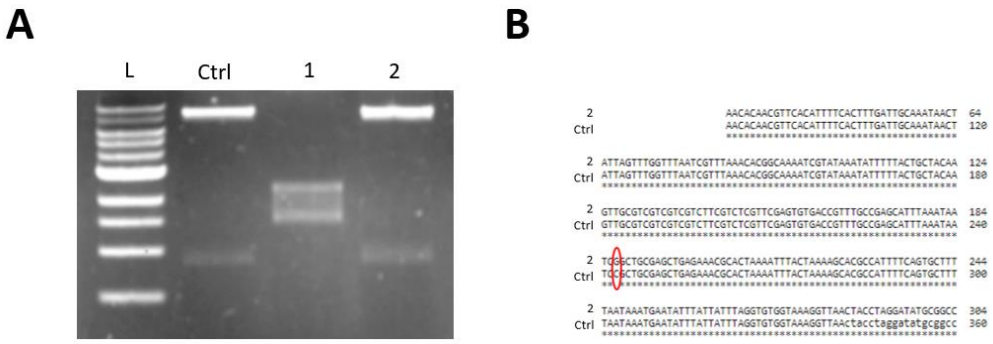

Figure 4.28 Cloning of PAM mutation containing DNA to pWS+HAup+HAdwn vector. (A) Analytical digestion with *Pme*I and *Afl*II. (B) Sequence alignment of Plasmid 2 and pWhiteStar + Homology Arm Up + Homology Arm Down pWS+HAup+HAdwn vector shows PAM mutation is present in the final construct (nucleotide in the red circle).

4.5.3. Generation of the gRNA vector for RMCE template

Two gRNAs which should target Cas9 to the upstream and downstream unconserved intergenic regions of *dmHexDC* were designed using FlyCRISPR Target Finder for the intergenic regions of *dmHexDC* (Figure 4.29).

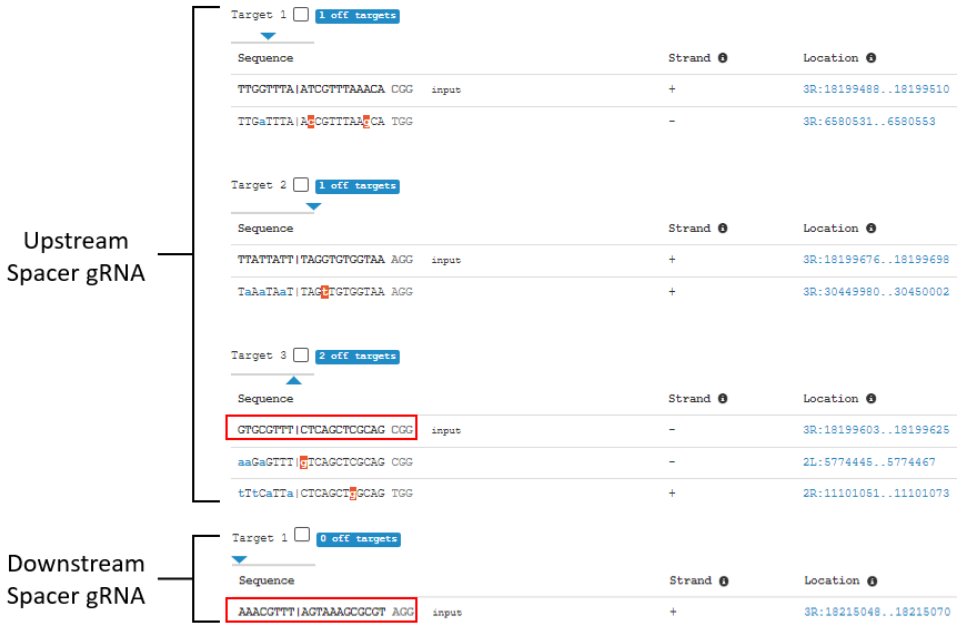

Figure 4.29. Spacer gRNAs were selected from upstream and downstream intergenic regions of *dmHexDC* by using Fly CRISPR Target Finder. 3 spacer gRNA sequences were evaluated from upstream intergenic region. Only the 3rd spacer gRNA contained no off-

targets on the 3rd chromosome, where *dmHexDC* is located. For downstream intergenic region, the top-scoring gRNA with no off-targets were selected.

We chose to implement site-directed mutagenesis approach to integrate spacer gRNA DNA to pCFD4. 63 bp forward and reverse primers that contained the chosen spacer gRNA DNA in their center were ordered. PCR was made to amplify 600 bp pCFD4 vector with spacer gRNA sequence. The product of this PCR was utilized as primers for site-directed mutagenesis to integrate the chosen spacer gRNA DNA to pCFD4 plasmid (Figure 4.30A). After site-directed mutagenesis the PCR products were digested by *Dpn*I and transformed into Dh5α competent *E. coli* cells. Plasmids were isolated from positive colonies. Successful integration would result in a *BstB*I site to be present at the junction of pCFD4 and spacer gRNA DNA. Analytical digestion was made on the isolated plasmids with *Kpn*I-HF and *BstB*I enzyme and consequently 2 bands below 400 bp were expected from mutant plasmids (Figure 4.30B).

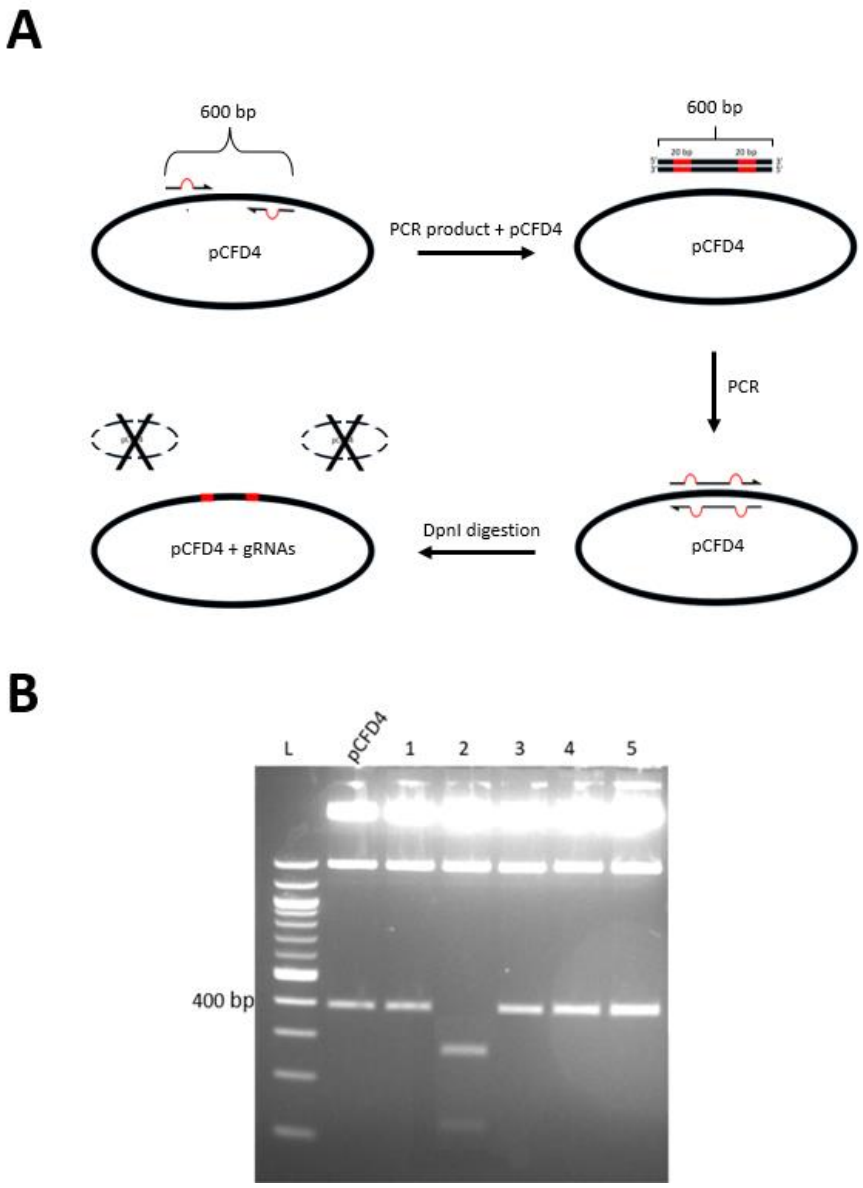

Figure 4.30 Analysis of gRNA integrated pCFD4 plasmids. (A) Schematic view of the strategy to integrate spacer gRNA sequences to pCFD4 vector. Red parts represent spacer gRNA DNA. (B) Isolated plasmids were digested by *Kpn*I-HF and *BstB*I enzymes. Plasmid with two bands below 400 bp was subjected to sequencing.

4.5.4. Injection and screening for mutant lines to generate a RMCE template

After the verification of gRNA sequences in pCFD4 + sgRNAs vector, the donor DNA and gRNA expression plasmids were isolated in large scale and prepared for injection. The constructs were injected to 480 Nos-Cas9 embryos at Genetivision Corporation. Males emerging from the injected embryos were selected and crossed to wild-type w^{1118} flies. Male F1 with red eyes were considered as putative mutant lines since they are the only ones that

can express the *white* gene and generate red eyes if the donor DNA vector was integrated into their genome. These flies were crossed until a stable stock was obtained (Figure 4.31A).

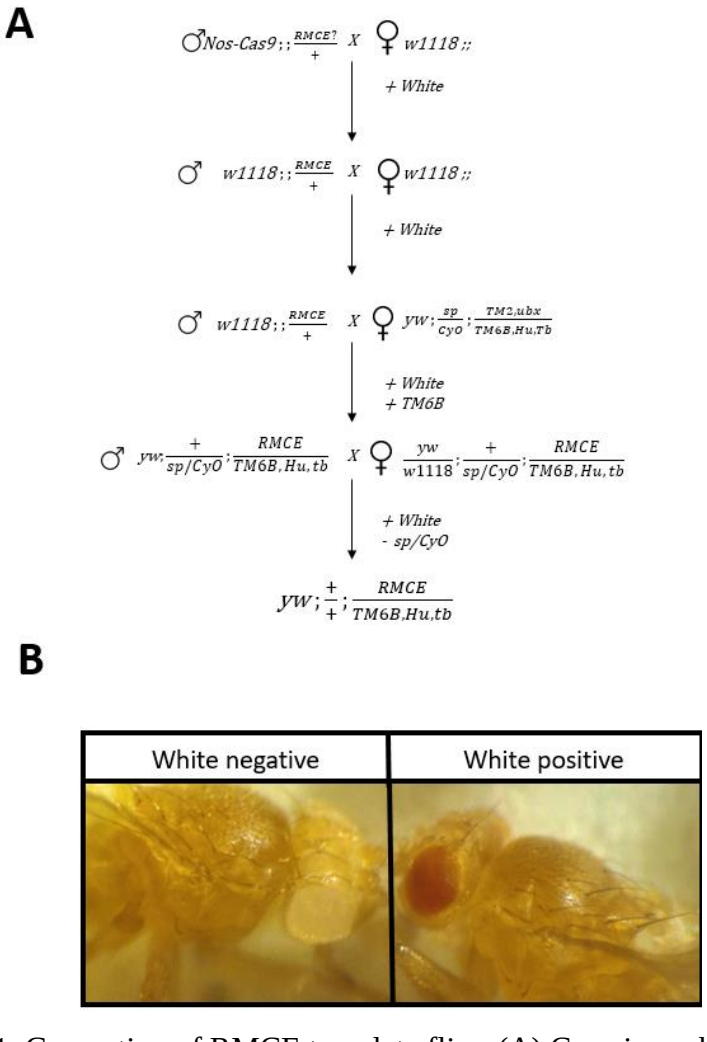

Figure 4.31. Generation of RMCE template flies. (A) Crossing scheme for RMCE template flies. (B) Screening for *white*-positive flies were made in *white*-negative background to identify the mutant lines.

