University of Sussex

# The molecular regulation of Hox gene RNA processing during *Drosophila* embryonic development

**Casandra Edelweiss Villava Robles**

Submitted for the degree of Doctor of Philosophy

University of Sussex

2014

# Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Casandra Edelweiss Villava Robles

UNIVERSITY OF SUSSEX

Casandra Edelweiss Villava Robles, Doctor of Philosophy

# The molecular regulation of Hox gene RNA processing during *Drosophila* embryonic development

## Summary

The Hox genes encode a family of developmental regulators that are essential for the normal patterning of the animal body axis. Their correct expression is controlled by a number of mechanisms including RNA processing, a molecular system that allows the formation of alternative mRNAs from a single gene. Previous work in the Alonso Lab has demonstrated that RNA processing of the *Drosophila* Hox gene Ultrabithorax (Ubx) by means of alternative splicing and alternative polyadenylation plays an important role during *Drosophila* development, but the mechanisms underlying these regulatory processes are not well understood. In this project we found that the *Drosophila* neural RNA binding protein ELAV has an important role in the control of both Ubx alternative splicing and polyadenylation during embryonic development. Furthermore, by conducting a series of *in vitro* experiments we demonstrate that ELAV is able to interact with two specific RNA elements located within Ubx intronic sequences and mutation of such elements abolishes the interaction. We also establish that embryos carrying a loss of function mutation in the elav gene produce lower levels of Ubx mRNA and protein suggesting a role of ELAV on Hox gene expression. Finally we investigated the roles that other *Drosophila* factors including RNA binding proteins, chromatin regulators and splicing regulatory proteins may have on Ubx RNA processing and found several potential regulators. All in all our work contributes to the understanding of the molecular basis of Hox RNA processing control during *Drosophila* development.

For Emyr

# Acknowledgements

I would like to thank Dr. Claudio Alonso for the opportunity to join his lab during these 4 years of intense research, as well as past and present members of the Alonso Lab. Ana, Stefan and Elvira for their support and friendship; Wan, Raul, Richard, Pedro and Joao for making this experience scientifically stimulating and fun.

Thanks to Dr. Matthias Soller for providing the *elav* constructs used in this thesis, Ali Mumtaz for the purification of the recombinant Elav protein and all I learned from him during his time in our lab. Thanks to Adrien Savy for cloning and mutagenise some elav binding sites and nice scientific chats.

I am very grateful to the members of the Couso Lab for their constant support with materials, equipment, protocols, critical thought and friendship: Juan Pablo, John, Sarah, Chris, Inyaki, Rose, Miguel Angel, Unum, Ali but particularly to Julie Aspden and Emile.

To the University of Sussex Drosophila Research groups for providing critical thinking for my research, interesting points of view and ideas, as well as great time during the yearly retreats.

A special thanks to my second supervisor Dr. Simon Morley for his support and comments on my research.

To my friends for keeping me sane all the way: Gabriel, Citlalli, Antonio, Erika, Luis, Alberto, Raul, Eugenia, Vanessa, Adrian, Isabel. And the rest of the Latin American gang!!

This project would not been possible without the financing from the Mexican Government through CONACyT and SEP postgraduate scholarships.

To the University of Sussex and the School of Life Sciences.

An eternal thanks for my family and friends back in Mexico for supporting me unconditionally, not forgetting me during these years so far away from home.

# Contents

# Chapter 1

# Introduction

## 1.1 Control of gene expression

Gene expression is a very complex process that is highly controlled at several levels. Spatial and temporal orchestration of the elements of gene expression is crucial for all cellular processes, as well as growth, development and function of higher organisms (Alonso and Wilkins, 2005), (Bhattacharjee et al., 2013).

During gene expression, genetic information flows from DNA to RNA through transcription. Transcription requires the recognition of a promoter sequence (or sequences) by factors and the transcription machinery in a sequence specific manner (Bhattacharjee et al., 2013). The recognition of such sequence modules will depend on the transcription factors and cofactors available at that time and in that particular cell or tissue. At this level, chromatin remodelling also plays an important and less understood role on overall transcription initiation. The spatial structure and compaction of the DNA around nucleosomes dictates the accessibility of a sequence elements to transcription factors (Alonso and Wilkins, 2005),(Jiang and Pugh, 2009). Studies have revealed that the presence of nucleosomes is reduced in intergenic DNA regions containing active promoters, potentially making them more accessible to the transcription machinery. However, regions tightly packed in nucleosomes can become accessible by the action of chromatin remodelling complexes, such as the SWI/SNF complex (Orphanides and Reinberg, 2002), (Carey, 2005), (Jiang and Pugh, 2009), (Euskirchen et al., 2012), and references therein.

## 1.2 Post-transcriptional gene regulation

The role of RNA is not just to transmit information, but to process it via several biochemical reactions that modify its order and structure. Post-transcriptional gene regulation refers to the processes at the RNA level and consist of mRNA of modifications such as the addition of the cap structure at the 5' end of the transcript, splicing, polyadenylation and editing that affects how this messenger is translated, transported and degraded, giving as a result, an extremely finely spatial and temporal regulated expression (Alonso and Wilkins, 2005), (Matlin et al., 2005), (Wang and Burge, 2008), (Millevoi and Vagner, 2010), (Licatalosi and Darnell, 2010). In this Introduction I will focus on splicing and polyadenylation only.

### 1.2.1 Splicing

Splicing is a posttranscriptional process where introns are removed from between exons, which are ligated in one non-interrupted sequence suitable for translation. When a pre-mRNA is just being transcribed, it contains exons separated by non coding introns, which are delimited by splice sites. Introns are delimited by 5' splice site (whose consensus sequence in mammals is AG-GU) and a 3' splice site (whose consensus sequence is AG-G). Introns also contain a conserved adenosine which constitutes the branching point, needed for the splicing reaction, followed by a conserved polypyrimidine tract. (Figure 1.1A). These sequences constitute the core elements for splicing, which will be recognised by the core splicing machinery (Jurica and Moore, 2003).

The splicing of an intron starts with the recognition of the 5' splice site by U1 snRNP forming the complex E (Figure 1.1B). The 3 splice site, in turn, is recognised by the subunit 35kDa from the U2AF complex (Matlin et al., 2005). SF1 splice factor recognises the branching point forming the complex A. Complex B forms with the addition of U2 snRNP and displacement of SF1. The catalytic complex is termed C and promotes the 2 trans - esterification steps, the first one to produce the intron lariat and the second one to join the exons together (Matlin et al., 2005). Auxiliary *cis* and *trans* factors are also important for the control of splicing, particularly in alternative splicing, explained below.

Most of the *cis* elements playing a role in splicing choices are located within 150 nt of the splicing sites. This implies that distance from splicing sites is important for the function of the element and it can be used to search for functional sites using bioinformatic and biochemistry tools (Brooks et al., 2011).

#### Alternative splicing control

Transcripts can undergo alternative splicing (AS), where different exon inclusion generates different isoforms from one single gene, increasing the number of 3' and 5' UTRs and possible proteins translated from it. Alternative splicing is a way of generating transcript diversity by using optional exons and sections of exons, it is important for the tissue specificity of gene expression and strikingly, more than 90 % of human genes undergo AS (Black, 2003), (Matlin et al., 2005), (Keren et al., 2010), (Nilsen and Graveley, 2010).

Alternative splice sites often have weaker or non canonical splice sites that can be selected by factors that can activate or inhibit them. Exon enhancers and silencers or intron enhancers and silencers sequences aid in the process of controlling alternative splicing (Figure 1.2A). These sequences are recognised by trans acting factors such as SR proteins (arginine-serine-rich proteins) or hnRNPs, to enhance or suppress the splicing of the alternative exon (Matlin et al., 2005), (Keren et al., 2010). As a general rule, SR proteins serve as splicing activators, by facilitating the recruitment of the splicing machinery whereas hnRNPs function as splicing repressors (Shepard and Hertel, 2009), (Nilsen and Graveley, 2010). Figure 1.2B shows two of the most common splicing types: cassette exon skipping (grey) and upstream splice site (pink). Many of these factors have tissue specific expression, resulting in a tissue specific control of alternative splicing. The combination of different factors and different stoichiometries constitute a sort of cellular code, for RNA processing control (Matlin et al., 2005).

### 1.2.2 3' end processing

Another very important event in pre-mRNA processing is 3 end formation, which is defined by the site of cleavage and polyadenylation of the 3' UTR. Polyadenylation consists in the cleavage and synthesis of an mRNA poly(A) tail by polyA

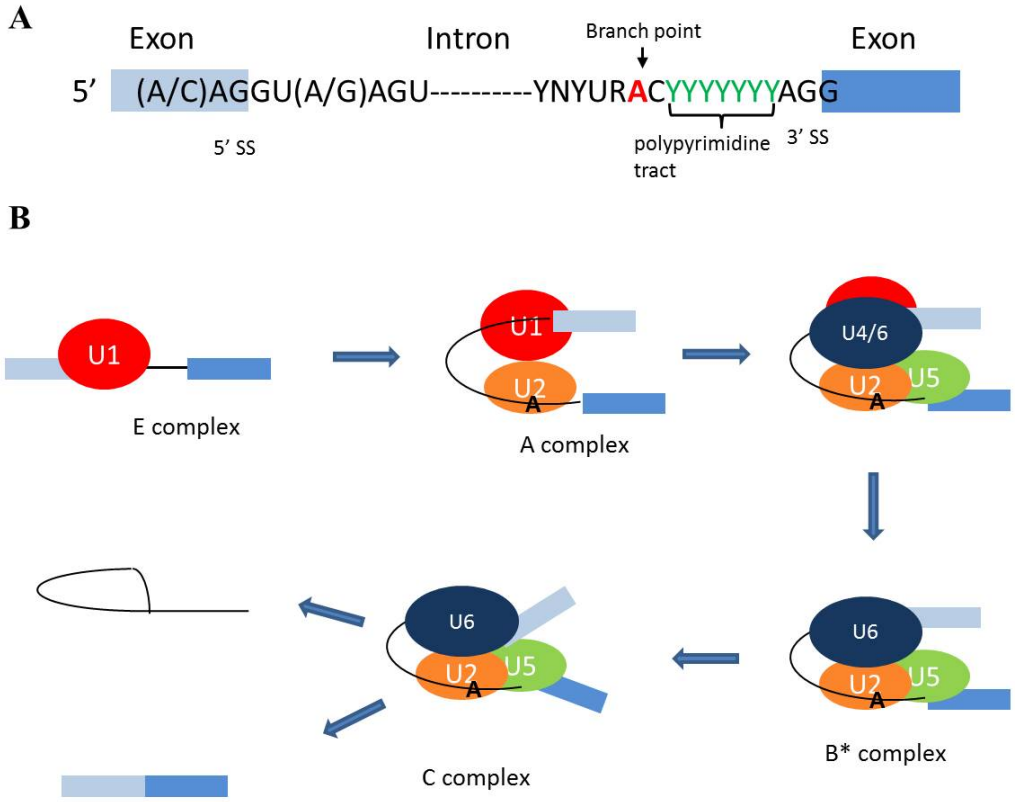


Figure 1.1: Basic mechanism of splicing. A) Representation of the core splicing *cis*-elements signals showing the exons in blue boxes. The splice sites are shown in the exon/intron junction, the branch point is marked by an arrow (red A) and the polypyrimidine tract is marked in green. B) Sequence of spliceosome recruitment and the formation of the intron lariat and binding of the exons. The 5' SS is first recognised by the U1 complex (E complex), next, the U2 complex binds the branch point and circularises the RNA (A complex). U4/6 and U5 are recruited (B* complex), then U1 and U4 are released to form the C complex, which catalyses the two trans-esterification steps to form the intronic lariat and join together the exons. 5' SS: 5' splice site, 3' SS: 3' splice site. Figure modified from (Jurica and Moore, 2003).