3 transgenic lines were identified as *white* positive and named as A, B and C lines. Since *dmHexDC*^{null} mutants were homozygous lethal and we aimed to replace *dmHexDC* with attP flanked *white*, we also expected the mutants with correctly integrated *white* allele to be homozygous lethal. As anticipated A and B lines showed homozygous lethal phenotype, however line C was homozygous viable for the *white*-containing allele. The correctness of the mutation was confirmed by PCR. We designed two sets of primers annealing to *white* and to intergenic regions upstream and downstream of the homology arms. Genomic DNAs were isolated from the three mutant lines and a wild type line and

subjected to PCR analysis with above-mentioned primers. No PCR products were present for the WT line as the intergenic region of *dmHexDC* is at the 3rd chromosome and the endogenous *white* gene is at the X chromosome. For lines A and B the products corresponding to the regions of integration were successfully amplified verifying the event of mutagenesis (Figure 4.32). PCR analysis of line C resulted in amplification of the PCR product of expected length only with one pair of primers suggesting that the integration of donor DNA occurred at the downstream intergenic sites and probably did not disrupt *dmHexDC.*

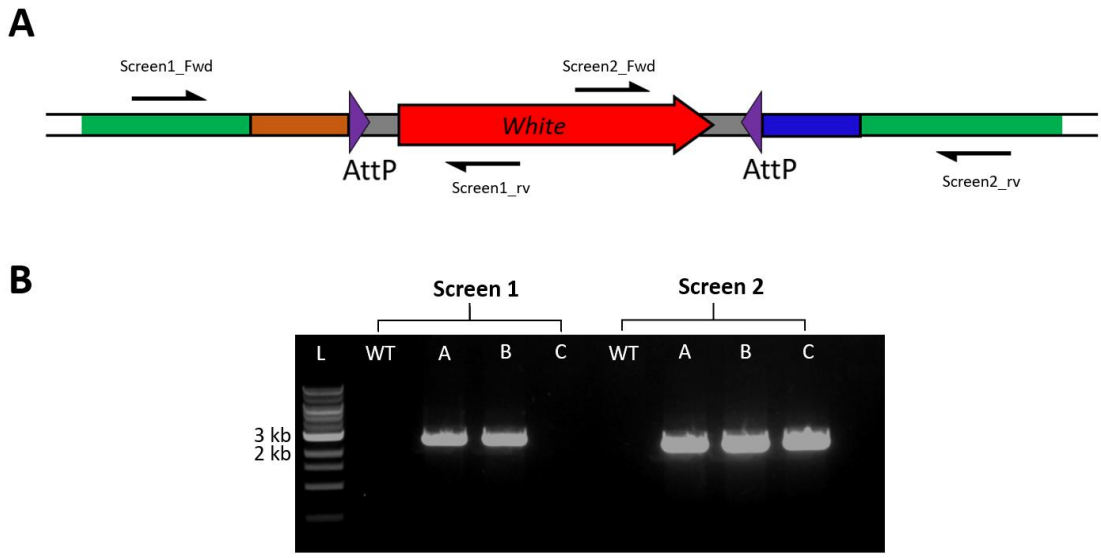

Figure 4.32. PCR analysis of RMCE template lines. (A) Overview of the screening PCRs. (B) The results of PCR. Both PCR resulted in no bands in WT as predicted and lines A and B gave the expected 2614 and 2322 bp bands for screen1 and screen 2 PCRs respectively. Only screen 2 PCR worked for line C.

4.6. Analysis of RMCE mutants

4.6.1. Flies that survive without *dmHexDC* in either allele were detected

Interestingly, we have detected flies that were homozygous viable for the RMCE Template mutant allele during the course of this study. This suggests that there may be other mutations present in the genome that impairs the viability of A and B lines and these mutations may have recombined during subsequent crosses. These were isolated and tested for the presence of *dmHexDC* loci via PCR with primers selected from inside the *dmHexDC* CDS (Figure 4.33).

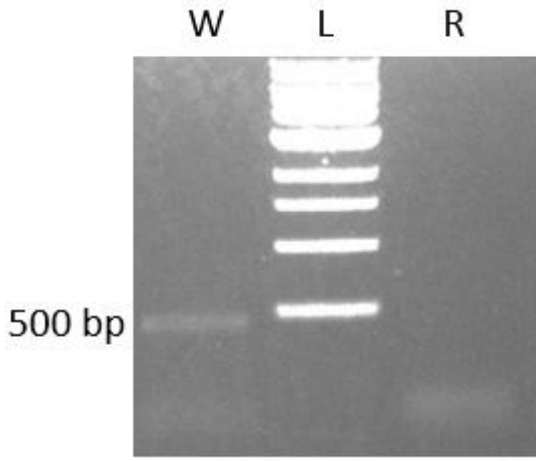

Figure 4.33. PCR analysis of mutant and wild-type lines. PCR results show that dmHexDC CDS is present in wild-type (W) but not in the homozygous RMCE template mutant (R).

PCR results showed that some flies somehow manage to survive without the presence of dmHexDC unlike their siblings. We are interested in the consequences the deletion of dmHexDC in photoreceptor cells in *Drosophila* development. Even though these flies seemed to adapt to the loss of *dmHexDC* and formed what seems to be normal adult eyes, we investigated the possible neuronal impairments in eye development as some phenotypes cannot be understood by looking at the external morphology.

4.6.2. Deletion of *dmHexDC* causes mild neuronal loss and disruption of the morphogenetic furrow

In order to understand the fate of photoreceptor cells located in the posterior of the eye imaginal disc, we performed an immunohistochemistry experiment labelling all neurons of wandering 3rd instar larval eye with Elav antibody and used Canton S flies as the wild-type control (Figure 4.34).

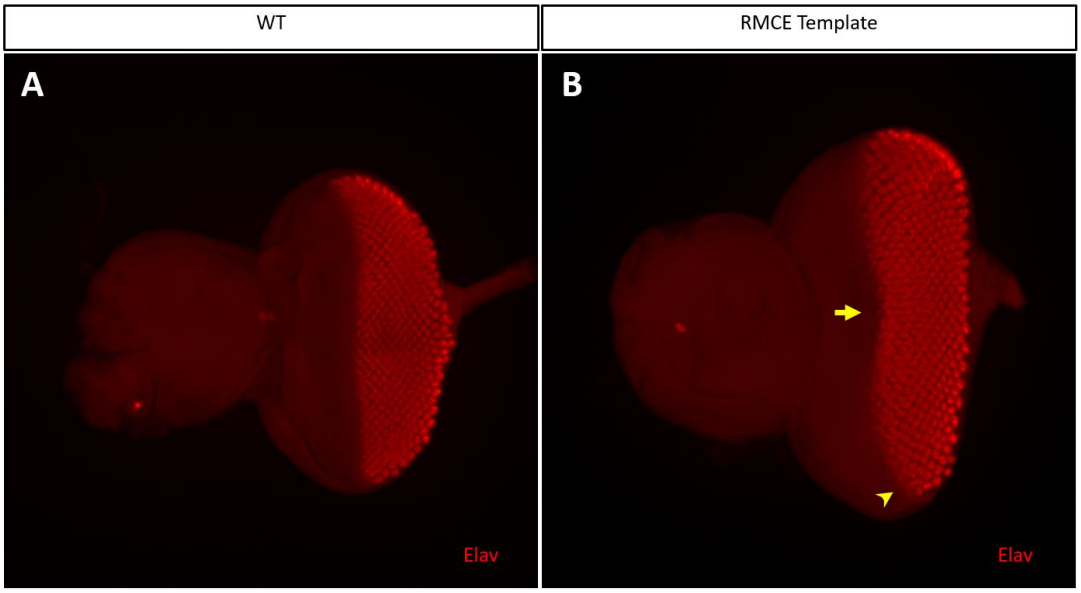

Figure 4.34. Comparison of photoreceptor differentiation between WT and dmHexDC deleted 3rd instar larvae eye imaginal discs. (A) WT neuronal pattern and morphogenetic furrow formation. (B) dmHexDC deleted larvae shows very mild to no neuronal loss at the dorsoventral midline (arrow) and the ventral side of morphogenetic furrow (arrowhead).

Labelling of the neurons with Elav in RMCE template flies showed that there is a loss of neurons at the ventral side of morphogenetic furrow of the eye imaginal disc. This suggests an impairment in differentiation of preneural cells fated to become photoreceptor neurons and supports the previous loss of function analysis in a much milder manner.

5. DISCUSSION

5.1. GH20 Hexosaminidases

Hexosaminidases are conserved in the 3 domains of life. In eukaryotes, glycoside hydrolysis activity of hexosaminidases are utilized in various ways. Hexosaminidases take part in metabolic activities, maturation and turnover of glycoconjugates. Research on hexosaminidase activity has mostly been influenced by the pathologies arising from HexA and HexB proteins. Mutations in both protein cause LSD by accumulation of GM2 gangliosides and for some time research on hexosaminidases were dominated by the investigation of hexosaminidases in GM2 gangliosidosis. In 1995, hexosaminidase activity was detected in *Drosophila* protein extracts. It was hypothesized that this hexosaminidase was responsible for simpler, paucimannose structure of the *Drosophila* glycome (Altmann et al., 1995). Later this hexosaminidase was identified as Fdl and this hypothesis was confirmed. Substrate specificity of Fdl was determined as GlcNAc and subcellular localization of fdl was determined to be the Golgi. Experiments with various substrates showed that Fdl has a role in N-glycan trimming and loss of Fdl function caused fusion of mushroom body β lobes in the CNS, indicating a role for hexosaminidases in neurodevelopment. For the first time, an invertebrate hexosaminidase was characterized to this extent. Later other hexosaminidases of *Drosophila melanogaster* were characterized. Recent work has shown that Hexo1 and Fdl take part in Rhodopsin 1 maturation proving further evidence that hexosaminidases are required for the proper function of neuronal proteins. The catabolism of chitin was attributed to Hexo1 and Hexo2. All *Drosophila* hexosaminidases identified so far are members of the GH20 family. Elucidation of the glycosylation elements such as hexosaminidases will allow us to further understand how the glycome of *Drosophila* and subsequently more complex eukaryotes, is regulated.

5.2. dmHexDC is Heterologously Expressed

In order to characterize dmHexDC, I cloned two forms of its CDS, full-length and transmembrane domain truncated, to pETM-20, a bacterial protein expression vector. The pETM-20 vector contains several N-terminal peptide tags, TrxA, 6xHis, and a Tobacco Etch Virus (TEV) protease target site. Furthermore, the pETM-20 vector contains a *lacI* gene. The product of *LacI* binds strongly to DNA and inhibits the expression of genes downstream

of the LacO region. However, lacI has a stronger affinity to allolactose. When allolactose is present, LacI will leave the DNA and bind to it.

In JM109 *E.coli* cells, expression of T7 RNA polymerase is regulated by lacI, also in the pETM-20 vector, there is a lacO region downstream of the strong T7 promoter and upstream of the cloned *dmHexDC* CDS. LacI is used to silence the transcription of dmHexDC until the *E. coli* culture reaches the log phase in which the production is most efficient. IPTG is a mimic of allolactose that cannot be metabolized by *E. coli* so that the concentration of IPTG is not altered during bacterial growth. Therefore when IPTG is introduced, T7 RNA polymerase can be expressed and transcribes *dmHexDC*.

Protein expression of the full-length dmHexDC is hindered, most probably because of its hydrophobic transmembrane domain. Strong expression of hydrophobic proteins can form insoluble aggregates due to interaction of their hydrophobic stretches and inhibit the growth of the organism. Fortunately, the expression of truncated dmHexDC did not exhibit such a problem and could successfully be expressed in JM109 *E. coli* together with the aforementioned protein tags. After the expression of truncated dmHexDC, in collaboration with Neş'e Bilgin, I tried to purify the truncated dmHexDC using IMAC. After purification, the N-terminus tags were going to be removed by TEV protease from the TEV protease target site. Interestingly however, we could not elute the samples from the purification column, which may be due to precipitation of the protein in columns. The proteins can of course be eluted by using harsh methods like detergents and urea, but these methods are destructive to the 3D structure of the protein. Since we wanted to assess the activity and substrate specificity of dmHexDC, these methods were not used to purify the protein.

5.3. Substrate Specificity of dmHexDC is Towards GalNAc

In order to elucidate the biochemical function of dmHexDC, I used the crude bacterial extracts for the enzyme test as described in Gutternigg, 2010. A downside to this approach is that we do not know if the N-terminus tags affect the activity or the thermostability of the protein. Also due to the N-terminus tags, the theoretical isoelectric point of dmHexDC shifts towards a more acidic value. I have decided to test the β-hexosaminidase activity of dmHexDC against GalNAc residues due to its sequence homology with human HEXDC therefore dmHexDC was tested against *pnp*-β-D-GalNAc between pH3.0-8.0. In all tested pH values, except for pH3.0-4.0, high levels of β-hexosaminidase activity was detected in comparison to the bacterial lysate that does not express truncated dmHexDC. This indicates

that dmHexDC is not active in extremely acidic conditions. The thermostability of the enzyme was also tested by preincubating it at 25°C, 37°C, 55°C, 70°C, and 90°C. As a result dmHexDC did not display any extreme thermostability features as it lost its activity at temperatures higher than 55°C and no change of activity was observed between 25°C and 37°C (see Appendix).

Even though the N terminus tags may shift the optimum pH slightly, the measurements of the crude extracts gave valuable information in terms of its possible *in vivo* activity. I observed that dmHexDC can be active in every cellular compartment, including lysosome (but with less efficiency). The pH optimum of 6.0 for GalNAc contradicts the hypothesis that dmHexDC is located in the lysosomes and supports the hypothesis that it is localized to the Golgi and/or endosomes. Furthermore, in comparison to FDL (Dragosits et al., 2015), dmHexDC seems to favor a more basic environment.

Even though I showed dmHexDC is a hexosaminidase that cleaves terminal βGalNAc residues, previous data on HEXDC and invertebrate hexosaminidases, *Drosophila* fdl, *Caenorhabditis elegans* HEX-2, HEX-3, HEX4 and *A. mellifera* fdl showed that there is the possibility that the substrate of dmHexDC is not limited to β-GalNAc residues. The mentioned hexosaminidases all cleave two substrates but not with the same efficiency. Human HEXDC and *Caenorhabditis* HEX-2, HEX-3, HEX4 enzyme prefer β-GalNAc residues but can also hydrolyze terminal β-GlcNAc with less efficiency. Fdl enzymes on the other hand catalyzes the hydrolysis of β-GlcNAc residues more efficiently than the β-GalNAc residues. In light of this information, we have obtained *pnp*-β-D-GlcNAc and I showed that dmHexDC also hydrolyses βGlcNAc residues between pH4.5 and 8.0 with less efficiency compared to β-GalNAc. Therefore, I concluded that dmHexDC cleaves both β-GalNAc and β-GlcNAc residues but its affinity is higher towards β-GalNAc. Thus, we have identified the first enzyme in *Drosophila melanogaster* that has high substrate specificity towards β-GalNAc. Comparison of pH optimums between β-GalNAc and β-GlcNAc revealed that dmHexDC cleavage of β-GlcNAc favors a more acidic environment. This trait is similar to fdl. This work also opens up the possibility that physiological substrates of dmHexDC can contain both hexosamines, meaning that dmHexDC can probably also contribute to the paucimannose dominated glycome of the *Drosophila melanogaster*.

5.4. Structure of dmHexDC is $(\beta/\alpha)_8$ Barrel

GH20 hexosaminidases have a conserved H/N-X-G-A/C/G/M-D-E-A/I/L/V motif which is also present in dmHexDC. It is well known that the hexosaminidase activity is established by an aspartic and glutamic acid dyad in this motif. Aspartic acid is theorized to electrostatically stabilize the substrate and glutamic acid catalyzes the hydrolysis by acting as a general acid/base residue. Previously, the structure of the GH20 domain enzymes have been investigated and it is now known that the majority of GH20 hexosaminidases fold into a $(\beta/\alpha)_8$-barrel topology. The β barrel is hidden inside the protein and surrounded by α helixes. Furthermore, it was shown that the present alpha helices at the N terminus of the $(\beta/\alpha)_8$-barrel topology is also required for the proper function of the enzyme.

In order to verify dmHexDC is a GH20 hexosaminidase with similar traits, I generated a homology model of dmHexDC by using MODELLER and I-TASSER. Due to unconserved amino acids at the N terminus and C terminus of the putative catalytic domain, a full model could not be constructed. The output in this case was only the catalytic domain of dmHexDC. I did not encounter the same problem with the I-TASSER server, but the transmembrane region that was not modelled was represented as a random coil in the output models. In both MODELLER and I-TASSER, bacterial hexosaminidases were used as templates due to the lack of the 3D structure data from invertebrate hexosaminidases.

The generated structure from I-TASSER resembled a $(\beta/\alpha)_8$-barrel topology, which contains eight β strands organized like a tunnel in the center instead of eight α helixes which resembles a doughnut. The dmHexDC model contained seven β strands and one coil in the catalytic pocket and eight alpha helixes surrounding a β tunnel. Considering that the homology model output did not contain high confidence scores, there was the possibility that some but not all attributes were assigned during homology modeling. Nevertheless, I used this homology model to further verify the location of important amino acid residues that interact with the substrate.

5.5. E347 is the General Acid/Base residue

I structurally superimposed dmHexDC on a bacterial hexosaminidase in complex with a hexosaminidase competitive inhibitor, NAG-thiazoline. I identified the 346^{th} aspartic acid and 347^{th} glutamic acid as catalytic residues as they contained the closest alpha carbons to NAG-thiazoline, suggesting the most direct interaction.

I continued to mutate the general acid/base residue, glutamic acid to validate that this is indeed the residue of importance for the catalytic activity of dmHexDC. To diminish the activity of dmHexDC, I mutated the E347 residue to alanine. Alanine is usually chosen for such mutations because it's secondary structure is deemed small enough and it does not interfere with the folding of the protein by electrostatic or steric effects (Lefèvre et al., 1997). Furthermore, the impact of alanine mutations on human hexosaminidase D was previously investigated in Alteen et al., 2016. The E347A mutant exhibited diminished catalytic activity towards β-GalNAc and β-GlcNAc. This validated the hypothesis that E347 is the general acid/base residue.

Previous studies made on the characterization of dmHexDC showed that dmHexDC was homozygous lethal in the 3rd instar larval stage. Interestingly, we observed homozygous lethality on the embryonic stages much more than 3rd instar larval stage. The fly line that was used had a blockage of transcription in *dmHexDC* (P.T). For this reason we decided to generate a dmHexDC mutant with diminished catalytic activity in order to increase the viability rate and allow us to observe developmental embryo defects. In Alteen et al., 2016, it was shown that mutation of the aspartic acid residue in the catalytic dyad diminishes the activity more than the disruption of the general acid/base residue. We considered that this might result in a *dmHexDC^{null}* phenotype more than a hypomorph phenotype.

5.6. Obstacles in Generation of Hypomorph

Since I aimed to investigate the effects of E347A mutation *in vivo*, I decided to generate the hypomorph by using CRISPR/Cas9 gene editing. To generate a specific mutant, I generated the tools required to induce the HDR pathway, meaning that I provided a homology DNA template and a gRNA to Cas9 expressing *Drosophila*. We selected the gRNA using flyCRISPR Target Finder tool and cloned the 20 nt sequence to a gRNA expression vector under the U6:3 RNA polymerase III promoter. The gRNA was designed in a way that DSB occurs 17bp away from the nucleotide that I intended to mutate. I cloned the donor DNA with 2 kb homology and mutated the catalytic site of the donor DNA. I also subjected the PAM site to a silent mutation to hinder the endonuclease activity of Cas9 against it. The flies that express Cas9 under the germ-line specific *nos* elements were injected with the generated constructs. I collected the flies and made a total of 450 crosses to be balanced with a potential generated CRISPR mutant. Two third of the flies died in the first crosses. To identify the CRISPR mutant in the rest of the 150 flies, I did a surveyor PCR to

analyze the presence of the catalytic mutation in the *Drosophila* genome. Before the surveyor PCR, I optimized the parameters for the PCR in a way that WT DNA would be amplified less as compared to the mutant by using mutant and WT donor DNA plasmids. These experiments gave the best results with primers prepared against the PAM site. In the same tubes, a control for PCR was also made. Since the 3' nucleotide in the primer is critical for the efficiency of the PCR, I selected the most intense bands and hypothesized that the fly lines containing these bands were CRISPR mutants. I have also collected lines that are homozygous lethal for the 3rd chromosome and hypothesized that in these lines, NHEJ occurred instead of HDR and possibly disrupted the product of *dmHexDC* by frameshift mutations.

In contrast to my expectation, neither the lines detected for dmHexDC nor the homozygous lines did not display any mutation in *dmHexDC*. It is possible that the optimization settings for the primer detecting the mutation were not the same when there is a mutant genome or mutant plasmid but most importantly the presence of homozygous lethal flies indicated that there is probably an unpredicted off-target of the selected gRNA. This shows that even tough *in silico* gRNA off target predictions are thought to be effective, there are instances that they cannot predict *bona fide* off targets (X. H. Zhang et al., 2015) and off-target effects in *Drosophila* has been reported before (Ren et al., 2014).

5.7. Utilization of dmHexDC Antibody did not Produce Consistent Results

The previous studies to identify the subcellular localization of dmHexDC relied on a BAC line, however the BAC line was not sufficient to rescue of the P.T. line. This suggests that either the BAC line is missing some regulatory elements, or the C-terminus eGFP tag results in overexpression of misfolded dmHexDC that cannot catalyze the hydrolysis of hexosamine residues.

Our group has generated a polyclonal peptide antibody against dmHexDC. Previous attempts to use this antibody resulted in poor stainings in the eye imaginal discs (Kıral, 2015). To study dmHexDC in neuronal tissues of the fly, I attempted to optimize the antibody with Western blot, but I could only obtain a band at ~180 kDa and showed that this band was an unspecific band with peptide blocking. Interestingly, the antibody specifically recognizes dmHexDC expressed heterologously in bacteria. So, I concluded that the antibody recognizes dmHexDC but it also recognizes unspecific proteins even when blocked suggesting that the antibody is not pure enough. Taking into account that low expression or

poor isolation of the transmembrane containing dmHexDC may also be one of the causes I did not observe a dmHexDC specific band in Western blot experiments. Nevertheless, I continued my experiments with IHC of eye imaginal disc and CNS or larvae. I preabsorbed the antibody with homozygous protein trap larval tissues to decrease the background but the results were similar in both tissues. I observed very poor signals from dmHexDC in eye imaginal discs. In the CNS however I observed strong and weak signals. The weak signals were not observed consistently in most experiments. A number of methods were tried for increasing the activity of the antibodies including fixation with methanol and acetone, various incubation times, buffers and temperature but no consistent result was obtained. To test the specificity of strong signals, I used stained homozygote P.T. CNS and I observed the signals in mutant tissue (see Appendix). I then stained late stage embryos of dmHexDC and observed similar signal pattern with BAC line. Yet in the end the studies we made to optimize the dmHexDC antibody for *in vivo* studies remained inconsistent (see Appendix).

The results I obtained from these experiments suggested that we needed a powerful tool to facilitate our *in vivo* investigations of dmHexDC function for future experiments by applying the RMCE system to *dmHexDC*.

5.8. *dmHexDC* is Excised and Replaced with *attp-White-attp*

Applicability of CRISPR/Cas9 in every region of the *Drosophila* genome with efficiency is in question. Furthermore, PCR based screening of putative mutant demands high mutation rates and efficient gRNA. Especially in the case of point mutation where identification of the mutant line can be a difficult task as I have experienced while attempting to isolate E347A mutant. There is a way to overcome these setbacks. The screening is much easier with the integration of a visible marker and editing the same region of the genome would also be much more efficient if site specific recombination sites are also integrated to the genome.

We have decided to generate a line that can easily be manipulated to answer every question in the context of our reverse genetics based approach on characterization of dmHexDC function *in vivo*. By this rationale I utilized the CRISPR/Cas9 technique to generate a RMCE template. Before generating this template I have carefully studied the intergenic region of dmHexDC due to the fact that non-functional inverted attR scars would be present in the genome after the cassette exchange, therefore I chose the most unconserved region. I chose the upstream and downstream intergenic regions as gRNA target sites

because integration of attP sites to the introns upstream and downstream of the protein coding exons may be problematic. Thus we excised out the entire *dmHexDC* genomic region and some intergenic regions (15.451 bp). This also means that we have excised *CG44158* gene along with *dmHexDC.* The function of *CG44158* gene is unknown. I could not find any implication on the function of its product based on sequence homology. Therefore we do not know how the knock-out of *CG44158* would manifest itself as phenotype during the analysis of the RMCE template lines in loss of function experiments. Restoring *CG44158* along with intron 2 with cassette exchange would be more reliable compared to the RMCE template flies to study the loss of function of *dmHexDC*.

I have generated the RMCE line by constructing a donor DNA and a gRNA expression plasmid. We chose pWhiteStar to be the template of the donor DNA. The main advantage to this plasmid is that it contains a *white* gene which will serve as a reporter after integration to the genome and the *white* gene is flanked by the attP sites which will serve as target for the most efficient site-specific recombinase, φC31. I have cloned the homology arms upstream and downstream of inverted attP sites in pWhiteStar vector in order to integrate it to the desired position. I encountered difficulties during site-directed mutagenesis PCR for the PAM mutation, therefore, I solved the problem by subcloning the mutation containing DNA to the donor DNA vector. Then, I generated the gRNA expression construct.