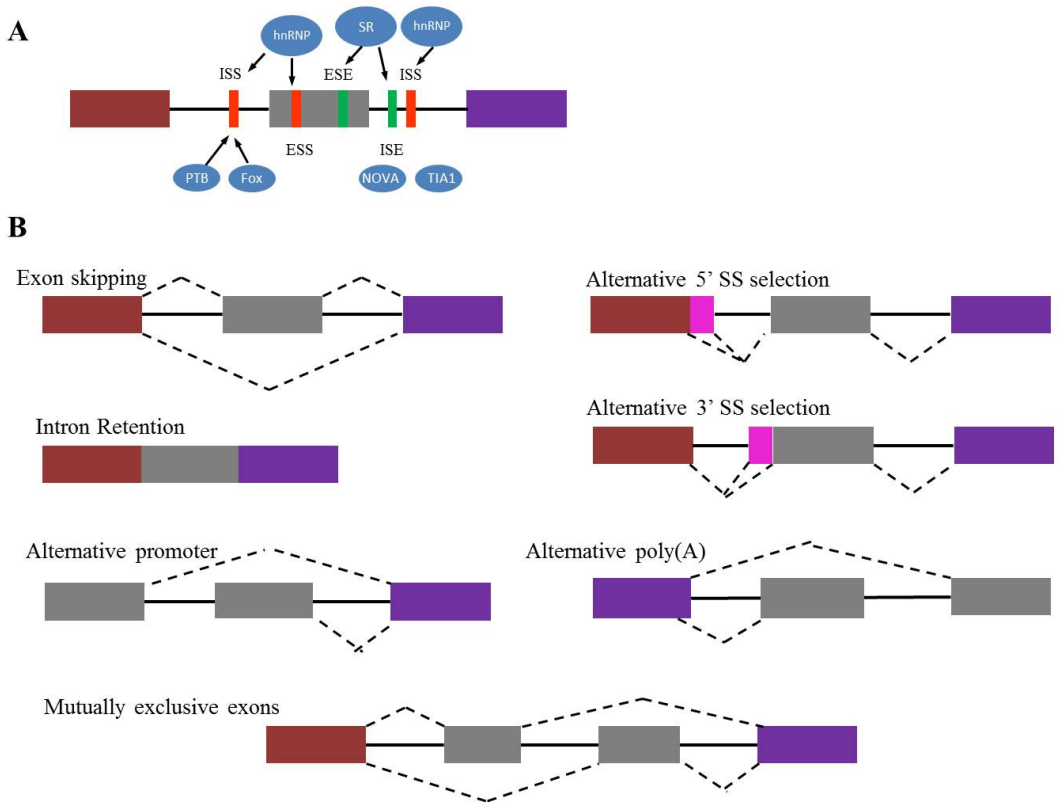


Figure 1.2: Alternative splicing. A) Position of *cis* elements that enhance or silence the splicing of the alternative exon (gray), the blue and purple boxes represent constitutive exons. The enhancing elements are represented with green bars whereas the repressing elements are represented by red bars. Potential trans factors are represented in blue ovals. B) The most common alternative splicing mechanisms. Alternative exon is shown in gray, constitutive exons in brown and blue. Exon skipping: The alternative exon is included when its 5' and 3' splicing sites are recognised and is not included when its splicing sites are skipped. Intron retention: the splicing machinery does not recognise the intronic splicing sites and the intron is included in the mature RNA. Alternative promoter: an exon is not included if a downstream promoter is used during transcription. Mutually exclusive exons: the inclusion of the first exon blocks the inclusion of the downstream exon, whereas the exclusion of the first exon allows the inclusion of the second. Alternative 5' or 3' SS selection: when a constitutive exon has two 5' SS and the downstream is used, a sequence is included in the exon; whereas the use of the upstream 3' SS leads to the inclusion of a fragment into the downstream exon. Alternative polyadenylation: the presence of a Polyadenylation site within an intron leads to the loss of the downstream exons. ISS: intronic splicing silencer, ESE: exonic splicing enhancer, ESS: exonic splicing silencer, ISE: Intronic splicing enhancer, PAS: poly adenylation site. Figure modified from (Jurica and Moore, 2003) and (Keren et al., 2010).

polymerase (PAP) enzyme. Having such sequence helps in translation initiation because it is recognised by the PABP that in turn binds the translation initiation complex, circularising the mRNA and promoting ribosome recruitment and recycling (Millevoi and Vagner, 2010). The polyA tail also protects the RNA from 3' exonucleases and extends its half life. In turn, removing this tail is the first step in degrading a transcript that is no longer needed (Bevilacqua et al., 2003),(Eulalio et al., 2009), (Alonso, 2012). The polyadenylation signal in mammals, AAUAAA (or the Proudfoot box) is one of the most conserved sequence elements known and usually lies 10-30 bases upstream of the cleavage/polyadenylation site (Colgan and Manley, 1997), (Proudfoot, 2011). A second element, a GU rich motif is 20-40 bases downstream of the cleavage site, however, this element is less conserved. Both elements, their sequence and distance between them, determine the strength of the site. The polyadenylation signal is recognised by the Cleavage and Polyadenylation Specificity Factor (CPSF) whereas the Cleavage stimulation Factor (CstF) enhances the RNA/CPSF interaction by binding to the GU rich region. The PolyA polymerase or PAP then adds up to 300 adenosines after cleavage of the RNA (Colgan and Manley, 1997) (Figure 1.3A).

There may be several polyadenylation signals in an mRNA. The particular balance between trans factors recognising the different elements described above, will determine which signal will be used and in turn, the 3' end (Figure 1.3B) (Takagaki et al., 1996). polyadenylation signals can be located at different positions of the 3' UTR or within introns, which when used, will lead to the elimination of the exons downstream, changing the carboxyl end of the protein translated from that message (Takagaki et al., 1996), (Soller and White, 2003), (Di Giammartino et al., 2011), (Mueller et al., 2013). 3' UTRs contain protein binding sites, miRNA target sequences and might form secondary structures recognisable by other factors. This information will direct the destiny of the mRNA in terms of stability, translation or localisation. Thus, the same coding sequence could have different translation or degradation rate, or even localisation if it has a different 3' UTR (St Johnston, 2005), (Andreassi and Riccio, 2009), (Millevoi and Vagner, 2010), (Alonso, 2012).

RNA localisation allows the spacial control of protein translation within a cell, important for many cellular processes such as cell signaling and polarization (An-

dreassi and Riccio, 2009), (Martin and Ephrussi, 2009), (Jung et al., 2014), (Parton et al., 2014). RNAs to be "localised" are transcribed in the nucleus and coated with proteins (forming mRNA-protein complexes) that will transport it and inhibit its translation until it reaches its final destination, where translation is enabled (Holt and Schuman, 2013). Translationally silenced RNAs can travel a considerable distance, for example in neurons, where they move from the nucleus to the axons or dendrites. For example, the *calcium/calmodulin-dependent protein kinase (CaMKII alpha)* and *activity-regulated cytoskeleton-associated protein (arc)* mRNAs are transported to dendrites as silenced RNAs in RNP granules by KIF5 (Kinesin Family) along microtubules (Hirokawa et al., 2010).

Another example is the mechanism of Drosophila oocyte patterning that involves the localisation of mRNAs such as *nanos (nos), bicoid (bcd), gurken (gk), oskar (osk)* (Johnstone and Lasko, 2001). Drosophila *bcd* mRNA contains an 3' UTR that forms a complex secondary structure of multiple stem-loops which are involved in the RNA localisation by *staufen (stau)* at *bcd* the anterior pole of the oocyte and early embryo. Other factors participate in the transportation of the bcd mRNA serving as a link to the microtubule network (*swallow (swa)* (Johnstone and Lasko, 2001).

### Regulation of alternative polyadenylation

The selection of one or another polyadenylation site is often determined by the concentration of the general polyadenylation factors, for example, a strong PAS will be used in low concentration factors whereas a weak site will be used only on high concentration of factors (Millevoi, 2009), (Di Giammartino et al., 2011). Another mechanism involved in PAS selection is provided by the presence of auxiliary factors that are expressed often in a tissue specific manner. An example of this type of factor is NOVA (Licatalosi et al., 2008), (Di Giammartino et al., 2011). Genome-wide High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) experiments using the human homologue of Pasilla (NOVA) revealed numerous NOVA-RNA interaction sites in 3' UTRs, in the vicinity of putative PAS and when these sites were mutated, the general sites of polyadenylation were modified (Licatalosi et al., 2008). When NOVA binds close to the PAS (within 30 nt) its role is inhibitory, presumably by interfering with the assembly of the polyadeny-

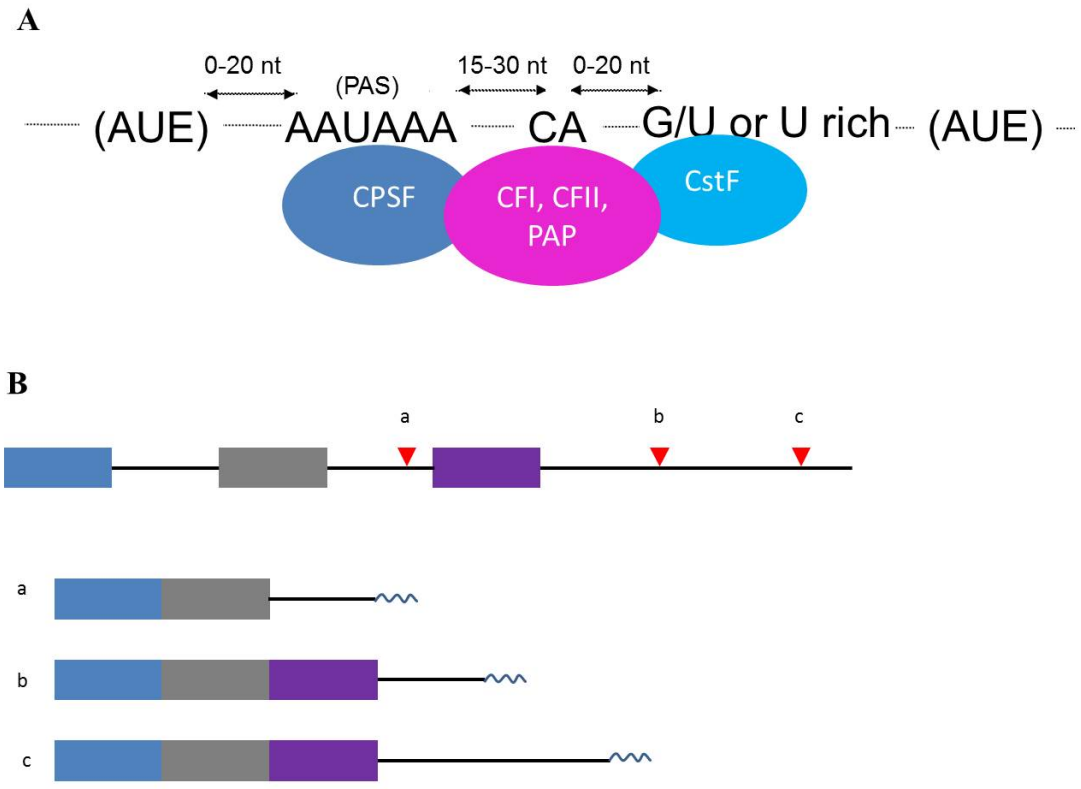


Figure 1.3: Polyadenylation and alternative polyadenylation. A) Position of *cis* elements needed for polyadenylation showing the polyadenylation signal (PAS), the auxiliary upstream element (AUE), the cleavage position (CA) and the auxiliary downstream element (ADE). The arrows show the distance between the elements. The trans elements are shown in ovals, dark blue for CPSF bound to the PAS, light blue for CstF recognising the G/U rich sequence and the cleavage site recognised by CFI, CFII and PAP. B) Figure modified from (Lutz, 2008).

lation machinery. When the NOVA binding site is further away from the PAS, it stimulates the use of such site (Licatalosi et al., 2008), (Di Giammartino et al., 2011).

### Post-transcriptional RNA regulation by microRNAs

Another class of *trans* regulators involves small RNAs such as microRNAs (miRNAs). miRNAs are short dsRNAs that regulate the stability and translation of mRNAs in a sequence specific manner (Ambros, 2003), (Huntzinger and Izaurralde, 2011) (Fig 1.4). The mirRNAs are gene encoded but they are also found within introns of other genes (He and Hannon, 2004). miRNAs are processed from longer transcripts (pri-miRNAs) by the RNaseIII-like Drosha in the nucleus into

pre-miRNAs of about 70 nt long and transported to the cytoplasm by Exportin-5 (He and Hannon, 2004) (Fig 1.4 A). Once in the cytoplasm, Dicer process the precursor hairpin (or pre-miRNA) further to an RNA duplex of about21-25 nt (Fig 1.4 B). The miRISC complex (miRNA Induced Silencing Complex) is loaded with the miRNA containing strand, which is used to select the target mRNA by base-pair interactions, retaining the target with partially or fully complementary sequence at sites often within the 3' UTR, but also found within the coding sequence and 5' UTR (Huntzinger and Izaurralde, 2011). Most of the miRNAs studied so far show a partially complementarity to its targets, leading to a reduction of the target protein production. This can be achieved by blocking the translation machinery or by recruiting RNA degradation machinery (Fig 1.4 C). miRNAs or siRNAs (small interference RNAs) having sequences fully complementary to their targets often cleave their target mRNAs followed by RNA degradation by XRN1 and exosome RNases (Fig 1.4 D).

miRNAs have important roles in very diverse biological functions including plant and animal development, cell proliferation and differentiation, apoptosis and metabolism (Ambros, 2003), (Kedde and Agami, 2008), (Jones and Newbury, 2010), (Huntzinger and Izaurralde, 2011). For example, the *bantam* miRNA controls the expression of the pro-apoptotic gene *hid*. Loss of *bantam* miRNA leads to up-regulation of *hid*, causing slow growth of larvae and loss of imaginal discs (Jones and Newbury, 2010).

In this context, having a shorter or longer 3 UTR can regulate the amount and type of *cis* elements regulating a given transcripts: having a longer 3' UTR with more miRNA binding sites may lead to more complex postranscriptional regulation (Alonso, 2012). The *Drosophila* Hox gene*Ultrabithorax* (*Ubx*) provides a good example of this mechanism. *Ubx* mRNA contains two functional PAS in its 3' UTR. The use of the proximal site produces a 'short' UTR that is mostly present at early embryonic stages. At later stages, however, the downstream second PAS is also used, resulting in the production of a "long" 3' UTR which harbours many putative miRNA binding sites. The production of the long 3' UTR makes the transcript "visible" or sensitive to miR-iab4/8 which leads to a reduction in *Ubx* mRNA levels in the posterior segments of the *Drosophila* embryo (Thomsen et al., 2010).