Initially, I strategized to clone the two designed spacer gRNA DNA to pCFD5 vector with Gibson assembly. The *Bbs*I enzyme was necessary to linearize pCFD5 and pCFD4 plasmids however due to problems with *Bbs*I enzyme at that time, I changed our strategy to integrate the spacer gRNA sequence with site directed mutagenesis approach. We chose to implement this approach to pCFD4 vector due to pCFD5 plasmid having repeats of tRNA sequences that would make the site directed mutagenesis PCR difficult. The main difference between these vectors is only two gRNAs can be expressed from pCFD4 and one gRNA is expressed under a less efficient RNA polymerase III promoter, U6:1 instead of U6:3. Thus, without using cloning, I have managed to integrate two spacer gRNA region to a gRNA expression plasmid by using the same primers described in pCFD4 online protocol. The method described in this study can be applied by anyone who wishes to integrate their gRNAs to pCFD4 vector without Gibson assembly. The advantage is that this method is cheaper but more colonies should be analyzed in order obtain the desired gRNA integrated plasmid.

5.9. Cassette Exchange Approach to Generate *dmHexDC* Mutants

We have generated a fly line that contains *attp-white-attp* instead of the *dmHexDC* gene region. These lines can be used for loss of function analysis and most importantly, flexible and efficient editing of the *dmHexDC* genomic region. After the generation of this line, *dmHexDC* can be edited in the most efficient way. Excised *dmHexDC* gene region in the RMCE template can be edited with any DNA of interest that is found between two attB sites. But most importantly the RMCE template should be rescued with the *dmHexDC* gene region in order to prove that the attR scars present in the genome do not cause any phenotype (Figure 5.5.1). Integration of tagged *dmHexDC* can show subcellular localization. To answer the question of interaction partners of dmHexDC, tagging dmHexDC for identification of its interaction partners, with engineered proteins such as APEX2 can be utilized. Effects of amino acid substitutions on *dmHexDC* activity *in vivo* can be studied and *dmHexDC* can be replaced with other GH20 hexosaminidases to study their homology.

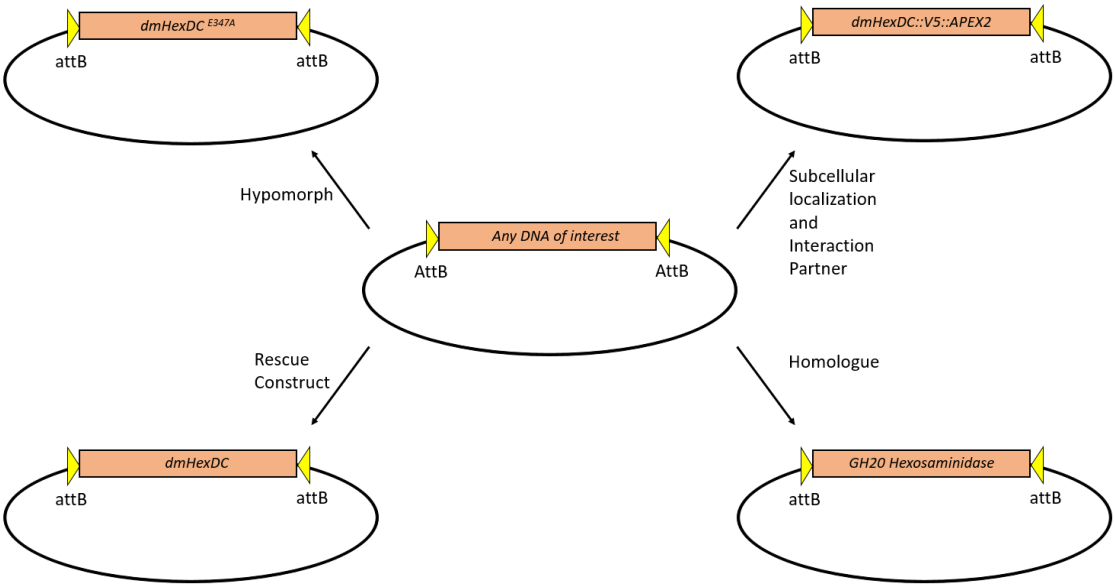

Figure 5.1. Scheme of possible constructs that can be integrated into the *dmHexDC* gene region. Virtually integration of any DNA to *dmHexDC* genomic region is possible.

5.10. Loss of *dmHexDC* Allele Showed a Very Mild Phenotype

Unexpectedly, we obtained both homozygous lethal and viable *Drosophila* lines for RMCE template mutants. We are not exactly confident on the reason behind this phenomenon. One explanation may be the gRNA off-target effect. Even though the integration of *attP-white-attP* was a success, other regions of the genome might be subjected to unspecific endonuclease activity by Cas9, which caused the mentioned homozygous

lethality. During crosses the unseen mutation can be recombined with the wild type DNA thus the damage for the *dmHexDC*-deficient allele is repaired. As a result, the lethality of the allele is lost. This also contradicts our previous hypothesis on the lethality of *dmHexDC* deficiency. I investigated to see the effects of dmHexDC deficiency in photoreceptor formation at 3rd instar eye imaginal disc and my results showed a much milder phenotype then the previous analysis made with the P.T. line that showed neuronal loss in a more extreme manner (Kıral, 2015).

The P.T. line is generated by insertion of YFP by P-element 1kb downstream of the 5'UTR of *dmHexDC*, which impaired the transcription of *dmHexDC*. Thus, the P.T. line was used in loss of function analysis. Although it is intriguing to see a much milder neuronal loss in RMCE Template flies, it is important to note that these flies somehow managed to overcome the loss of *dmHexDC* and thus expected to show less severe phenotypes. In another note, the true nature of P.T. line. It is not quite known as there can be other disruptions in the genome. In addition, it is possible that accumulation of YFP products may elevate the phenotype that was observed from P.T line significantly.

Unfortunately, time constraints in this study did not allow for the investigation of the difference between the RMCE template and P.T. lines.

5.11. Role of dmHexDC in R7 Photoreceptors

A question that has not been fully addressed in this study is the reason behind the presence of dmHexDC specifically in R7 photoreceptors in eye imaginal disc. It is now obvious that dmHexDC has a role in the glycosylation pathway, either in glycan trimming in Golgi and/or glycan metabolism in the lysosome. Consequently, the activity of dmHexDC may result in post translational modification of a transcription factor required in downstream regulation of R7 specific genes since all glycoproteins have either GalNAc or GlcNAc, which are both substrates for dmHexDC. Theoretically identification of upstream elements of dmHexDC can also be very insightful regarding the pathway dmHexDC partakes in.

The AC887 enhancer trap line contains a *Gal4* in the first intron of dmHexDC. Exon 1 and exon 2 of *dmHexDC* contains two putative promoter sites in their upstream. This is in accordance with the information annotated by FlyBase, which specifies that dmHexDC has two transcripts. One includes exon 1 and the other exon 2 but they are not present in the same transcript, even though both of them are UTRs. A promoter pulldown assay specific for the promoter region upstream of exon 1 may give valuable information on the

transcription factor that takes role in R7 specific transcription of dmHexDC, Since the R7 specific Gal4 is expressed by the exon 1 downstream of the transcript starting from the upstream of exon 1.

APPENDIX A: VECTOR MAPS

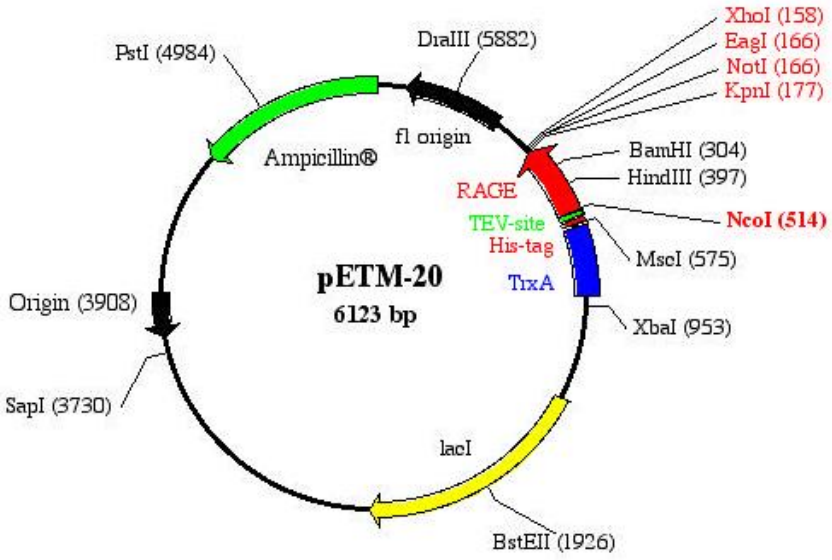

Figure A.1. Vector map of pETM-20

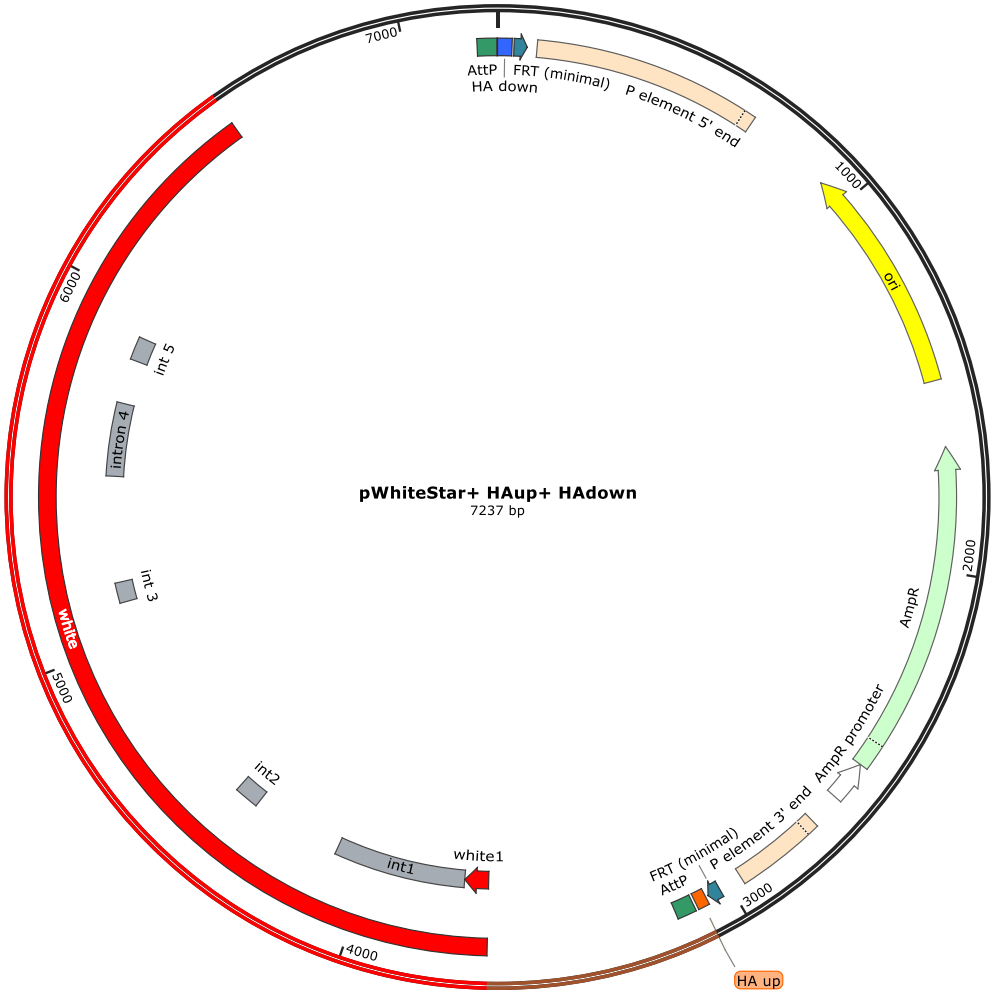

Figure A.2. Vector map of pWhiteStar +HAup+HAdown

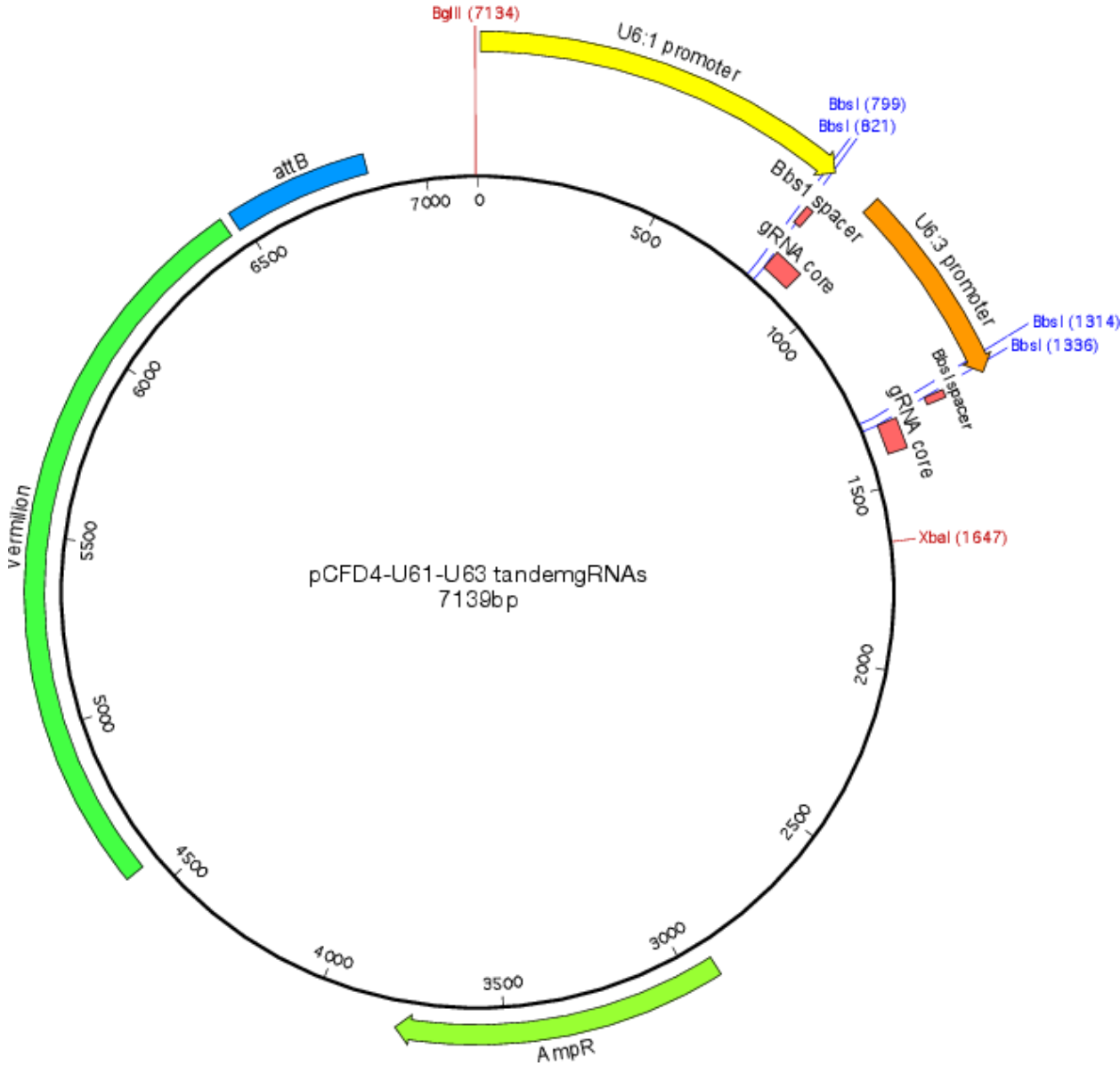

Figure A.3. Vector map of pCFD4 +gRNAs

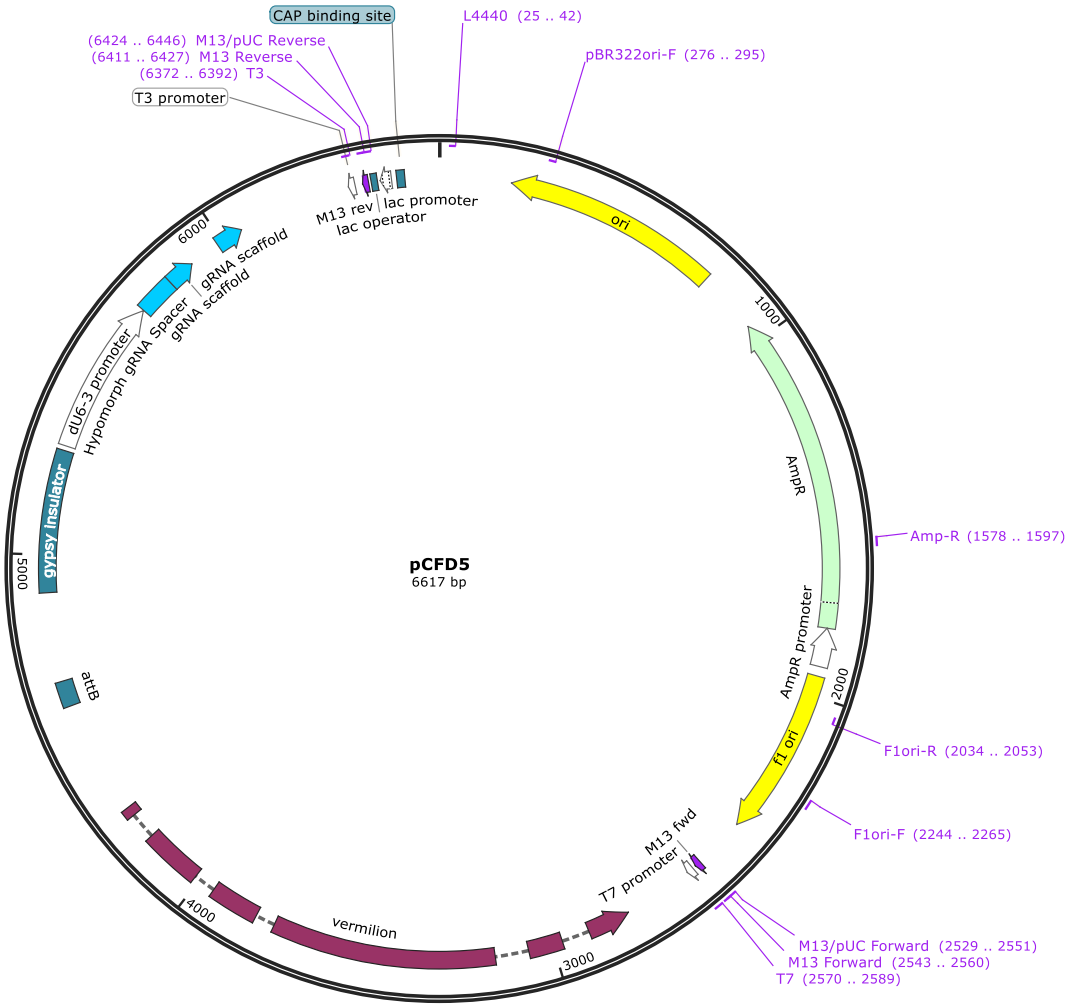

Figure A.4. Vector map of pCFD5+ Hypomorph gRNA

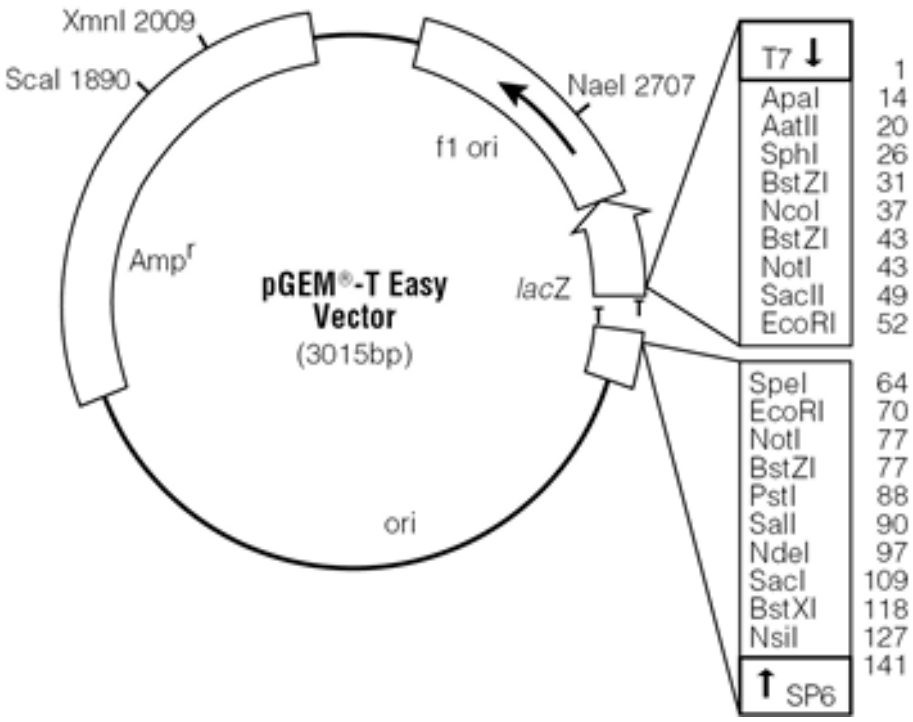

Figure A.5. Vector map of pGEM-T Easy

APPENDIX B: PROTEIN STRUCTURE

A

```
ATG CAT TCC AAC TTA TGG ATC TTT ATT TGG CGC CGC AAA TTA AGC TTT TTA CTG GTA GTA TCC GGG CTG GTT TTA CTG GGC CTT
TGG ACC TGG GCG ATT CTT ATC GAT ACC ACG GCG AGC GCA CAA CCA CCG GCC GCA CCG ATT TCT AAC AGT GAG CGC CAA CAA CAG
CAG CAG CAG TCA ACG GGC TAC CAA CAA GTT CAC ATT ATT CAA AAA CTT ATT GAG CAG GCG AAC CGT GTG TTG CGT GCT GAA CGC
TTG AAG GAC AAT GCT GTT GAA AAG TCA CCA TCT CCA GTA CAA CGC CTG GGA AAT GTA CGC TCT TTG CGT CCT ACC CAA GCG AAA
AAA GCC CCT TTG CCC ATG TCG GAA TCG TTT TTA TTC GAA CAA GAA CGT CTG AAG ATT TTG GAG CAA CAG CAA CGT CAA CGT AAT
GCG GGA ATG GAA GAA CTT AAC GCG AAC GCG GAG GAC GAA CGC ATG TTG AGT TCC GCT GAA CGC CAG ACC CAG TAT GAA CAT GAA
CTG CAG CGT ATG GGG GTT CCC GTC GTT GCG GGA ATT GGT CCT GAT GGC CCG CGT GCC CCC GCG GAG CGC CTT GTA CAT TTG GAC
TTG AAA GGG GCT CCA CCA AAG TTG TCA TTT TTG AAG CAG TTA CTG CCT GTT TTA CGT GCT CTT GGC GCG ACA GGC TTG TTA ATC
GAA TAC GAA GAT ATG TTT CCC TAT TCG GGA GTA TTG CAG CCA TTG GCT GCT CAT AAT GCT TAC AAG GAG GAT GAG CTT CGT GAC
TTT CTG GAG TGT GCC GCC TTG CAC GGT TTA TCG GTG ATG CCC CTG GTC CAA ACC TTT GGT CAT ATG GAG TAC GTA TTA AAA CTG
AGC GGA TTC GAG CAG TTA CGC GAA CTT GCT GAA AGC CCC CAA TCA ATC TGC CCA AGT CAA CCC CAA TCC ATG GCC TTG CTG GAG
CAG ATG CTT ACC CAG GTC ATC GAG TTA CAT AGC CAA TGT CTT GGA CAA GCT ACG CCA GCA ACA AAC GTT CCT GTA GCC CGC ATT
CGT TTC ACG CAC ATT CAT ATT GGG TGT GAT GAG GTT CAA CGT ATG GGG GAG TGT ACA ACA TGC CGC CAA CGT CTT CGT TCT GAA
TTG TTC CTG TCC CAT GTC GTA TCT ATG GCC CAC TTT ATC CGT CGT CAA TGG CCG CAC CTG GGG GTT GTA ATT TGG GAT GAC CAA
CTG CGT AGT ATG TCT CTG TCG GAA CTG CAA CAT TCG CAG GTC GGC AGC TAC GTC GAA CCT ATG GTA TGG GTA TAC GCA TCG GAT
ATC TAC CGC TTC ATC CAG CCC CAG CTT TGG GAT ACT TAC GCC AAG GTA TTT CCT TCC GCG TGG ACT GCC TCT GCG TTC AAA GGC
GCC TTC GGG GAA TCC CTT TTG GTG CCA CCT TTG CAA CAC CAC TTA GAG AAC AAT ATC CGC TGG TTA GCT GTT ATC GCT AAG GAA
GGA GGA CGT TTT AGC AAG GGC CTT CGC GGT TTG GCA TTG ACT GGT TGG CAA CGC TAT GAC CAT TTT GCC GTT TTG TGT GAA CTT
CTG CCT GTT GGG ATT CCT TCG CTT ATG ACG TCC CTG AGT ACG GTC TCT AAG GGC TAC TTT AGC ACC AAC CCA CGC GAT AAT GAA
CTG TTA CGC GTT TTA CAG TGT GTA TTT CAA CCA GAT TCA CGC CGC AGT GGC CGC CCG TGG CTT GAG TTA CAC CCG AAT GCT CAT
CAC TCT CAG CTT TTC GCG GTC TGC AAT TAC CCC GGA CAC TTG GTA TTT AAA TAT GCT TTG CGT CTT TTC GAT AAG CTG GCC GAG
ATC CGT CTG TAC CTG CAA CAA ACC CGC GAC CAA TCG GCT TGG CTT TCA GAT TAC AAC GTG CGT CAT AAC TTC TCT TCG CCG CTG
CGC GTG CGT GAG CTT ACC GCC CGC ACT CCG ATG CTT ATC GAG GAG TTG CGT GCA ATG GCG CGC GAA GCA CAG CAA CTG TTG TGG
GAG GTT TAC GAC GAG TAC ACA GTG ACA GAA TTC GTA GAA CAA CAC ATT TAT CCT ACG ATT GAG GCA CTT CAG CGT CAG CTG GCG
CGC GCA GAG ATG TTA CTG CAA CGC CGC ACT TGG CCG CAA CGC CCA CTT CCG CTT CCA TTA AAA GTC CAG GAA GAC ATG GGG CTT
GTC ACT CAT CAA CAA GAG CAA CAG TGA
```

B

Accepted - Low Complexity (Scores less than 7)

Some complexities exist but we do not anticipate problems with this sequence.