AU-rich elements (AREs) are *cis* elements frequently present in 3' UTR that me-

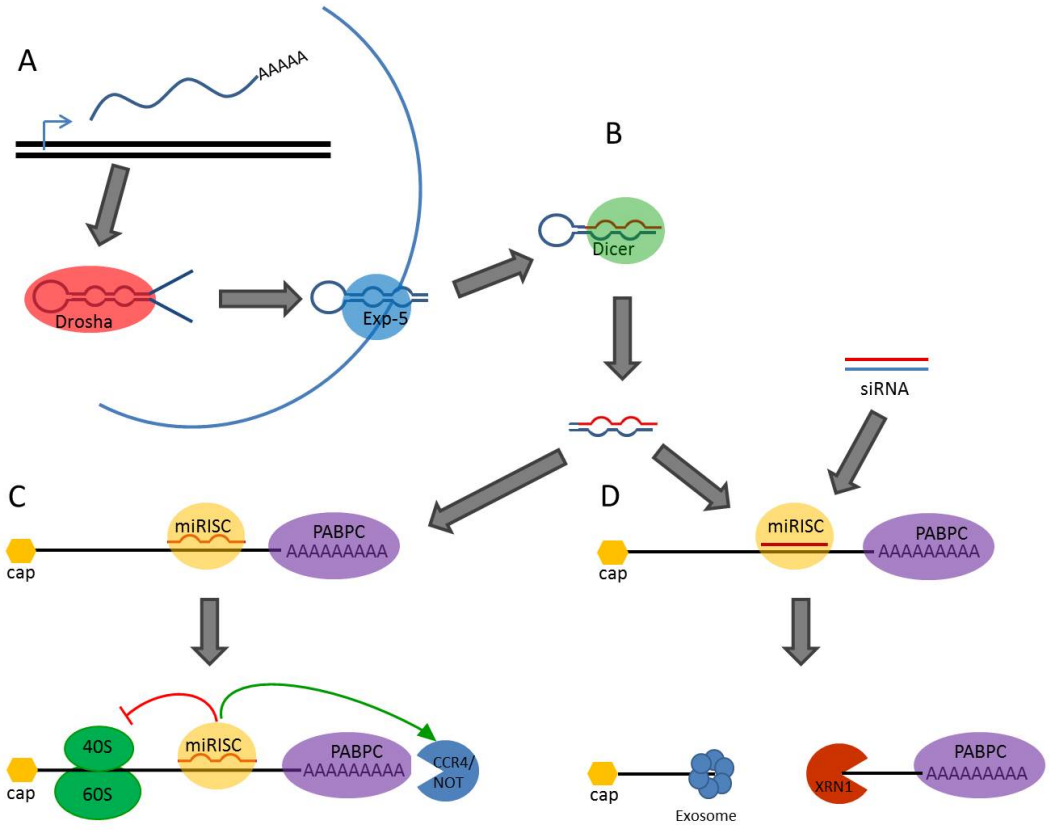


Figure 1.4: microRNA biogenesis and mode of action. A) Although some are encoded within introns from other genes, most microRNAs are transcribed by RNApolII and are capped and polyadenylated (pri-miRNAs). The transcript form secondary structures called hairpins, which are cropped by an RNaseIII enzyme called Drosha (in red), resulting in pre-miRNAs of around 70 nt. The pre-miRNAs are transported to the cytoplasm via Exportin-5 (in blue). B) Once in the cytoplasm, the pre-miRNA is processed by Dicer (in green) to produce the miRNA duplex, of 21-25 nt. The RISC (RNA induced silencing complex, in yellow) is loaded with the strand containing the miRNA sequence (in red) and is used to identify its mRNA target. The target mRNA harbours microRNA target sites mainly within the 3 UTR, but sometimes along the coding sequence or 5 UTR. C) When the complementarity between the microRNA and target RNA is partial, showing complementarity of at least the seed sequence (nucleotides 2-7 from the 5 end of the miRNA) silencing of the target RNA can occur mainly through blocking translation or recruiting RNA degradation machinery (CCR4/NOT deadenylases). D) In plants and some cases in animals, miRNA have complete complementarity to their targets, leading to the cleavage of the mRNA and subsequent degradation by the exosome and XRN1 RNase. This mechanism has been used as a tool to repress specific mRNAs using artificial dsRNAs (or siRNAs). Figure modified from (He and Hannon, 2004), (Filipowicz et al., 2008) and (Huntzinger and Izaurralde, 2011).

diate mRNA binding protein recognition for a diversity of functions. The most well characterised known function is RNA degradation. In humans, the Tristetraprolin protein binds an ARE element in the tumor necrosis factor-alpha mRNA and promotes its deadenylation and degradation (Lai et al., 2005), (Barreau et al., 2005), (Spasic et al., 2012). AREs can also prevent mRNA degradation mediated by the HuR protein (Barreau et al., 2005), (Simone and Keene, 2013). mRNA localisation is frequently mediated by *cis* elements within the 3' UTR and the use of different PAS can generate transcripts with or without localisation signals, radically affecting the translation of such transcript (Colegrove-Otero et al., 2005), (Di Giammartino et al., 2011).

The use of alternative polyadenylation sites is a widespread phenomenon so that more than 50% of the human genes exhibit more than one polyadenylation isoform, and in many cases, these alternative isoforms are conserved in mouse (Di Giammartino et al., 2011). Global studies have revealed that changes in polyadenylation patterns can be related to cellular processes like proliferation and differentiation. Cells in high proliferation, de-differentiation and in some cancers have general UTR shortening whereas high differentiated cells (like neurons) and during late development show a general lengthening of 3' UTRs (Di Giammartino et al., 2011).

In summary, the establishment of the length of the 3' UTR, by alternative polyadenylation, determining its *cis* elements, is a crucial event that will shape the expression of the gene (Alonso, 2012),(Merritt et al., 2008).

## 1.3 Crosstalk among RNA processing events

More and more evidence now suggests that mRNA processing events are highly interconnected and coupled. (Smith and Valcárcel, 2000), (Maniatis and Reed, 2002); (Licatalosi and Darnell, 2010). Each of these processing events is controlled by the concerted action of a number of *trans*-acting factors that are present in the cell in a tightly controlled stoichiometry. Furthermore, many of the sequence elements can affect more than one RNA processing event, resulting in coordination and interdependence of RNA processing events (Millevoi and Vagner, 2010).

A striking example is the coupling of transcription and mRNA splicing. It has

been shown that transcription and splicing occur simultaneously (Goldstrohm et al., 2001) and that RNApol-II transcription affects the splicing pattern of certain mRNAs as well as 3 end formation and mRNA export (Reed et al., 2010), (Millevoi and Vagner, 2010).

RNApol-II has a central role in the coordination of mRNA processing, where its carboxy-terminal domain (CTD) is key (Steinmetz, 1997), (Shukla and Oberdoerffer, 2012). Some components of the splicing and polyadenylation machinery are able to physically interact with the CTD, making them available immediately during transcription. This plataform may facilitate the interaction between factors afecting different aspects of RNA processing and distant sequences within the transcript. The CTD consists on a repeated sequence of amino acids (YSPTSPS) that can extend up to 50 repetitions. Each unit can be dynamically phosphorylated giving it differential properties. The phosphorylated form has a higher negative charge that might facilitate the interaction and binding of SR proteins and spliceosome components, and actually splicing is enhanced when the CTD is phosphorylated (Steinmetz, 1997). Likewise, two of the core components of the polyadenylation machinery, the CPSF and the CstF (Steinmetz, 1997) interact directly with the CTD to regulate 3' end processing (Phatnani and Greenleaf, 2006). CPSF in turn, interacts with the transcription factor TFIID (Di Giammartino et al., 2011). In humans, histone modification enzymes also can affect splicing, where histone markers facilitate the recognition of splicing *cis*-elements, for example, the H3acetyl mark is recognised by the acetyltransferase GCN5 which helps to recruit U2 snRNP, facilitating intron recognition (Hnilicová et al., 2011), (Luco et al., 2011), (Enroth et al., 2012).

Another model for RNApol-II effect on splicing and particularly alternative splicing is a kinetic model (de la Mata et al., 2003). In this model, when transcription is slow, there is enough time for the recruitment of factors that might recognise weak splice sites, promoting their use, thus enhancing exon inclusion. When transcription is faster, stronger sites are preferably recognised, which enhances exon skipping (Batsché et al., 2006). This model works on the assumption that transcription and splicing occurs simultaneously. However, not all splicing events are co-transcriptional an data suggests that introns located towards the 3' end of the

transcript, are more frequently spliced after transcription has finished (Kornblihtt et al., 2004).

The rate of transcription of the RNApol-II can be affected by recruitment of factors that affect polymerase II speed. The Brahma protein is a component of the SWI/SNF chromatin remodelling complex making genes more accessible for RNApol-II, potentially enhancing its processivity. The SWI/SNF complex has also a role in transcription elongation, having direct contacts with the transcription machinery and with spliceosome components. Brahma, however, has a chromatin independent role in slowing the RNApol-II during transcription elongation allowing exon inclusion (Batsché et al., 2006), (Luco et al., 2011).

The speed of RNApol-II also affects 3' end formation. In the *polo* transcript, a slow elongating polymerase allows enough time for the assembly of the polyadenylation machinery in the first PAS, facilitating its use whereas a fast polymerase would transcribe fast enough to expose both PAS to the polyadenylation machinery, promoting the use of distal PAS (Pinto et al., 2011).

All of these lines of evidence points to a central role of RNApol-II in controling and modulating almost every aspect of pre-RNA processing, from transcription to splicing and 3' end formation. It is likely that the polymerase functions as a platform for the factors that will model the RNA processing within the very packed and complex cellular environment, saving time and energy and avoiding mistakes.

Some protein factors also link mRNA processing events independently of RNApol-II. For example, several splicing factors also interact with the polyadenylation machinery providing a link between splicing and polyadenylation: U2AF65 recognises and binds the polypyrimidine tract of the last intron and recruits the CFI complex promoting the use of the nearest PAS (Millevoi, 2009) (Millevoi and Vagner, 2010),(Di Giammartino et al., 2011).

## 1.4 RNA processing during development

During the embryonic development of higher eukaryotes, gene expression and RNA processing are highly dynamic and controlled. *Trans* acting factors including RNA binding proteins (RBP), play a key role in this process. RBPs are involved in every

stage of gene transcription, processing, translation and degradation (Gamberi et al., 2006).

RBPs are proteins that often contain a domain for RNA interaction, such as RNA Recognition Motif (RRM), hnR NPK Homology (KH) domain (Gamberi et al., 2006).

RBPs are important in the control of processing of transcripts during cell differentiation and many of them are tissue specific, (Kalsotra and Cooper, 2011), (Taliaferro et al., 2011). The result of this control is the establishment of regulatory networks that control specific cellular functions, differentiation or development (Kalsotra and Cooper, 2011). For example, during the transition of mouse pluripotent embryonic stem cells to embryoid body, near 100 alternative splicing events occur (Salomonis et al., 2010). Some of these alternative splicing events modify important domains of proteins involved in the maintenance of pluripotency, progression of the cell cycle and lineage specification (Salomonis et al., 2010). For example, the human Forkhead box (FOX) transcription factor 1 is important in the establishment of cell types of cardiomyocites during heart development (Zhang et al., 2010). During cell differentiation, a cassette exon shows increased inclusion wich modifies the FOX1 DNA binding specificity (Gabut et al., 2011).

Genome-wide studies have revealed that most of the transcripts in *Drosophila* experience alternative polyadenylation and that there are tissue-specific trends (Smibert et al., 2012). For example, a general shortening of 3' UTRs is observed in testis and a general lengthening in the nervous tissue (Smibert et al., 2012). Other studies have demonstrated that global 3' UTR shortening is associated with cellular proliferation, which may be mediated by increased levels of the polyadenylation machinery (Elkon et al., 2012). In several mammalian cancers, where the generation of shorter 3' UTRs leads to the loss of micro-RNA regulatory sequences, there is increased mRNA stability and increased protein production (Flynt and Lai, 2008), (Mayr and Bartel, 2009), (Bava et al., 2013), (Li and Lu, 2013). For example, the 3' UTR of the human transcription factor MYB contains conserved miR-15a/16 and miR-150 target sites. The loss of its 3' UTR leads to overexpression and protein production of MYB (Persson et al., 2009).

The nervous tissue is a complex system that requires precise control of gene ex-

pression to allow proper cell differentiation, neuron morphology and wiring of neuronal circuits (Norris and Calarco, 2012). To achieve this, neuronal transcriptome is enriched with regulatory motifs and trans-acting factors that regulate transcription and post transcriptional processing generating a very high molecular diversity (Norris and Calarco, 2012).

In neuronal tissue, alternative splicing events can control the presence or absence of protein domains, this allows the control of the presence or position of protein interaction surfaces, allowing the formation of protein complexes of diverse functions. Human neuronal specific cassette exons have been identified in many proteins important for endocytosis functions (Ellis et al., 2012).

Understanding how mRNA processing is regulated during development sheds light on the understanding how different developmental pathways are controlled in a dynamic context. Recent studies have demonstrated that global mRNA polyadenylation pattern changes with developmental progression. This has been demonstrated in mouse embryos where mouse genes tend to express mRNAs with longer 3 UTRs as embryonic development progresses (Ji et al., 2009).

### 1.4.1 Elav/HuR family

Some proteins can affect both alternative splicing and alternative polyadenylation, however, little is known about the molecular mechanism mediating the crosstalk between different RNA processing events. A protein factor proved to affect alternative splicing and polyadenylation is the Elav protein (embryonic lethal abnormal vision) (Yao and Samson, 1992), (Sakai et al., 1994), (Koushika et al., 1996). The *elav* gene is located in the X chromosome and produces a 50 kD protein of 483 amino acids that contains an A/Q rich amino terminal region and three RRMs (RNA recognition motif). The *elav* transcript contains two introns and three exons, one of the introns is located in the 5' UTR (Yao and White, 1991), (Yao and Samson, 1992). Elav protein was the first neural-specific RBP described in *Drosophila melanogaster* and expressed in the nucleus of post-mitotic neurons throughout all *Drosophila* stages (Figure 1. 5).

In humans, there are four Elav-like proteins, HuR (human antigen R) (the most similar to Elav), HuB, HuC and HuD. HuR is ubiquitously expressed whereas HuB,

C and D are expressed solely in the CNS. HuR proteins have been described to bind AREs via RRM domains. Elav-like proteins contain 3 RRM which are very conserved (Antic and Keene, 1997), (Pascale and Govoni, 2012).