Total Complexity Score: 0.7

Figure B.1. dmHexDC CDS was analyzed by codon optimization tools. (A) dmHexDC CDS. (B) Output result from Integrated DNA Technologies CodonOpt tool predicts expression of *dmHexDC* CDS would not result in complications during heterologous protein production in *E.coli*.

Table B.2. Top 5 Structures with homology to dmHexDC from protein data bank that are used to construct homology model of dmHexDC *in silico*.

PDB ID	**Description**
2epo	N-acetyl-B-D-glucosaminidase (GCNA) from *Streptococcus gordonii*
5A6K	GH20C, Beta-hexosaminidase from *Streptococcus pneumoniae* in complex with Gal-NGT
2EPL	N-acetyl-B-D-glucosaminidase (GCNA) from *Streptococcus gordonii*
5A6A	GH20C, Beta-hexosaminidase from *Streptococcus pneumoniae* in complex with NGT
2EPL	N-acetyl-B-D-glucosaminidase (GCNA) from *Streptococcus gordonii*

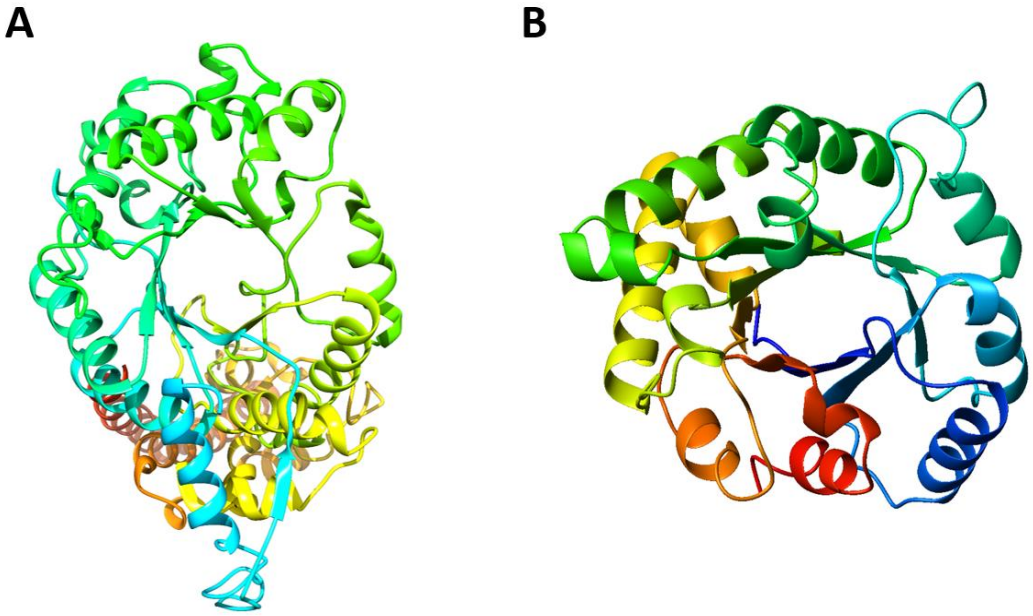

Figure B.3. Comparison of dmHexDC homology model and the structure of a typical $(\beta\alpha)_8$. (A) dmHexDC homology model. (B) Structure of Triose Phosphate Isomerase, a typical $(\beta\alpha)_8$ enzyme.

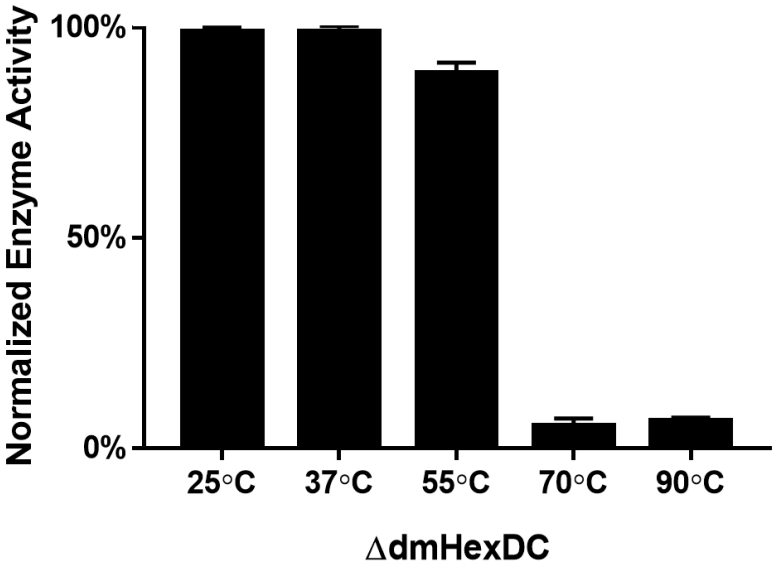

Figure B.4. Thermostability of truncated dmHexDC was assessed by preincubation at 25°C, 37°C, 55°C, 70°C and 90°C. Major loss of activity was observed in temperatures higher than 55°C (**** $p<0.0001$).

APPENDIX C: SURVEYOR PCR

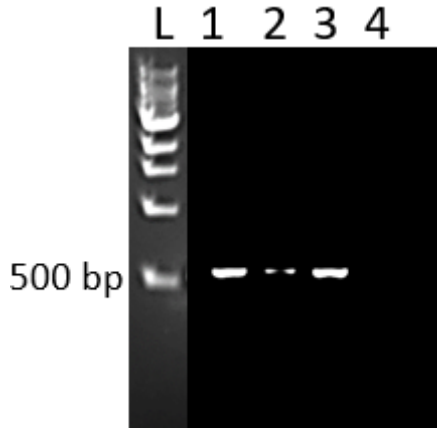

Figure C.1. Primer optimized for Surveyor PCR were experimented against PAM mutant containing plasmids WT plasmids.

L: Ladder, 1: with primer for PAM mutant and PAM mutant plasmid, 2: with primer for PAM mutant and WT mutant plasmid, 3 with primer for WT and WT plasmid 4, with primer for WT and PAM mutant plasmid template

APPENDIX D: LOCALIZATION OF dmHexDC

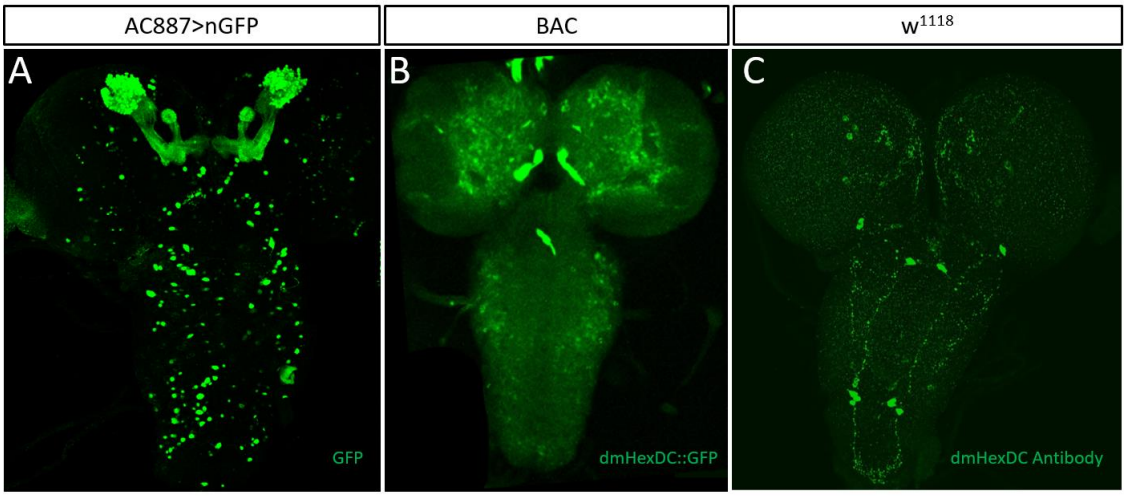

Figure D.1. Immunohistochemistry analysis of dmHexDC localization in *Drosophila* 3rd instar CNS. (A) dmHexDC nuclear staining of dmHexDC expressing cells shows nGFP expression in Kenyon cells and ventral nerve cord. (B) Overexpression of dmHexDC::eGFP. (C) Staining of WT tissue with dmHexDC antibody shows localization in ventral nerve cord and optic lobes.

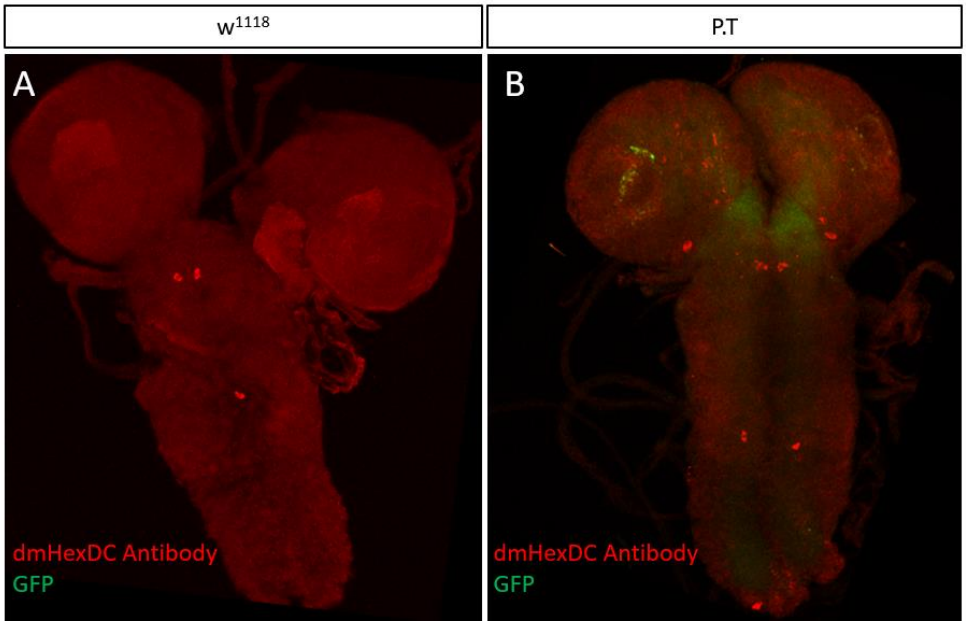

Figure D. 2. Immunohistochemistry analysis of dmHexDC localization in CNS by using dmHexDC antibody. (A) Localization in signals in WT CNS. (B) dmHexDC signals observed from *dmHexDC*^{null} genotype.