In humans, HuR proteins are able to increase the stability of their target mRNA in the cytoplasm, although in some cases, they are able to decrease their translation (Pascale and Govoni, 2012). HuD also plays a role in translation initiation, as a protein needed in mRNA circularisation (Fukao et al., 2009). The third RRM domain has polyA binding activity whereas the hinge region recognises eIF4A protein, helping in the circularisation of the mRNA. However, it is unclear how widespread this role in translation is (Fukao et al., 2009).

Other two members of this family in *Drosophila* have been described: RBP9 and FNE (found in neurons). RBP9 shares 52% identity with Elav whereas FNE shares 61%. RBP9 is expressed in neurons after mid-pupal stages and in oocytes but not in embryonic stages (Yao and Samson, 1992), (Samson and Chalvet, 2003). FNE has a similar expression pattern to Elav, although its expression in embryos is delayed, and FNE protein is present in the cytoplasm. Interestingly, FNE and RBP9 interact (Samson and Chalvet, 2003). Elav protein has served as a very important marker in the study of the *Drosophila* nervous system and differentiation of neurons during development. However, little is known about the role of the Elav protein itself in neuron determination and survival. Indeed, this work together with that of colleagues in the the Alonso Lab and our collaborators at the University of Mainz adds to this open question (see below).

During *Drosophila* embryonic CNS development, some neurons project their axons across the midline and contact their targets in the other side (Marillat et al., 2004). These axons are called "commissural" axons. Elav is important in directing these axons to cross the midline and form "commissures". *elav* mutant embryos have morphological defects in the nervous tissue, the commissures are thinner and their pattern seems affected, particularly towards the posterior region. (Jimenez and Campos-Ortega, 1987), (Simionato et al., 2007). The *elav* mutant does not induce cell death but affects the ability of the axons to extend and cross the midline. Neurons differentiate normally but the axonal projections are not directed properly. A closer study of the molecular mechanisms for this shows that Elav interacts with

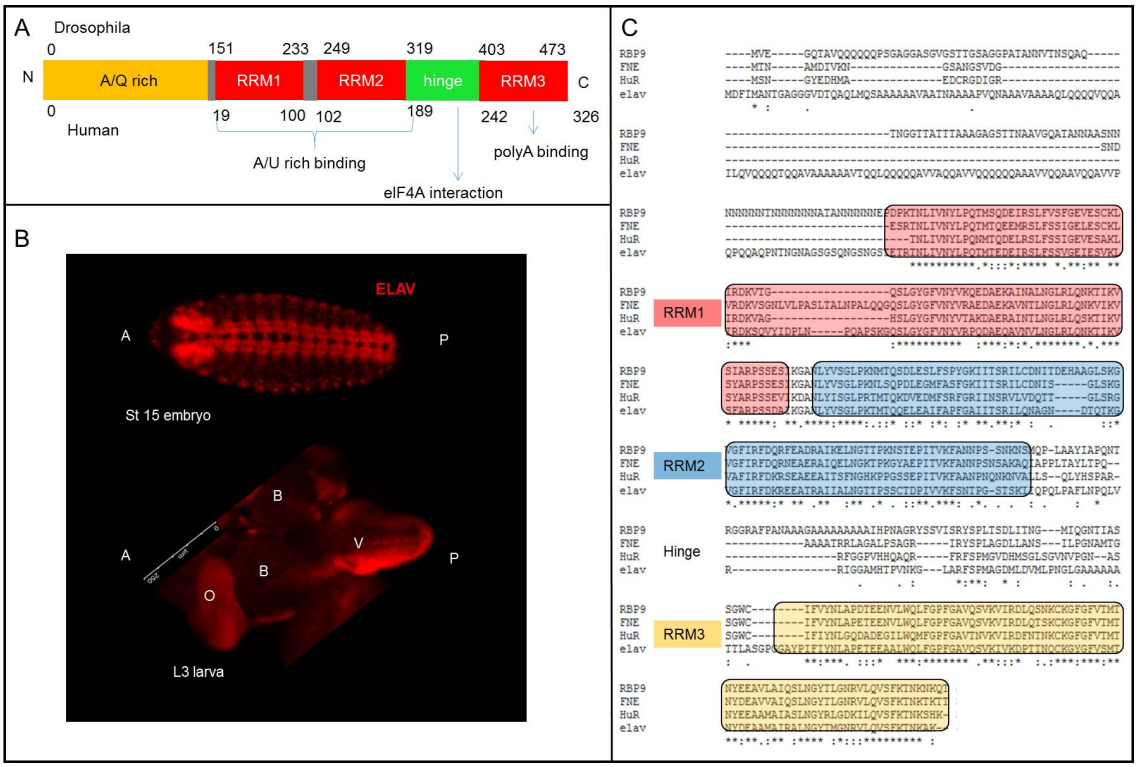


Figure 1.5: Elav protein structure and expression. A) Elav protein domain organisation, amino terminal A/Q rich domain and the three RRM domains, of which the last one is separated by a hinge. The RRM1 and 2 bind A/U rich RNA, and in humans the third RRM interacts with the polyA whereas the hinge interacts with eIF4A, promoting mRNA circularisation and translation. The numbers represent the aminoacid position in *Drosophila* (above) and Human (below). B) Elav immunostaining showing protein expression at embryonic stage 15 and CNS at larval stage 3. A, anterior. P, posterior. C) Aminoacid alignment of *Drosophila* elav, RBP9, FNE and human HuR proteins. Boxes indicate the RRM domains.

the mRNA of *commisureless* (*comm*), which is a member of the Robo/Slit signalling pathway that guides axons to cross the midline (Marillat et al., 2004), (Simionato et al., 2007). *comm* mRNA and protein levels are reduced in the *elav* mutant. However, the precise molecular mechanism of Elav controling *comm* mRNA is unclear. Other studies trying to elucidate the molecular function of Elav, have shown that it is involved mainly in post transcriptional processing, particularly splicing and polyadenylation of *armadillo* (*arm*), *Neuroglian* (*ng*), *erect wing* (*ewg*) and itself (Borgeson and Samson, 2005), (Soller et al., 2008), (Soller et al., 2010).

*armadillo* encodes the *Drosophila* homolog of the vertebrate beta catenin, which is an adhesion molecule and transcriptional regulator of Wnt/wingless pathway. Elav stimulates inclusion of exon 6 resulting in a neuronal specific *arm* mRNA isoform (Loureiro and Peifer, 1998), (Koushika et al., 2000), (Soller et al., 2008).

*Neuroglian* (*ng*)is a gene that encodes two protein isoforms by alternative splicing, an ubiquitously expressed 167 kDa peptide and a neuronal 180 kDa one. The presence of Elav in neuronal cells directs the production of the neuronal isoform. Ectopic expression of Elav produces the neuronal isoform of *ng* (Koushika et al., 1996), (Koushika et al., 2000). Elav binds a U-rich region in the last intron, either affecting splicing directly or preventing the use of the closely located PAS (Lisbin et al., 2001).

The best characterised Elav regulated mRNA processing event is the *erect wing* (*ewg*) mRNA, which is a neuronal transcription factor involved in CNS development and formation of flight related musculature (Fleming et al., 1983), (Koushika et al., 1999). *ewg* mRNA is ubiquitously transcribed and undergoes alternative processing in a tissue-specific manner. The neuronal isoform is polyadenylated in the downstream polyadenylation sites located in the 3' UTR (Arrows in Figure 1.6A). This transcript leads to full length active protein (Koushika et al., 1999). However, the non-neural mRNA isoforms are polyadenylated within intron 6, loosing the downstream exon (Figure 1.6A). The transcripts lacking the last coding exon are not translated into proteins. Elav has been demonstrated to be the responsible factor directing the use of the polyadenylation site located within the 3' UTR and in turn, promoting the retention of the last exon (Soller and White, 2003), (Soller and White, 2005) (Figure 1.6B). The Ewg protein is a key transcription factor during

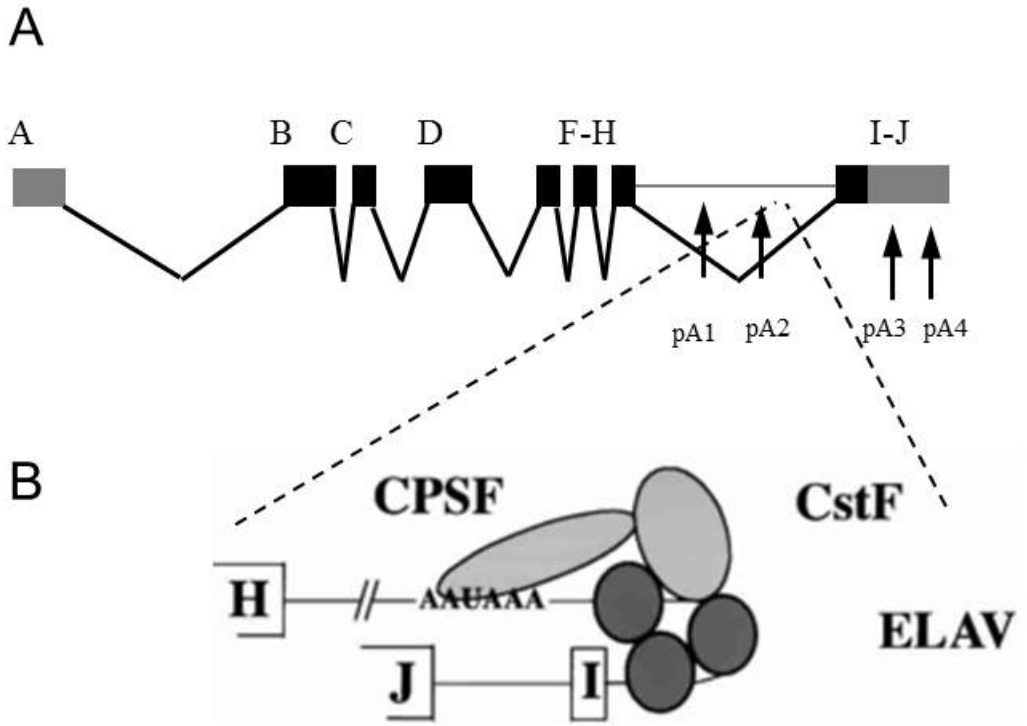


Figure 1.6: *ewg* polyadenylation is regulated by Elav. A) Exon and intron organisation of the *ewg* transcript, grey boxes represent non coding exons and black boxes represent coding exons. Arrows mark polyadenylation sites (pA). B) Model of the mechanism of Elav role in *ewg* polyarenylation site choice, binding directly as a multimer to a sequence downstream pA2, interacting with CstF to block cleavage on this site. Figure taken from (Soller and White, 2003).

the development of the *Drosphila* CNS and EWG protrein expression is sufficient to rescue an *elav* mutant. This suggests that ewg is a major direct target of Elav (Haussmann et al., 2008).

Elav also regulates polyadenylation site selection of its own mRNA. A screen using a modified protocol of *in vitro* selection found several Elav binding sites, the strongest of which maps to a non coding exon in the 3' end of *elav* mRNA (Borgeson and Samson, 2005).

RNA-seq studies have demonstrated that after the maternal-to-zygotic transition (MZT) in *Drosophila* Elav coordinates the the use of more distal polyadenylation sites of a set of genes in the CNS, generating transcripts with longer 3' UTR (Hilgers et al., 2011), (Hilgers et al., 2012), (Smibert et al., 2012). Not all of these transcripts are neuronal, but the observed lengthening of 3'UTRs is restricted to the CNS

and there is a high proportion of RNA binding proteins displaying longer UTRs (Smibert et al., 2012). Interestingly, some of them show a sequential lengthening of the 3'UTRs, presumably using a more distal polyadenylation site as development progresses (Hilgers et al., 2011). *elav* transcript itself shows this behavior along with *fne*, *Brain tumor (brat)*, *mei-P26* and *argonaute1 (ago1)*. These long 3' UTRs are enriched in regulatory elements, compared to the corresponding short sequences, including miRNA binding sites, polyadenylation signals and protein binding sites, presumably for a more complex regulation during CNS development. However, the effect on protein production have not yet been studied (Hilgers et al., 2011), (Smibert et al., 2012). Elav ectopic expression is sufficient to promote the lengthening of the 3' UTR outside neuronal tissue (Hilgers et al., 2012). RNA-IP experiments revealed that Elav binds sites near polyadenylation signals, suggesting that Elav acts in a similar way to the mechanism described in the *ewg* transcript (Soller and White, 2003), (Hilgers et al., 2012).

*Ubx* and other HOX genes also undergo 3' lengthening at late stages of development (Thomsen et al., 2010) but the mechanism for this remains unknown.

In *ewg* and *neuroglian* mRNA, the effect of Elav on alternative splicing and polyadenylation come together because Elav influences polyadenylation site selection between a proximal (within intron 6) and distal site (within 3' UTR). In the absence of Elav, the intronic PAS is used, so the polyadenylation process cleaves the RNA within the intron, losing the downstream coding exon. When Elav is present, the PAS in the intron is blocked, promoting the use of a downstream PAS located in the 3' UTR and in consequence, the last coding exon is retained. However, Elav also affects splicing independently of polyadenylation, as described for other mRNAS, for example, in the *armadillo* (Koushika et al., 2000) and *fas* mRNAs (Izquierdo, 2010).

It remains unclear the mechanisms by which Elav can affect both splicing and polyadenylation in the same transcript when the PAS sites are within the 3' UTR and not in an intron.

## 1.5 Hox genes

During *Drosophila* embryonic development, a key event is the determination of segment identities along the body axis. This allows the development of body structures in their correct position. One gene family involved in this process is represented by the Hox genes (Figure 1.7). The Hox genes encode evolutionary conserved transcriptional regulators that trigger different developmental programs at specific spatial coordinates along the antero-posterior body axis of animals (McGinnis and Krumlauf, 1992), (Alonso, 2002). In *Drosophila*, the Hox genes have a peculiar orderly arrangement along chromosome 3 that shows a strong correlation (termed colinearity) with the order of expression within the *Drosophila* body axis (Hueber et al., 2010) (Figure 1.7).

Mutations affecting Hox gene activity often result in major morphological alterations; for example homeotic transformations observed in *Drosophila* i.e. Antennapedia and Bithorax phenotypes(Garcia-Bellido et al., 1976), (Lewis, 1978), (Hughes and Kaufman, 2002) (Figure 1.8). In spite of their general importance, many aspects of Hox gene function remain unclear. At early stages of *Drosophila* embryonic development, cells within each segment acquire positional information by the combination of the expression of several genes deployed in a cascade. These genes encode for transcription factors and are expressed in the following order: maternal, gap and pair-rule. At this stage, the embryo is divided in 14 segments. The Hox genes are expressed depending on the relative expression of those genes along the AP axis and then they direct the morphogenetic complexity within each segment (Maeda and Karch, 2006), (Lewis, 1978). However, it is not clear how Hox gene activity is regulated at the cellular level during the progressive development of segments. Regulation of Hox gene transcriptional patterns, RNA processing and microRNA contribute to the establishment of segmental cellular identity (Bender et al., 1983) (Ronshaugen et al., 2005), (Bushati et al., 2008), (Bender, 2008).

It has recently been shown that alternative splicing (AS) plays an important role in Hox specificity. Specific splice isoforms of the *Drosophila* Hox gene *Ubx* have distinct DNA binding specificities. The different isoforms activate transcription with different efficiency genes in the embryonic mesoderm and have different ways to pattern muscles in *Drosophila* embryos (Reed et al., 2010), (de Navas et al., 2011)

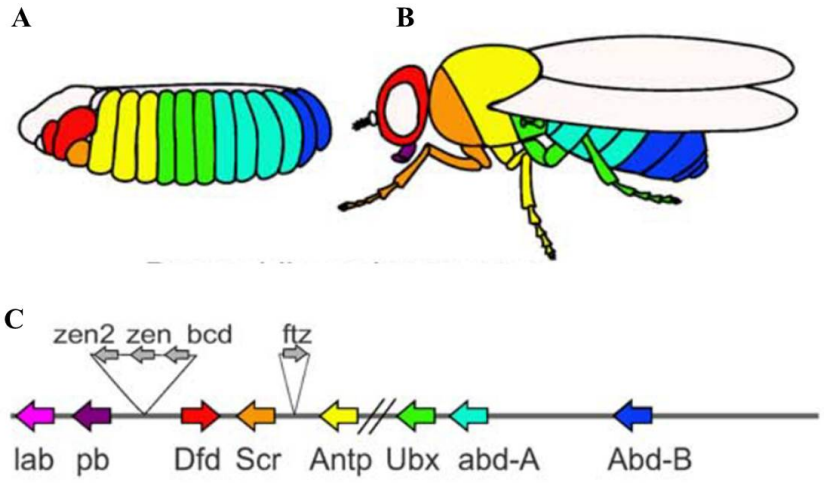


Figure 1.7: The Hox genes. A) *Drosophila* embryo showing body segments, the colours represent the domains of expression of the different hox genes, shown in C. B) *Drosophila* adult showing the expression domains of Hox genes. C) Representation of the chromosomal position of the Hox genes. lab: labial, pb: proboscipedia, Dfd: deformed, Scr: sex combs reduced, Ant: antennapedia, *Ubx*: *Ultrabithorax*, abd-A: abdominal-A, abd-B: abdominal B. Figures taken from (Hueber et al., 2010).

(Figure 1.8).

### 1.5.1 *Ultrabithorax*

*Drosophila* Ubx protein is a transcription factor and is the product of one of the three Hox genes in the bithorax complex. *Ubx* guides the development of embryonic parasegments 5 and 6, which correspond to the third thoracic and first abdominal segments (Akam and Martinez-arias, 1985), (White and Wilcox, 1985) (Figure 1.8B). During adult *Drosophila* development *Ubx* controls the differences between wings and halteres (the small dorsal appendages in the posterior thorax) (Garcia-Bellido et al., 1976), (Lewis, 1978). Ubx protein is expressed in the haltere throughout development (White and Wilcox, 1985) and its absence leads to a complete homeotic transformation of the haltere into a wing (Figure 1.8C) (Bender et al., 1983). The Ubx protein is also responsible for the patterning of segments within musculature and the central nervous system (CNS) (Kornfeld et al., 1989) and it is expressed in the CNS, mesoderm and epidermis at late embryonic stages (Figure 1.8B).

The *Ubx* transcript produces several protein isoforms through AS which differ from one another by the presence of the two optional microexons (mI and mII)

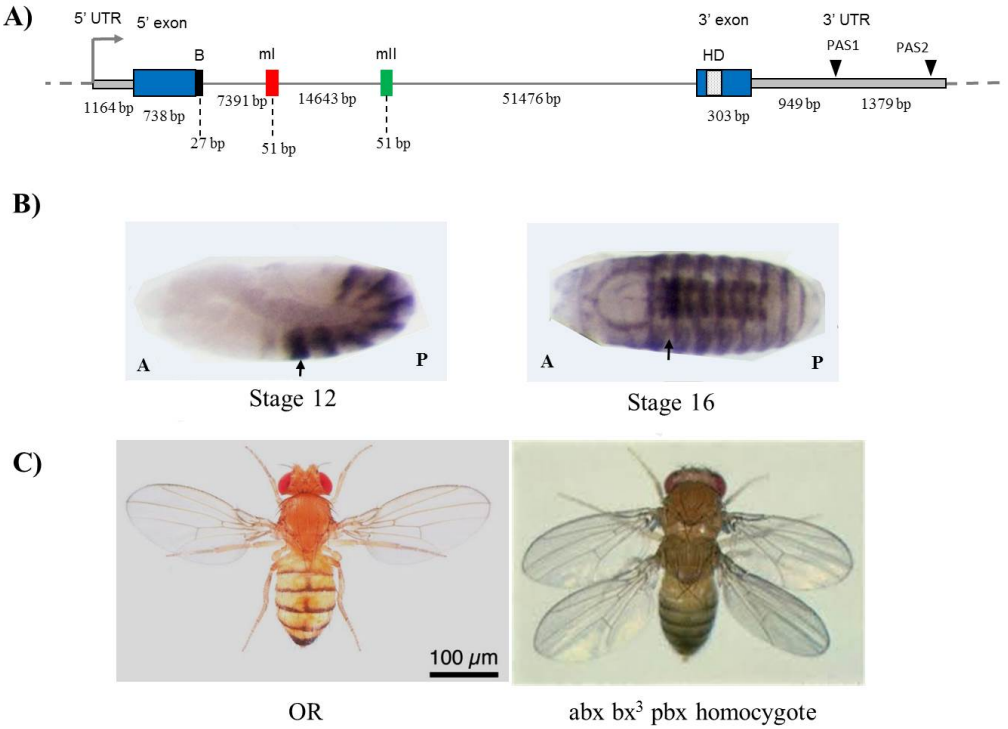


Figure 1.8: The *Ultrabithorax* gene. A) Representation of the *Ubx* gene. Boxes represent coding exons, constitutive exons in blue, microexons I and II in red and green respectively. Gray bars represent 5' and 3' UTRs and lines represent introns. The numbers represent the sizes of each element in bp. Arrowheads represent the sites of polyadenylation. B) Enzimatic anti-Ubx immunostaining at stage 12 and 16 of embryonic development showing high Ubx expression in the CNS and in epidermis.A: anterior, P: posterior. C) Wild type adult fly showing normal wings and halteres. abx bx3 pbx homocygote showing a very strong haltere to wing transformation. Adult flies pictures taken from (Lewis, 1998).

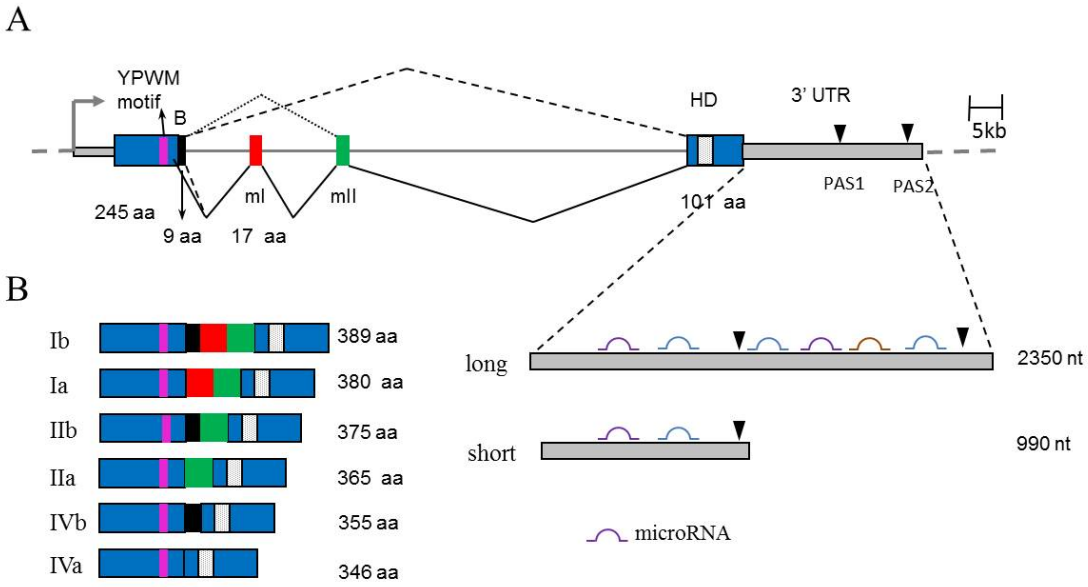


Figure 1.9: RNA processing of *Ultrabithorax*. A) Scheme representing the transcript organisation. Boxes represent coding exons, constitutive exons in blue, microexons I and II in red and green respectively. Gray bars represent 5' and 3' UTRs and lines represent introns. Dotted lines represent the alternative splicing routes. B) Splicing and polyadenylation isoforms produced. The scheme on the left represent the six isoforms produced by the inclusion or exclusion of the alternative exons. On the right the two polyadenylation isoforms is shown. Arrowheads mark position of cleavage and polyadenylation.

and by the use of a proximal or distal donor splice site in the 5' exon (B element). Alternative splicing then produces six protein isoforms: Ib (B element, mI and mII), Ia (mI and mII), IIb (B element and mII), Ia (mII), IVb (B element, no microexons) and IVa (no microexons) (Figure 1.9B).

The difference between the protein isoforms produced is relatively small, so elucidating if they are functionally different has proven to be a challenge. Initial research produced little evidence for a differential role of each protein isoform: a chromosomal inversion around mII that produces only the IV isoform (Ubx $^{MX17}$ does not show any morphological difference at embryonic or larval stages to wild type flies (Bisturia et al., 1990). This result suggests that the Ubx IV isoform is able to successfully substitute *Ubx* normal expression at those stages, in the epidermal tissue. Moreover, over expressing individual *Ubx* isoforms under a heat shock promoter (hsp 70) leads to a similar thoracic to abdominal transformation in the ventral cuticle in instar 1 larvae (L1) suggesting the interchangeability of the isoforms (Mann and Hogness, 1990). However, the same hsp 70 promoter-Ubx constructs show a different effect of Ia or IVa Ubx isoforms in the peripheral nervous system: Ia isoform overexpression leads to an abdominal transformation of PNS at parasegments 3, 4 and 5 whereas IVa isoform does not, suggesting a differential role (Mann and Hogness, 1990). New evidence has shown that the 3' exon contains the homeodomain (HD) and the 5' exon contains the YPWM motif (Figure 1.9) which plays an important role in the protein interaction between Ubx and Extradenticle, enhancing Ubx DNA specificity and binding. The inclusion or exclusion modifies the distance of these two domains in the Ubx protein, potentially changing its properties (Passner et al., 1999),(Reed et al., 2010) Figure 1.9A and 1.10).

In S2 cells, *Ubx* isoforms IIa and IVa show a slightly stronger ability to activate an *Ubx* target reporter (Krasnow et al., 1989). Recent studies have confirmed that protein isoforms of Ubx behave differentially *in vitro* and *in vivo*. An *in vivo* experiment using a *Ubx* mutant that produced a strong haltere-to-wing transformation was rescued by the expression of specific splicing isoforms of *Ubx* and the results showed that Ia isoform has a higher ability to rescue the haltere phenotype than the IVa isoform (Figure 1.10 C-H) (Reed et al., 2010), (de Navas et al., 2011).

*Ubx* mRNA is also alternatively polyadenylated. Two polyadenylation sites have

been characterised. The use of the first PAS (proximal) produces a 3.2 kb transcript that appears at 3 hAEL and the use of the second PAS (distal) 4.3 kb one that appears at 6 hAEL and peaks at 12-15 hAEL (O'Connor et al., 1988), (Kornfeld et al., 1989). Bender's group managed to isolate neuroblasts and culture them for 18 h and found that they produced more long UTR isoforms compared to isolated myoblasts. Splicing isoform analysis in the same cells revealed that the I isoforms are more abundant in the myoblasts and the IV isoforms in the neuroblasts, suggesting a tissue specific distribution of the isoforms rather than segment specific (O'Connor et al., 1988).