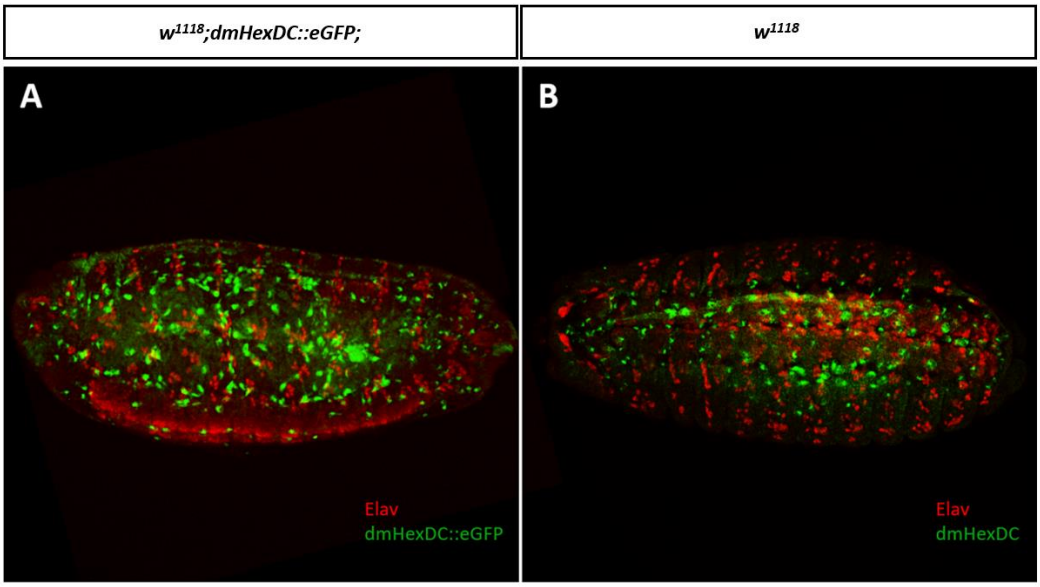

Figure D.3. dmHexDC is expressed throughout the late stage *Drosophila* embryo. (A) GFP was observed from dmHexDC::eGFP overexpression line (n=5). (B) Signals are detected similar to dmHexDC::eGFP embryos with dmHexDC antibody (n=4).

REFERENCES

Alteen, M. G., V. Oehler, I. Nemčovičová, I. B. H. Wilson, D. J. Vocadlo, and T. M. Gloster, 2016, "Mechanism of Human Nucleocytoplasmic Hexosaminidase D", *Biochemistry*, Vol. 55, No. 19, pp. 2735–2747.

Altmann, F., H. Schwihla, E. Staudacher, J. Glossl, and L. Marz, 1995, Insect cells contain an unusual membrane-bound β-N-acetylglucosaminidase probably involved in the processing of protein N-glycans, *Journal of Biological Chemistry*.

Andrade, C., 1993, "(Received July 21, 1992)", Vol. 23, No. 2, pp. 724–742.

Bakr Shabbir, M. A., H. Hao, M. Z. Shabbir, H. I. Hussain, Z. Iqbal, S. Ahmed, A. Sattar, M. Iqbal, J. Li, and Z. Yuan, 2016, "Survival and Evolution of CRISPR-Cas System in Prokaryotes and Its Applications", *Frontiers in Immunology*, Vol. 7, No. SEP, pp. 1–14.

Barrangou, R., 2015, "The Roles of CRISPR-Cas Systems in Adaptive Immunity and Beyond", *Current Opinion in Immunology*, Vol. 32, , pp. 36–41.

Bassett, A. R., and J. Liu, 2014, "CRISPR/Cas9 and Genome Editing in Drosophila", *Journal of Genetics and Genomics*, Vol. 41, No. 1, pp. 7–19.

Başdağ, Bilal. 2018. "Analysis of Cellular Pathways Involved in dmHexDC Function", M.Sc. Thesis, Boğaziçi University.

Burda, P., and M. Aebi, 1999, "The Dolichol Pathway of N-Linked Glycosylation", *Biochimica et Biophysica Acta - General Subjects*, Vol. 1426, No. 2, pp. 239–257.

Cain, J. A., N. Solis, and S. J. Cordwell, 2014, "Beyond Gene Expression: The Impact of Protein Post-Translational Modifications in Bacteria", *Journal of Proteomics*, Vol. 97, , pp. 265–286.

Chen, Y. W., J. W. Pedersen, H. H. Wandall, S. B. Levery, S. Pizette, H. Clausen, and S. M.

Cohen, 2007, "Glycosphingolipids with Extended Sugar Chain Have Specialized Functions in Development and Behavior of Drosophila", *Developmental Biology*, Vol. 306, No. 2, pp. 736–749.

Cho, K. O., and K. W. Choi, 1998, "Fringe Is Essential for Mirror Symmetry and Morphogenesis in the Drosophila Eye", *Nature*, Vol. 396, No. 6708, pp. 272–276.

Chu, S. S., C. L. Christensen, and F. Y. M. Choy, 2018, "The Impact of GC Content on CRISPR/Cas9 Gene Editing: Implications for Mucopolysaccharidosis Type IIIB", *Molecular Genetics and Metabolism*, Vol. 123, No. 2, pp. S32.

Chuai, G., Q. Wang, and Q. Liu, 2016, "In Silico Meets In Vivo : Towards Computational CRISPR-Based SgRNA Design", *Trends in Biotechnology*, Vol. 35, No. 1, pp. 1–10.

Comer, F. I., and G. W. Hart, 2000, "O-Glycosylation of Nuclear and Cytosolic Proteins. Dynamic Interplay between O-GlcNAc and O-Phosphate", *Journal of Biological Chemistry*, Vol. 275, No. 38, pp. 29179–29182.

Cui, Y., J. Xu, M. Cheng, X. Liao, and S. Peng, 2018, "Review of CRISPR/Cas9 SgRNA Design Tools", *Interdisciplinary Sciences: Computational Life Sciences*, Vol. 10, No. 2, pp. 455–465.

Dastsooz, H., M. Alipour, S. Mohammadi, F. Kamgarpour, F. Dehghanian, and M. Fardaei, 2018, "Identification of Mutations in HEXA and HEXB in Sandhoff and Tay-Sachs Diseases: A New Large Deletion Caused by Alu Elements in HEXA", *Human Genome Variation*, Vol. 5, No. December 2017, pp. 18003.

Davie, K., J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Ł. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. B González-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, N. Konstantinides, M. Fiers, P. Verstreken S. Aerts, 2018, "A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain", *Cell*, pp. 982–998.

Davies, Gideon J & Sinnot, M. L., 2008, " Sorting the diverse: The sequence-based classifications of carbohydrate-active enzymes ", *The Biochemical Society*, No. August, pp. 26–32.

Deltcheva, E., K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, and J. Vogel, 2011, "UKPMC Funders Group UKPMC Funders Group Author Manuscript Factor RNase III", Vol. 471, No. 7340, pp. 602–607.

Dersh, D., Y. Iwamoto, and Y. Argon, 2016, "Tay-Sachs Disease Mutations in HEXA Target the Chain of Hexosaminidase A to Endoplasmic Reticulum-Associated Degradation", *Molecular Biology of the Cell*, Vol. 27, No. 24, pp. 3813–3827.

Dragosits, M., S. Yan, E. Razzazi-Fazeli, I. B. H. Wilson, and D. Rendic, 2015, "Enzymatic Properties and Subtle Differences in the Substrate Specificity of Phylogenetically Distinct Invertebrate N-Glycan Processing Hexosaminidases", *Glycobiology*, Vol. 25, No. 4, pp. 448–464.

Dreos, R., G. Ambrosini, R. Groux, R. C. Perier, and P. Bucher, 2017, "The Eukaryotic Promoter Database in Its 30th Year: Focus on Non-Vertebrate Organisms", *Nucleic Acids Research*, Vol. 45, No. D1, pp. D51–D55.

Eswar, N., B. Webb, M. A. Marti-renom, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, and A. Sali, 2006, “Comparative Protein Structure Modeling Using Modeller” *Curr Protoc Bioinformatics*, *Chapter 5*, Unit-5.6.

Fernandes, M. J. G., S. Yew, D. Leclerc, B. Henrissat, C. E. Vorgias, R. A. Gravel, P. Hechtman, and F. Kaplan, 1997, "Identification of Candidate Active Site Residues in Lysosomal β- Hexosaminidase A", *Journal of Biological Chemistry*, Vol. 272, No. 2, pp. 814–820.

Gaj, T., 2014, "ZFN, TALEN and CRISPR/Cas Based Methods for Genome Engineering", *Trends in Biotechnology*, Vol. 31, No. 7, pp. 397–405.

Gerken, T. A., O. Jamison, C. L. Perrine, J. C. Collette, H. Moinova, L. Ravi, S. D. Markowitz, W. Shen, H. Patel, and L. A. Tabak, 2011, "Emerging Paradigms for the Initiation of Mucin-Type Protein O-Glycosylation by the Polypeptide GalNAc Transferase Family of Glycosyltransferases", *Journal of Biological Chemistry*, Vol. 286, No. 16, pp. 14493–14507.

Gill, D. J., J. Chia, J. Senewiratne, and F. Bard, 2010, "Regulation of O-Glycosylation through Golgi-to-ER Relocation of Initiation Enzymes", *Journal of Cell Biology*, Vol. 189, No. 5, pp. 843–858.

Gill, D. J., K. M. Tham, J. Chia, S. C. Wang, C. Steentoft, H. Clausen, E. A. Bard-Chapeau, and F. A. Bard, 2013, "Initiation of GalNAc-Type O-Glycosylation in the Endoplasmic Reticulum Promotes Cancer Cell Invasiveness", *Proceedings of the National Academy of Sciences*, Vol. 110, No. 34, pp. E3152–E3161.

Gutternigg, M., 2010, "Mammalian Cells Contain a Second Nucleocyoplasmic Hexosaminidase", *The Biochemical journal*, Vol. 419, No. 1, pp. 83–90.

Hagen, K. G. T., L. Zhang, E. Tian, and Y. Zhang, 2009, "Glycobiology on the Fly: Developmental and Mechanistic Insights from Drosophila", *Glycobiology*, Vol. 19, No. 2, pp. 102–111.

Hales, K. G., C. A. Korey, A. M. Larracuente, and D. M. Roberts, 2015, "Genetics on the Fly: A Primer on the Drosophila Model System", *Genetics*, Vol. 201, No. 3, pp. 815–842.

Harada, Y., R. Buser, E. M. Ngwa, H. Hirayama, M. Aebi, and T. Suzuki, 2013, "Eukaryotic Oligosaccharyltransferase Generates Free Oligosaccharides during N-Glycosylation", *Journal of Biological Chemistry*, Vol. 288, No. 45, pp. 32673–32684.

Heidmann, D., and C. F. Lehner, 2001, "Reduction of Cre Recombinase Toxicity in Proliferating Drosophila Cells by Estrogen-Dependent Activity Regulation", *Development Genes and Evolution*, Vol. 211, No. 8–9, pp. 458–465.

Hema Thanka Christlet, T., and K. Veluraja, 2001, "Database Analysis of O-Glycosylation Sites in Proteins", *Biophysical Journal*, Vol. 80, No. 2, pp. 952–960.

Hille, F., and E. Charpentier, 2016, "CRISPR-Cas : Biology , Mechanisms and Relevance".

Huai, C., G. Li, R. Yao, Y. Zhang, M. Cao, L. Kong, C. Jia, H. Yuan, H. Chen, D. Lu and Q. Huang, 2017, "Structural Insights into DNA Cleavage Activation of CRISPR-Cas9 System", *Nature Communications*, Vol. 8, No. 1, pp. 1–9.

Huang, Y., C. Mccann, A. Samsonov, D. Malkov, and G. D. Davis, 2017, "Modulation of Genome Editing Outcomes by Cell Cycle Control of Cas9 Expression", pp. 1–26.

Irvine, K. D., 1999, "Fringe, Notch, and Making Developmental Boundaries", *Current Opinion in Genetics and Development*, Vol. 9, No. 4, pp. 434–441.

Kaçmaz, Güner. 2013. "Characterization of a Novel R7-Specific Gene in the Drosophila Visual System.", M.Sc. Thesis, Boğaziçi University.

Kent G. Golic and Susan Lindquist, 1989, "The FLP Recombinase of Yeast Catalyzes Site-Specific Recombination in the Drosophila Genome", Vol. 59, , pp. 399.

Kıral, Ferdi Rıdvan. 2015. "Functional Characterization of a Novel Hexosaminidase, CG7985, In Drosophila Eye Development.", M.Sc. Thesis, Boğaziçi University.

Kumar, J. P., and D. F. Ready, 1995, "Rhodopsin Plays an Essential Structural Role in Drosophila Photoreceptor Development", *Development*, Vol. 121, pp. 4359–4370.