*Ubx* shows a correlation between splicing and polyadenylation patterns: *Ubx* isoforms retaining both microexons (i.e. *Ubx* I) are more associated with shorter 3 UTR tails determined by a proximal pA (S), while splicing forms in which the microexons are spliced out (i.e. *Ubx*-II and *Ubx*- IV) possess longer 3 UTR sequences (L) (O'Connor et al., 1988), (Kornfeld et al., 1989). This pattern of splicing and polyadenylation isoforms is conserved among distant Drosophilids, suggesting an important function for development (Bomze and Lopez, 1994).

*Ubx* expression levels can be also regulated posttranscriptionally by the miR-iab4/iab8 microRNAs (miRNAs) acting on specific target sites located in *Ubx* distal 3 UTR (Bender et al., 1983), (Ronshaugen et al., 2005), (Bender, 2008),(Bushati et al., 2008), (Tyler et al., 2008). Bioinformatic studies revealed that most of these miRNA target sequences are positioned within the *Ubx* 3 UTR tail precisely in the region that is affected by the alternative pA process described above (Patraquim et al., 2011)(Figure 1.11) and that they are evolutionarily conserved. Altogether, these observations suggest that *Ubx* AS, pA and miRNA-regulated *Ubx* expression levels may be functionally interlinked during development. This possibility would concur with the new molecular views of gene expression as an extensively coupled network instead of a sequential assembly line (Maniatis and Reed, 2002). Recent work highlights the relevance of these views in the context of animal development and its evolutionary variation (Alonso and Akam, 2003), (Alonso and Wilkins, 2005).

Proper identity establishment of segments during development is crucial for the correct diversification of neuronal circuits that run along the anterior-posterior axis. This ensures the integration of sensory signals coming from the environment,

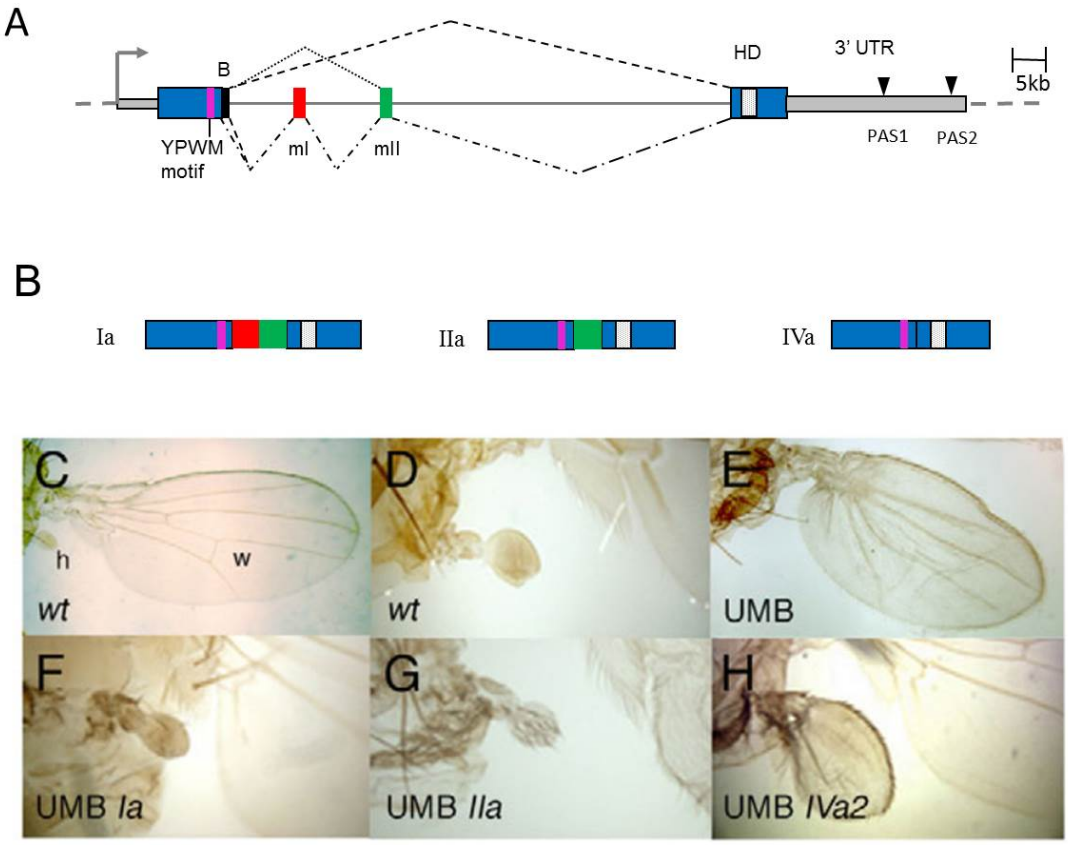


Figure 1.10: Alternative splicing of *Ultrabithorax*. A) Representation of the *Ubx* transcript showing the constitutive exons in black, the 5' exon contains the YPWM motif shown as a blue bar, the 3' exon contains the homeodomain, represented with a white bar (HD). The gray bars represent the alternative microexons mI and II. 5'and 3' UTRs are shown as thick gray lines. The arrow at the 5' end marks the position of transcription initiation. B) The *Ubx* isoforms produced by the alternative inclusion of mI and mII only. C-H) Rescue experiment of an *Ubx* mutant using different Ubx protein isoforms. C) Wild type fly showing normal wing and haltere. D) Higher magnification picture showing the normal wild type haltere. E) *Ultrabithorax* mutant background (UMB) used for the rescue experiments, shows a haltere to wing transformation, picture at the same magnification as D. F) UMB fly over expressing the Ia isoform showing a total rescue of the wing phenotype. G) UMB fly over expressing the IIa isoform showing a less strong rescue of the wing phenotype. H) UMB fly over expressing the IVa isoform showing a very weak rescue of the wing phenotype. Pictures taken from (de Navas et al., 2011).

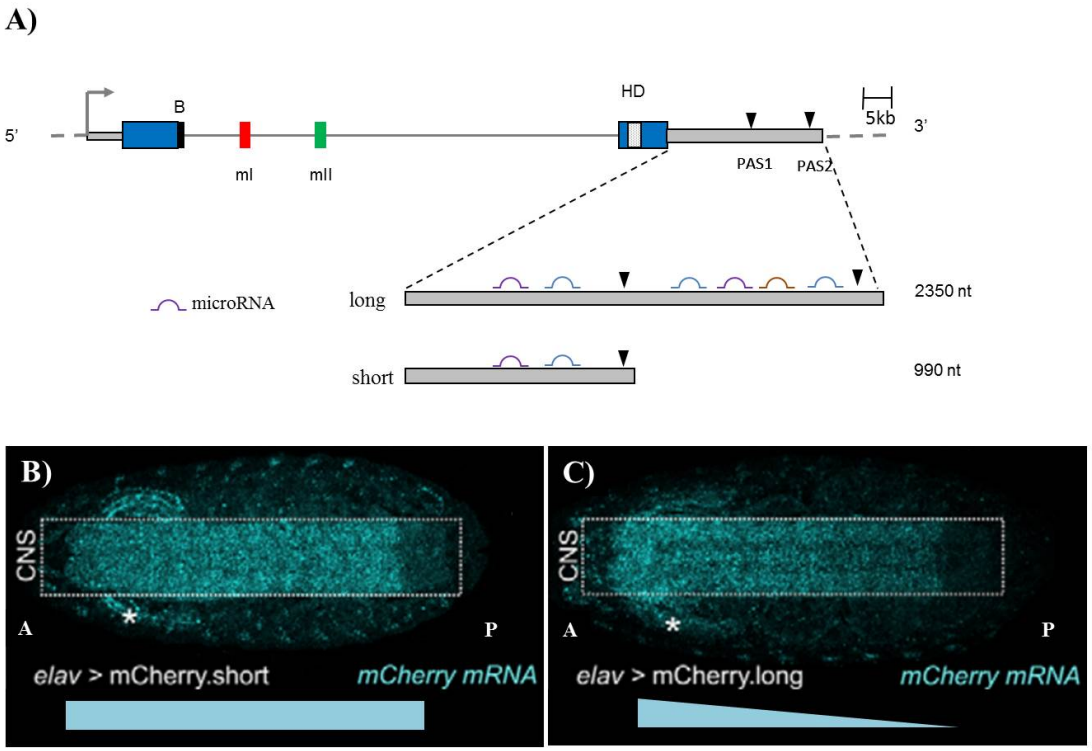


Figure 1.11: The longer 3' UTR from *Ubx* contains regulatory elements/ A) Scheme representing the alternative polyadenylation of *Ubx* resulting in a long 3' UTR of 2350 nt or a shorter 3' UTR of 990 nt. The longer portion of the 3' UTR contains more putative miRNA binding sites. B) mCherry fused to the short 3' UTR of *Ubx* shows a uniform expression when an *elav* driver is used. C) When the long 3' UTR is used, a down regulation is observed towards the posterior region of the embryo. Blue figures at the bottom represent the construct expression profile. A, anterior; P, posterior; asterisk marks unspecific signal from the salivary glands. Taken from (Thomsen et al., 2010).

needed for locomotion and behavioral decisions. Each hemisegment produces 30 neuronal stem cells or neuroblasts (NB) that reads the signals from the neuroectoderm to acquire individual lineage identities. These identities are similar but still have segment-specific identity and are called serial homologs. This fine tuned identity is achieved by neuroblast specification, differential cell death and proliferation (Rogulja-ortmann and Technau, 2008), (Rogulja-Ortmann et al., 2008), (Estacio-Gómez et al., 2013), (Dixit et al., 2013). Given the role of Hox genes in determining anteroposterior axis it is likely that they are responsible for the establishment of neuroblast identity in each segment. For example, the NB1-1 lineage is specified by *Ubx* and Abdominal A and it produces neurons and glial cells in abdominal segments but in thoracic segments, it only produces neurons. In abdominal segments, *Ubx* is needed for the suppression of the thoracic NB6-4 lineage. *Ubx* also activates segment specific cell death, eliminating the GW neuron fate from lineage NB7-3 in T3 and abdominal segments (Rogulja-Ortmann et al., 2008).

*Ubx* offers a very good tool to explore what factors are involved in the control of alternative splicing and polyadenylation during embryonic development of *Drosophila*.

## 1.6 Aims of the study

Proper RNA processing is crucial for the correct expression and function of genes. However, little is known about the precise mechanisms that underlie the production of specific splicing and polyadenylation isoforms during development. The aim of this study was to increase our understanding of the factors involved and their mechanisms in the establishment of mRNA splicing and polyadenylation isoform patterns during embryonic development. This project used the *Ubx* gene to explore the role of some candidate factors in the control of alternative splicing and polyadenylation and describes the direct interaction between *Ubx* mRNA and one of these proteins (Elav), demonstrating that the mechanism for its control involves a direct contact within intronic sequences.

Elav plays important roles in the mRNA processing of other mRNA targets (*ewg* and *Neuroglian*) in the CNS during embryonic and larval stages. Its expression

partially overlaps *Ubx* expression in the CNS at embryonic stages, making them likely to interact, so the first aim was to test if they do. I first tested whether Elav affects the mRNA processing of *Ubx* by RT-PCR experiments using RNA extracted from whole embryo *elav* null mutants. I asked if Elav affects the splice site choice between the alternative exons changing the balance of isoforms produced. I asked if Elav plays a role in polyadenylation site selection as well, using the same experimental strategy. I also explored other possible effects of the *elav* null mutant in the *Ubx* gene, such as mRNA abundance and protein production.

Once the interaction between Elav and *Ubx* mRNA was established, I sought to determine the nature of such interaction. For this, I asked wether Elav is able to bind directly to elements within the *Ubx* mRNA *in vitro*. For this, a bioinformatic search for potential Elav binding sequences was carried out by Pedro Patraquim, member of the Alonso Lab.

A third aim was to extend the search of candidate factors that also contribute to the finely controlled processing of *Ubx* mRNA during embryonic stages of development. For this we asked whether selected candidate factors affect the splicing and polyadenylation patterns of *Ubx* mRNA, using the same experimental strategy used for the Elav factor.

# Chapter 2

# Materials and Methods

## 2.1 *Drosophila melanogaster* culture

Drosophila *melanogaster* flies were kept in plastic vials or bottles with standard molasses food, at 25°C or room temperature. Oregon Red was used as wild type in all the experiments.

### 2.1.1 Collection of embryos

Apple juice agar plates were used to collect embryos from collection cages. 5 cm diameter plates were used with small collection cages and 9 cm diameter plates with big collection cages. For 500 ml of apple juice agar, 12 g of agar, 12 g of sugar, 125 m of apple juice were combined. Agar was melted in a microwave and poured in the plates. Live yeast paste was spread on the plate and covered with a piece of tissue before use.

## 2.2 Embryo dechorionation and fixation

Embryos were collected in a small funnel net and yeast was rinsed with tap water. Dechorionation was done with 50% bleach for 5 minutes. Embryos were rinsed with tap water. Embryos for antibody staining were fixed using 0.66 mL 4% (v/v) paraformaldehyde in PBS and 2 mL of heptane in vigorous agitation for 20 minutes. Lower phase was discarded and 1.5 ml methanol was added, and vortexed for one minute. All liquid was removed and embryos were rinsed with 1 ml methanol four

times and stored at -20°C in methanol.