Lee, J., T. Tonozuka, and M. Jayaram, 1997, "Mechanism of Active Site Exclusion in a Site-Specific Recombinase: Role of the DNA Substrate in Conferring Half-of-the-Sites Activity", *Genes and Development*, Vol. 11, No. 22, pp. 3061–3071.

Lee, T., 2014, Generating mosaics for lineage analysis in flies, *Wiley Interdisciplinary Reviews: Developmental Biology*, Vol. 3, pp. 69-81

Lefèvre, F., M. H. Rémy, and J. M. Masson, 1997, "Alanine-Stretch Scanning Mutagenesis: A Simple and Efficient Method to Probe Protein Structure and Function", *Nucleic Acids Research*, Vol. 25, No. 2, pp. 447–448.

Leinekugel, P., S. Michel, E. Conzelmann, and K. Sandhoff, 1992, "Quantitative Correlation between the Residual Activity of β-Hexosaminidase A and Arylsulfatase A and the Severity of the Resulting Lysosomal Storage Disease", *Human Genetics*, Vol. 88, No. 5, pp. 513–523.

Léonard, R., D. Rendić, C. Rabouille, I. B. H. Wilson, T. Préat, and F. Altmann, 2006, "The Drosophila Fused Lobes Gene Encodes an N-Acetylglucosaminidase Involved in N-Glycan Processing", *Journal of Biological Chemistry*, Vol. 281, No. 8, pp. 4867–4875.

Liu, T., M. Xia, H. Zhang, H. Zhou, J. Wang, X. Shen, and Q. Yang, 2015, "Exploring NAG-Thiazoline and Its Derivatives as Inhibitors of Chitinolytic β-Acetylglucosaminidases", *FEBS Letters*, Vol. 589, No. 1, pp. 110–116.

Liu, X., A. Homma, J. Sayadi, S. Yang, J. Ohashi, and T. Takumi, 2016, "Sequence Features Associated with the Cleavage Efficiency of CRISPR/Cas9 System", *Scientific Reports*, Vol. 6, pp. 1–9.

Lyne, R., R. Smith, K. Rutherford, M. Wakeling, A. Varley, F. Guillier, H. Janssens, W. Ji, P. Mclaren, P. North, D. Rana, T. Riley, J. Sullivan, X. Watkins, M. Woodbridge, K. Lilley, S. Russel, M. Ashburner, K. Mizuguchi and G. Micklem, 2007, "FlyMine: An Integrated Database for Drosophila and Anopheles Genomics", *Genome Biology*, Vol. 8, No. 7.

Maier, T., N. Strater, C. G. Schuette, R. Klingenstein, K. Sandhoff, and W. Saenger, 2003, "The X-Ray Crystal Structure of Human β-Hexosaminidase B Provides New Insights into Sandhoff Disease", *Journal of Molecular Biology*, Vol. 328, No. 3, pp. 669–681.

Mark, B. L., D. J. Vocadlo, S. Knapp, B. L. Triggs-Raine, S. G. Withers, and M. N. G. James, 2001, "Crystallographic Evidence for Substrate-Assisted Catalysis in a Bacterial β-Hexosaminidase", *Journal of Biological Chemistry*, Vol. 276, No. 13, pp. 10330–10337.

Mazzoni, E. O., A. Celik, M. F. Wernet, D. Vasiliauskas, R. J. Johnston, T. A. Cook, F. Pichaud, and C. Desplan, 2008, "Iroquois Complex Genes Induce Co-Expression of Rhodopsins in Drosophila", *PLoS Biology*, Vol. 6, No. 4, pp. 825–835.

Mcmahan, Z. H., and F. M. Wigley, 2015, "HHS Public Access", Vol. 5, No. 3, pp. 355–370.

Nern, A., B. D. Pfeiffer, K. Svoboda, and G. M. Rubin, 2011, "Multiple New Site-Specific Recombinases for Use in Manipulating Animal Genomes", *Proceedings of the National Academy of Sciences*, Vol. 108, No. 34, pp. 14198–14203.

Nienhaus, U., 2013, "On Growth and Form and Patterning of Drosophila Wing Imaginal Discs", pp. 1–107.

Okajima, T., and K. D. Irvine, 2002, "Regulation of Notch Signaling by O-Linked Fucose", *Cell*, Vol. 111, No. 6, pp. 893–904.

Öztürk, A., 2010, "Characterization of Genes Involved in Photoreceptor Differentiation", M.Sc. Thesis, Boğaziçi University.

Pandey, U. B., and C. D. Nichols, 2011, "Human Disease Models In", Vol. 63, No. 2, pp. 411–436.

Pásztói, M., B. Sódar, P. Misják, K. Pálóczi, Á. Kittel, K. Tóth, K. Wellinger, P. Géher, G. Nagy, T. Lakatos, A. Falus and E. I. Buzás, 2013, "The Recently Identified Hexosaminidase D Enzyme Substantially Contributes to the Elevated Hexosaminidase Activity in Rheumatoid Arthritis", *Immunology Letters*, Vol. 149, No. 1–2, pp. 71–76.

Pizette, S., C. Rabouille, S. M. Cohen, and P. Therond, 2009, "Glycosphingolipids Control the Extracellular Gradient of the Drosophila EGFR Ligand Gurken", *Development*, Vol. 136, No. 4, pp. 551–561.

Rajalingam, D., C. Loftis, J. J. Xu, and T. K. S. Kumar, 2009, "Trichloroacetic Acid-Induced Protein Precipitation Involves the Reversible Association of a Stable Partially Structured Intermediate.", *Protein science : a publication of the Protein Society*, Vol. 18, No. 5, pp. 980–993.

Rath, D., L. Amlinger, A. Rath, and M. Lundgren, 2015, "The CRISPR-Cas Immune System: Biology, Mechanisms and Applications", *Biochimie*, Vol. 117, , pp. 119–128.

Raymond, C. S., and P. Soriano, 2007, "High-Efficiency FLP and ΦC31 Site-Specific Recombination in Mammalian Cells", *PLoS ONE*, Vol. 2, No. 1, pp. 1–4.

Ren, X., Z. Yang, J. Xu, J. Sun, D. Mao, Y. Hu, S. J. Yang, H. H. Qiao, X. Wang, … J. Q. Ni, 2014, "Enhanced Specificity and Efficiency of the CRISPR/Cas9 System with Optimized SgRNA Parameters in Drosophila", *Cell Reports*, Vol. 9, No. 3, pp. 1151–1162.

Rosenbaum, E. E., E. Vasiljevic, K. S. Brehm, and N. J. Colley, 2014, "Mutations in Four Glycosyl Hydrolases Reveal a Highly Coordinated Pathway for Rhodopsin Biosynthesis and N-Glycan Trimming in Drosophila Melanogaster", *PLoS Genetics*, Vol. 10, No. 5,.

Sandhoff, K., and K. Harzer, 2013, "Gangliosides and Gangliosidoses: Principles of Molecular and Metabolic Pathogenesis", *Journal of Neuroscience*, Vol. 33, No. 25, pp. 10195–10208.

Schwarz, F., and M. Aebi, 2011, "Mechanisms and Principles of N-Linked Protein Glycosylation", *Current Opinion in Structural Biology*, Vol. 21, No. 5, pp. 576–582.

Scott, H., and V. M. Panin, 2015, "N-Glycosylation in Regulation of the Nervous System",

Glycobiology, Vol. 24, , pp. 367–394.

Shmakov, Sergey; Sitnik, Vassilii; Makarova, Kira; Wolf, Yuri; Severinov, Konstantin; Koonin, E., 2017, " The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes", *mBio*, Vol. 8, No. November, pp. 1–18.

Smith, M. C. M., and H. M. Thorpe, 2002, Diversity in the serine recombinases, *Molecular Microbiology*.

Song, G., M. Jia, K. Chen, X. Kong, B. Khattak, C. Xie, A. Li, and L. Mao, 2016, "CRISPR/Cas9: A Powerful Tool for Crop Genome Editing", *Crop Journal*, Vol. 4, No. 2, pp. 75–82.

Sontheimer, L. A. M. and E. J., 2011, "NIH Public Access", *CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea*, Vol. 11, No. 3, pp. 181–190.

Stanley, P., and T. Okajima, 2010, "Roles of Glycosylation in Notch Signaling", *Current Topics in Developmental Biology*, Vol. 92, No. C, pp. 131–164.

Ungar, D., 2009, "Golgi Linked Protein Glycosylation and Associated Diseases", *Seminars in Cell and Developmental Biology*, Vol. 20, No. 7, pp. 762–769.

van Kooyk, Y., H. Kalay, and J. J. Garcia-Vallejo, 2013, Analytical tools for the study of cellular glycosylation in the immune system, *Frontiers in Immunology*.

Venken, K. J. T., K. L. Schulze, N. A. Haelterman, H. Pan, Y. He, M. Evans-Holm, J. W. Carlson, R. W. Levis, A. C. Spradling, R. A Hoskins and H. J. Bellen, 2011, "MiMIC: A Highly Versatile Transposon Insertion Resource for Engineering Drosophila Melanogaster Genes", *Nature Methods*, Vol. 8, No. 9, pp. 737–747.

Vilain, S., R. Vanhauwaert, I. Maes, N. Schoovaerts, L. Zhou, S. Soukup, R. da Cunha, E. Lauwers, M. Fiers, and P. Verstreken, 2014, "Fast and Efficient Drosophila Melanogaster Gene Knock-Ins Using MiMIC Transposons", *G3;*

Genes|Genomes|Genetics, Vol. 4, No. 12, pp. 2381–2387.

Wang, J., T. Zhang, R. Liu, M. Song, J. Wang, J. Hong, Q. Chen, and H. Liu, 2017, "Recurring Sequence-Structure Motifs in (Bα)8-Barrel Proteins and Experimental Optimization of a Chimeric Protein Designed Based on Such Motifs", *Biochimica et Biophysica Acta - Proteins and Proteomics*, Vol. 1865, No. 2, pp. 165–175.

Wenger, D. A., S. Coppola, and S. L. Liu, 2003, "Insights into the Diagnosis and Treatment of Lysosomal Storage Diseases", *Archives of Neurology*, Vol. 60, No. 3, pp. 322–328.

Werz, D. B., R. Ranzinger, S. Herget, A. Adibekian, C. W. Von der Lieth, and P. H. Seeberger, 2007, "Exploring the Structural Diversity of Mammalian Carbohydrates ("glycospace") by Statistical Databank Analysis", *ACS Chemical Biology*, Vol. 2, No. 10, pp. 685–691.

Yamasaki, Y., L. Tsuda, A. Suzuki, and K. Yanagisawa, 2018, "Induction of Ganglioside Synthesis in Drosophila Brain Accelerates Assembly of Amyloid β Protein", *Scientific Reports*, Vol. 8, No. 1, pp. 4–10.

Yao, X., X. Wang, X. Hu, Z. Liu, J. Liu, H. Zhou, X. Shen, Y. Wei, Z. Huang, … H. Yang, 2017, "Homology-Mediated End Joining-Based Targeted Integration Using CRISPR / Cas9", *Nature Publishing Group*, Vol. 27, No. 6, pp. 801–814.

Young, C. L., Z. T. Britton, and A. S. Robinson, 2012, "Recombinant Protein Expression and Purification : A Comprehensive Review of Affinity Tags and Microbial Applications", pp. 620–634.

Zhang, R., V. L. Y. Yip, S. G. Withers, and B. Columbia, 2010, "Mechanisms of Enzymatic Glycosyl Transfer", *Comprehensive Natural Products II Chemistry and Biology*, pp. 385–422.

Zhang, X. H., L. Y. Tee, X. G. Wang, Q. S. Huang, and S. H. Yang, 2015, "Off-Target Effects in CRISPR/Cas9-Mediated Genome Engineering", *Molecular Therapy* -

Nucleic Acids, Vol. 4, No. 11, pp. e264.

Zhang, X., W. H. Koolhaas, and F. Schnorrer, 2014, "A Versatile Two-Step CRISPR- and RMCE-Based Strategy for Efficient Genome Engineering in Drosophila", *G3: Genes, Genomes, Genetics*, Vol. 4, No. 12, pp. 2409–2418.

Zhou, Z. D., U. Kumari, Z. C. Xiao, and E. K. Tan, 2010, "Notch as a Molecular Switch in Neural Stem Cells", *IUBMB Life*, Vol. 62, No. 8, pp. 618–623.

Zhu, L., H. Mon, J. Xu, J. M. Lee, and T. Kusakabe, 2015, "CRISPR / Cas9-Mediated Knockout of Factors in Non-Homologous End Joining Pathway Enhances Gene Targeting in Silkworm Cells", *Nature Publishing Group*, Vol 5, November, pp. 1–13.