## 2.3 Whole embryo antibody staining

Fixed embryos in methanol were rinsed with PBT (0.1% (v/v) Triton X-100 PBS) and washed 3 times for 10 minutes. Primary antibody was incubated overnight at 4°C in gentle agitation (anti-*Ubx* diluted in PBT 1:20, anti-GFP 1: 300, rat anti-Elav 1:300). Embryos were rinsed 3 times with 1 ml PBT and washed 3 times for 10 minutes. Secondary antibody was incubated for 2 hours at room temperature (anti mouse 1:1000, anti rabbit 1:1000). DAPI was added at 1:200 for the last 10 minutes of incubation. Embryos were rinsed 3 times in PBT and washed 3 times for 20 minutes, then rinsed 2 times in PBS, all liquid was removed and 70% Glycerol in PBS was added. Embryos were mounted in 70% (v/v) Glycerol in PBS on poly-lysine slides.

## 2.4 RNA extraction

RNA was extracted using the TRIzol® reagent (Invitrogen Cat#15596-026). At least 50 dechorionated embryos were homogenized and lysed using a sterile pestle in a 1.5 ml eppendorf tube containing 100 $\mu$l of TRIzol® reagent. After homogenization, volume was adjusted to 500 $\mu$l or 1000 $\mu$l depending on the number of embryos. 200 $\mu$l of chloroform per ml of TRIzol® were added and vortexed for 15 seconds.The tubes were incubated at room temperature for 5 minutes and then centrifuged to separate phases. The resulting supernatant was transferred to a sterile, nuclease-free tube for precipitation using 500 $\mu$l of isopropanol per ml of initial TRIzol®. The RNA was precipitated by centrifugation at 4°C for 30 min at top speed in a top table centrifuge. The supernatant was decanted and the pellet dried. RNA was resuspended in 20-50 $\mu$l of milliQ sterile water to a final concentration of 3 $\mu$g/$\mu$l.

### 2.4.1 DNA digestion

DNA was eliminated from the RNA samples by digestion using DNaseI enzyme from New England Biolabs (Cat#M0303S) at 37°for 10 minutes. Reaction was stopped

by heating the sample at 75°C for 10 minutes.

## 2.5 cDNA synthesis

cDNA was synthesized using the RETROscript® kit (Ambion Cat#AM1710M) following the manufacturer's instructions for the Two-step RT-PCR with heat denaturation of RNA protocol. One microgram total RNA was combined with either random decamers or oligo(dT) primers and nuclease-free water, then denatured at 85°C before the addition of the remaining RT reagents: 10X RT buffer, dNTPs, RNase inhibitor, and the M-MLV Reverse Transcriptase. Reverse transcription of cDNA was carried out at 42°C for 1 hour. The reaction was stopped by inactivating the reverse transcriptase at 92°C for 10 minutes.

## 2.6 Primer design and Polymerase Chain Reaction (PCR)

Primers were designed using the tool Primer3 (http://bioinfo.ut.ee/primer3/) using the default settings. For *Ubx* Universal and Distal, primers previously published were used (Thomsen et al., 2010), (Burnette et al., 1999). Primer stocks were resuspended in TE buffer (10mM Tris, 1mM EDTA, pH 8.0) at 100 $\mu$M, working solutions 10 times dilutions in distilled water. PCR reactions were carried out using one unit of Taq polymerase (New England Biolabs - Cat#M0273S), 1 $\mu$l of each forward and reverse primers, 0.5 $\mu$l dNTP mix (10 mM each), 1$\mu$l cDNA, in a final volume of 25 $\mu$l of 1 X Taq buffer. A typical reaction used the following program:

| | |
|---|---|
| 1 cycle | DNA denaturation at 94°C - 5 min |
| 29 cycles | denaturation at 94°C- 30 seconds |
| | annealing at 51-58°C- 30 seconds |
| | extension at 72°C- 30 seconds to 1 min |
| 1 cycle | final extension at 72°C- 10 minutes |
| Hold | at 4°C |

## 2.7 Agarose gel electrophoresis

DNA visualisation was carried out using agarose/ethidium bromide gel electrophoresis. 1% agarose gels were prepared in SB buffer (Sodium borate 10mM, pH 8.2) (Brody and Kern, 2004) and 0.5 $\mu$M ethidium bromide. DNA ladder New England Biolabs 100 pb (Cat #N3231S) or 1 kb (Cat#N3232S) were used as reference. Gels were run at 10 Volts per cm. The analysis of band intensities was carried out with ImageJ software. The image from the gel obtained from the gel documentation station was open in the ImageJ software, the background signal was substracted and the bands were measured by drawing a square around them and measuring its integrated density, which gives the sum of pixel values within the selected area. The values obtained were copied in an Excel spreadsheet for statistical analysis and graph design.

## 2.8 Gene cloning

TA cloning was performed with Taq PCR products and pGEMT-Easy vector systems (Promega Cat#A1360) following manufacturer's instructions. Ligation was performed overnight at room temperature and bacteria were transformed by heat shock at 42°C for one minute. Recovery was allowed at 37°C with 1 ml LB media and plated on ampicilin LB plates. Cloning for sequencing was performed with PCR products using Phusion high fidelity enzymes from New England Biolabs (Cat# M0530S) and A overhangs were added with one unit of Taq polymerase, 0.2 mM dATP (or dNTP mix) at 72°C for 30 minutes.

## 2.9 Site-Directed Mutagenesis

Primers used were designed using the Quick Change II site-directed mutagenesis online tool from Agilent technologies and were adjusted to 50 bp long. They were synthesised by Sigma using standard protocols. The Phusion enzyme (NEB Cat# M0530S) was used for the amplification reaction under the following conditions: 50 to 100 ng of template plasmid was used, mixed with 1 $\mu$l of each oligo (10 nMol ), 10 $\mu$l of 5X Phusion buffer, 1 $\mu$łof dNTPs 10 mM, 1 $\mu$l of the Phusion polymerase and

distilled water up to 50 $\mu$l. The amplification conditions were as follows: 95°C for 2 minutes, then 95°C for 30 seconds, 55°C for 1 minute, 68°C for 2 min/kb. These last three steps were repeated 20 times. Then, 1 $\mu$l of DpnI restriction enzyme was added for 1.5h at 37°C to digest the template bacterial produced plasmid. A no primer and a no DpnI control were performed. 1 $\mu$l of the reaction was used to transform 30 $\mu$l of competent E. coli Dh5$\alpha$ bacteria using standard protocols. Colonies were picked and minipreps performed to be send to sequence with T7 oligo primer (MWG Eurofins). The identified colonies were then cultivated in 50 ml LB media for midiprep preparation.

## 2.10 Mini and midipreparation of plasmid DNA

For plasmid purification, the QIAprep Spin Miniprep (Cat#27106) or Midiprep (Cat#12143) were used according to the manufacturer's instructions. The final plasmid was eluted in a volume of 50 $\mu$l of distilled water.

## 2.11 SDS Polyacrylamide gel electrophoresis

Mini-PROTEAN® Tetra Cell (Biorad Cat#165-8000) system was used. The glass plates were washed with soap and then rinsed with tap water, distilled water and ethanol. They were cast in the casting system with the gel mix using 37.5:1 acrylamide to bisacrylamide stabilized solution (ProtoGel from National Diagnostics) as shown below.

| Component | Resolving gel (12% acrylamide) | Stacking gel (4% acrylamide) (4% acrylamide) |
|---|---|---|
| dd H2O | 4.35 ml | 3.2 ml |
| 1.5M Tris-HCl, pH 8.8 | 2.5 ml | - |
| 0.5M Tris-HCl, pH 6.8 | - | 1.25 ml |
| 10% SDS | 100 $\mu$l | 50$\mu$l |
| Acrylamide/bisacrylamide (37.5:1, 40% sol.) | 3 ml | 500 $\mu$l |
| 10% Ammonium Persulfate | 50$\mu$l | 25$\mu$l |
| TEMED | 5 $\mu$l | 5$\mu$l |

Both mixes were prepared without APS and TEMED. These components were added to the resolving gel, mixed and poured. 50% isopropanol was layered on top of the gel to nivelate the gel. When the resolving gel was polymerized, the isopropanol was removed and APS and TEMED were added to the stacking gel, mixed and poured, placing the comb on top. Samples were mixed with 2X Protein Loading Buffer and boiled for 5 min or incubated at 85°C for 10 mins. The gel was run at 200V for 45 mins.

### 2.11.1 Buffers

**Running buffer**

For 1L, 10X buffer, the following was combined:

Tris: 30.27 g, glycine: 144.13g, SDS: 10g, pH 8.8.

**Protein Loading Buffer**

| for 10 mL | volume | final concentration |
|---|---|---|
| SDS 10% | 4ml | 4 % (w/v) |
| Tris-HCl 0.5M pH 6.8 | 2 ml | 100 mM |
| Bromophenol blue 0.5% | 2 ml | 0.1% (w/v) |
| Glycerol | 2 ml | 20% (v/v) |

## 2.12 Western Blot analysis of proteins

The proteins were transfered at 100 V for at least 2 hours to a nitrocellulose membrane (GE Healthcare Hybond ECL Membrane Cat#RPN203D). The membrane was stained temporarily with Ponceau red (Sigma Cat #P7767) to control for the efficiency of the transfer. The Ponceau was washed with distilled water and the membrane was blocked with 10% (w/v) non fat dry milk in PBT 0.05% Tween for 1h at room temperature and shaking. The primary antibody was diluted in 5% (w/v) milk in PBT 0.05% Tween and incubated overnight at 4 °C and shaking. The antibodies used were diluted as follows: anti *Ubx* (FP3.38)-1:100, anti-Elav-1:400, anti-H3P 1:1000. The primary antibody was washed with PBT 0.05% Tween 3 times for 20 minutes and then the secondary antibody was incubated in 5% (w/v) milk in PBT 0.05% Tween for 2h at room temperature and shaking. The secondary antibodies were diluted in PBT 0.05% Tween as follows: anti mouse 1:1000, anti-rabbit 1:500, anti-rat 1:1000. The signal was developed using the Amersham ECL Western Blotting Detection Reagents (Cat#RPN2109), used as indicated by the manufacturer. The membrane was exposed to films (Amersham Hyperfilm ECL Cat#28906835) for 5 min, 30 min and 2 hours.

## 2.13 Nuclear extract

I used a protocol designed by the Laboratory of Donald Rio (University of California, Berkeley) as described in the Rio Lab Protocols Compendium 2007, provided by Julie Aspden "hnRNPs-enriched Nuclear Extract".

### 2.13.1 Solutions

**HNEB1**

| final concentration | for 100 ml | stock solution |
|---|---|---|
| 15 mM HEPES pH 7.6 | 1.5 ml | 1M (pH 5.5) |
| 10 mM KCl | 0.33 ml | 3M |
| 350 mM sucrose | 11.9 gr | powder |
| 2.5 mM MgCl2 | 0.25 ml | 1 M |
| 0.1 mM EDTA | 0.02 ml | 0.5 M |

The solution was adjusted to pH 7.6 with KOH.

**HNEB2**

| final concentration | for 50 mL | stock solution |
|---|---|---|
| 10 mM Tris pH 7.5 | 2 ml | 1 M |
| 100 mM NaCl | 1 ml | 5M |
| 2.5 mM MgCl2 | 0.125 ml | 1M |
| 0.5% Triton X-100 | 1.25 ml | 100% |

Add protein inhibitor prior to use (ROCHE Cat#04693124001)

### 2.13.2 Procedure

1-2 mg of overnight embryos were collected on big apple juice agar plates, the eggs were recovered from the plates, rinsed in baskets and dechorionated as described. Embryos were transfered to a 50 ml falcon tube with PBS, all liquid was removed and weight of embryos was determined. Embryos were resuspended in 2 ml of HNEB1 per gram and transferred to a medium dounce. Falcon tubes were rinsed with 2ml per gram of embryos and mixed to embryo suspension. The embryos were smashed with a loose pestle, using 5 strokes on ice. The lysate was then filtered with one layer of miracloth. The dounce was rinsed with 1 ml per gram of embryos and filtered as well. The filtrate was then transfered to a 15 ml glass centrifuge tube and centrifuged in a Sorvall at 9,100 rpm (10,000g) in a SS-34 rotor for 15 minutes at 4°C. The supernatant was decanted and disposed. The pelleted nuclei were resuspended in 1ml of HNEB2/gram of embryos. The very solid yolk pellet was

eliminated. The nuclei were transfered to the douncer again and smashed with the tight pestle with 15 strokes on ice. A volume of HNEB2/30% sucrose equal to that of the smashed nuclei was placed in a glass tube and the smashed nuclei overlayed on top. The nuclei were centrifuged for 15 min in a Beckman SWI28 rotor with no brake, at 6,350 rpm (4,800 g). Four fractions were obtained; the top one was eliminated (lipids) and the second one was recovered (nuclear extract). Glycerol was added to 15% (v/v) and 40 $\mu$l aliquots were made and frozen in liquid N2.

## 2.14 Labelling of RNA with $^{32}$P

500$\mu$l of digested gel purified plasmid was mixed with 2$\mu$l T7 polymerase buffer, 0.75 $\mu$l of unlabelled ATP, CTP, GTP; 0.5 $\mu$l RNAse inhibitor; 1.5 $\mu$l T7 polymerase; 2$\mu$l of $^{32}$P UTP (3000Ci/mmol, BLU007H250UC, Perkin Elmer); water was added up to 20 $\mu$l. The reaction was incubated at 37°C for 2 hours.

### 2.14.1 Gel purification of labelled probe

The *in vitro* transcription reaction (20 $\mu$l) was mixed with 30 $\mu$l of water and non incorporated rNTPs were separated using illustra ProbeQuant G-50 micro columns (GE Healthcare). The probe was purified for a second time using a denaturing polyacrylamide gel. To prepare the gel, 20 g of urea were dissolved in 18.2 ml distilled water at 50°C. 2 ml of 10X TBE, 5 ml of 40% acrylamide 19:1 Acrylamide to bisacrylamide stabilised solution (National Diagnostics) were added and mixed well. 400 $\mu$l of 10% ammonium persulfate and 25 $\mu$l TEMED were added at the end and poured immediately. When the gel was polymerised, it was mounted to the tank and 1L of 1X TBE was added. Wells were flushed before loading the samples and run the gel at 400 V for 1h. To determine probe yield and specific activity, I used the method described by Ambion (http://www.biocompare.com/Application-Notes/42363-Determining-RNA-Probe-Specific-Activity-And-Yield/). This method first calculates the theoretical maximum yield of RNA given the individual nucleotide concentration considering rUTP as the limiting component. Afterwards, the incorporation efficiency is calculated by measuring the cpm of labelled RNA before and after purification. With this information, it is possible to calculate the

theoretical amount of labelled RNA produced. To confirm that the probes were evenly quantified, I resolved 1$\mu$l of RNA sample in a denaturing polyacrylamide gel and exposed it to a phosphoimager plate.

## 2.15 Recombinant Elav production

The recombinant GST-*Elav* protein was obtained using a pGEX-GST vector (kindly donated by M. Soller) were the *Elav* open reading frame was inserted downstream of the GST tag (Soller and White, 2003). Another member of the Alonso Lab, Ali Mumtaz carried out the purification using the following protocol: BL21 bacteria transformed with the pGEX-GST-*Elav* construct was cultured at 37°C until it reached OD 600 in LB medium and then bacteria were induced at 30°C with 0.4 mM IPTG for 3 hours. The culture was then centrifugated and the bacterial pellet was resuspended in Lysis buffer (PBS, 1% Triton X, 1mM DTT, 0.1 mg/ml lysozyme and Protease Inhibitor cocktail from Roche). Cells were flash frozen and thawed at 37°C 10 times and then lysed at 4°C on a wheel. The lysate was then sonicated 3 times for 10 seconds in a water bath sonicator. The lysate was centrifuged and the supernatant recovered and incubated with pre-equilibrated 50% (v/v) slurry of anti-GST agarose beads (Invitrogen Cat #G2879)(0.5 ml bead volume per litre of culture) for 2:30 hours at 4°C on the wheel. The beads were washed three times with 10 volumes of beads of PBS and with PreScission Protease Buffer (50mM Tris-HCl, 100mM NaCl, 1mM EDTA, 1mM DTT pH 8.0) 2 times. Protein was eluted with 160U of PreScission Protease (GE Healthcare Cat #27-0843-01) per ml of beads equilibrated with PreScission Protease Buffer overnight at 4°C. Supernatant was cleared from residual protease and GST by incubation with pre equilibrated agarose Protein A beads. Protein concentration was determined by Bradford assay and purity by SDS-PAGE and Coomassie staining.

## 2.16 UV crosslinking of proteins to RNA

700 pM of uniformly labelled RNA was incubated with 5X binding buffer (400 mM KCl, 100 mM HEPES pH 7.6, 20 mM MgCl2, 25% glycerol and 5 mM DTT), 1 $\mu$l of tRNA 1mg/ml and 38 $\mu$g/ul of recombinant Elav or 40% Nuclear Extract in 10

$\mu$l. The mix was incubated at room temperature for 10 mins, divided in 2 samples, one UV-crosslinked for 12 minutes on ice in a Stratalinker and the negative control kept on ice, without crosslinking. The samples were digested with 0.5 $\mu$g RNAse A at 37°C for 15 mins. Samples were run on a SDS-PAGE and the gel was dried and exposed to a phosphoimager plate over night (Walker et al., 1998), (Soller and White, 2003).

## 2.17 Electrophoretic Mobility Shift Assay (EMSA)

The sizes of the probes used are described in the following table:

| Probe | Size1 | Size2 |
|---|---|---|
| *ewg* | 235 | 157 nt |
| *ewg* as | 257 | 160 nt |
| site3 | 215 nt | 151 nt |
| site5 | 300 | - |
| site8 | 171 nt | 157 nt |
| site13 | 200 nt | 153 nt |
| site16 | 192 nt | 155 nt |

EMSA experiments were performed as described in (Soller and White, 2003). The first step is to prepare the Buffer mix, as follows:

### 2.17.1 Buffer mix

1 $\mu$l of 10X binding buffer (500 mM Tris-HCl pH 7.5, 400 mM KCl, 350 mM NaCl), 0.5 $\mu$l of BSA 1 mg/ml, 0.05 $\mu$l of DTT 0.1M and 0.5 $\mu$l of RNAse inhibitor, per reaction.

The second step is the preparation of the RNA mix:

### 2.17.2 RNA mix

For a 100 pM RNA reaction and 50 g/L tRNA, the following was mixed: 3 $\mu$l of 300 pM labelled RNA, 0.25 $\mu$l of 1 mg/ml tRNA and 1.75 $\mu$l water. The mix was incubated at 65 °C for 5 mins and allow to renature at room temperature for

10 mins. A master mix per RNA probe was prepared to be mixed with different concentrations of Elav.

Then, the reactions were assembled with different quantities of recombinant Elav protein, on a chilled rack over a bucket of ice as shown in the next table:

| Probe | tRNA | *Elav* nM | Buffer mix | RNA mix | *Elav* protein | Water |
|---|---|---|---|---|---|---|
| *ewg* | 25 ng/$\mu$ | 0 | 2.05 $\mu$l | 5 $\mu$l | 0 $\mu$l | 2.95 $\mu$l |
| *ewg* | 25 ng/$\mu$ | 200 | 2.05 $\mu$l | 5 $\mu$l | 1.35 (1:10)$\mu$l | 1.6 $\mu$l |
| *ewg* | 25 ng/$\mu$ | 400 | 2.05 $\mu$l | 5 $\mu$l | 2.7 (1:10)$\mu$l | 0.25 $\mu$l |
| *ewg* | 25 ng/$\mu$ | 600 | 2.05 $\mu$l | 5 $\mu$l | 0.41 $\mu$l | 2.54 $\mu$l |

The reactions were incubated at room temperature for 20 min. The 10$\mu$l reaction was mixed with 3 $\mu$l 50% (v/v) glycerol and loaded onto a 4% (80:1 acrylamide/bisacrylamide) polyacrylamide native gel and run at room temperature at 250 V in 0.5 X TBE. Gels were dried in a gel dryer and exposed on a phosphoimager plate overnight (Walker et al., 1998), (Soller and White, 2003). The images then were analysed with ImageJ.

# Chapter 3

# The role of Elav in *Ubx* RNA processing

## 3.1 Introduction

*Ubx* RNA processing during *Drosophila* development is a very complex and regulated process. As described in the introduction, the *Ubx* transcript contains two constitutive exons (E3' and E5') and three alternative exons (B element, mI and mII, black, red and green respectively in Figure 3.1A) which are used to generate a family of 6 splicing isoforms. Isoform I contains both mI and II, isoform II contains only mII and IV does not contain any microexon. The isoforms containing the B element are named b, whereas the isoforms lacking it are named a. A graphic description of the splicing isoforms is shown in Figure 3.1A.

*Ubx* transcrip also undergoes alternative polyadenylation by the use of a proximal or distal Polyadenylation Site (PAS), generating a "short" 3' UTR of 1kb or a "long" one of 2.4 kb (Figure 3.1B)

A detailed analysis of *Ubx* alternative splicing and polyadenylation during embryonic development carried out by a former member of the Alonso Lab, Lucy Williams using semiquantitative RT-PCR approach (shown in the supplemental material in (Thomsen et al., 2010)) is partly reproduced in Figure 3.2. In this analysis, *Ubx* mRNA appears at 3 hAEL (hours after egg layed) with mainly splicing isoform Ia and isoform IIa showing at very low levels. Both isoforms increase in abundance along development and at 11 hAEL isoform IVa appears. By 21 hAEL the three

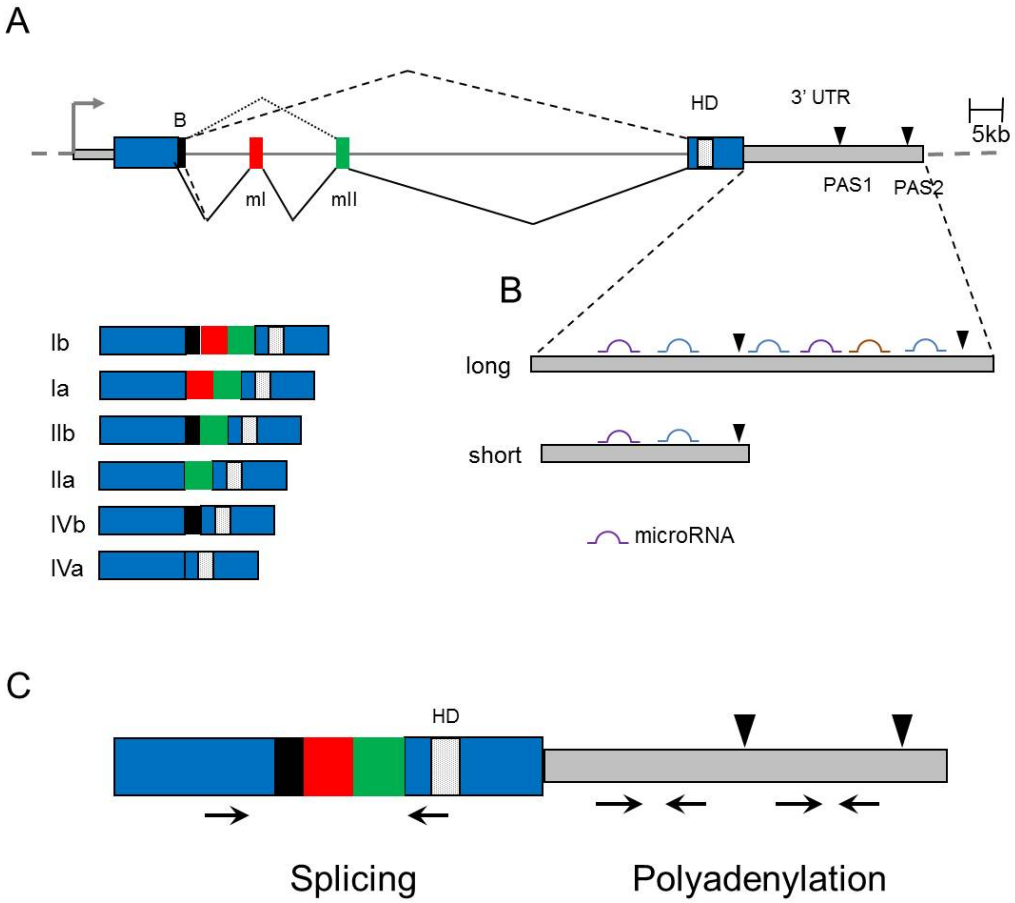


Figure 3.1: *Ubx* alternative splicing and polyadenylation. A) The *Ubx* transcript produces six splicing isoforms by the inclusion or exclusion of the B element (in black), microexon I (in red) or microexon II (in green) (mI and mII respectively), named Ib to the isoform containing all the elements, Ia containing mI and II, IIb contains the B element and mII, whereas IIa contains only the mII, IVb contains only the b element and IVa does not include any alternative exons. B) *Ubx* transcripts produces a short or long UTR by the use of a proximal or a distal Polyadenylation Site (PAS1 or 2 respectively). The use of the proximal PAS produces a "short" UTR of around 1kb and the use of the distal PAS produces the "long" 3' UTR of around 2.4 kb. The longer the UTRs the bigger number of cis elements present in the transcript that could potentially determine its post translational regulation, for example, by microRNAs. Figure modified from (Kornfeld et al., 1989). C) Position of primers along the *Ubx* transcript for PCR exploring splicing and polyadenylation patterns. HD, Homeodomain; mI and II, microexon I and I; PAS, polyadenylation site.

isoforms are equally abundant.

The polyadenylation analysis carried out by Lucy Williams was done specifically for each splicing isoform, with primers designed on the exon-exon junction for each isoform and before and after the proximal polyadenylation site (scheme to the right in Figure 3.2 B). The proximal PAS is the main polyadenylation site used in all *Ubx* transcripts at 3hAEL so most of them have a short UTR. At 5 hAEL a very small quantity of long 3' UTR starts appearing associated with isoform Ia. As development progresses, the long UTR is more abundant and its association shifts to IIa at 13 hAEL, with of most of the long UTR associated to IVa isoform at 17 to 21 hAEL.

*Ubx* RNA generates more diversity of isoforms towards the late stages of development and previous research have found that within the nervous system there is a different control of RNA processing (O'Connor et al., 1988). What are the factors that shape the splicing and polyadenylation patterns within the nervous system? The RNA binding protein Elav emerges as a very strong candidate regulator as it is a neuronal specific factor and its role on the splicing and polyadenylation of other genes has been very well documented.

Previous work in the lab done by Dr. Stefan Thomsen and Elvira Lafuente demonstrated a role of Elav in *Ubx* polyadenylation choice. They used a null mutant for *elav* (*elav*5/FM6) and monitored the choice of the proximal or distal PAS of *Ubx* by *in situ* hybridisation (ISH) using a probe against the distal region of the 3' UTR (Figure 3.3C-D). The results showed a clear shift from the distal PAS to the proximal PAS in the *elav* mutants at stage 16 (Rogulja-Ortmann et al., 2014). Consistent with this result, over expression of Elav at earlier stages results in an up regulation of the longer UTR (Figure 3.3 E-F).

In this chapter I describe the role of the Elav protein in *Ubx* splicing and polyadenylation. Using RT-PCR as an experimental strategy, I show that Elav helps to shape the patterns of *Ubx* isoforms achieved during embryonic development in *Drosophila*.

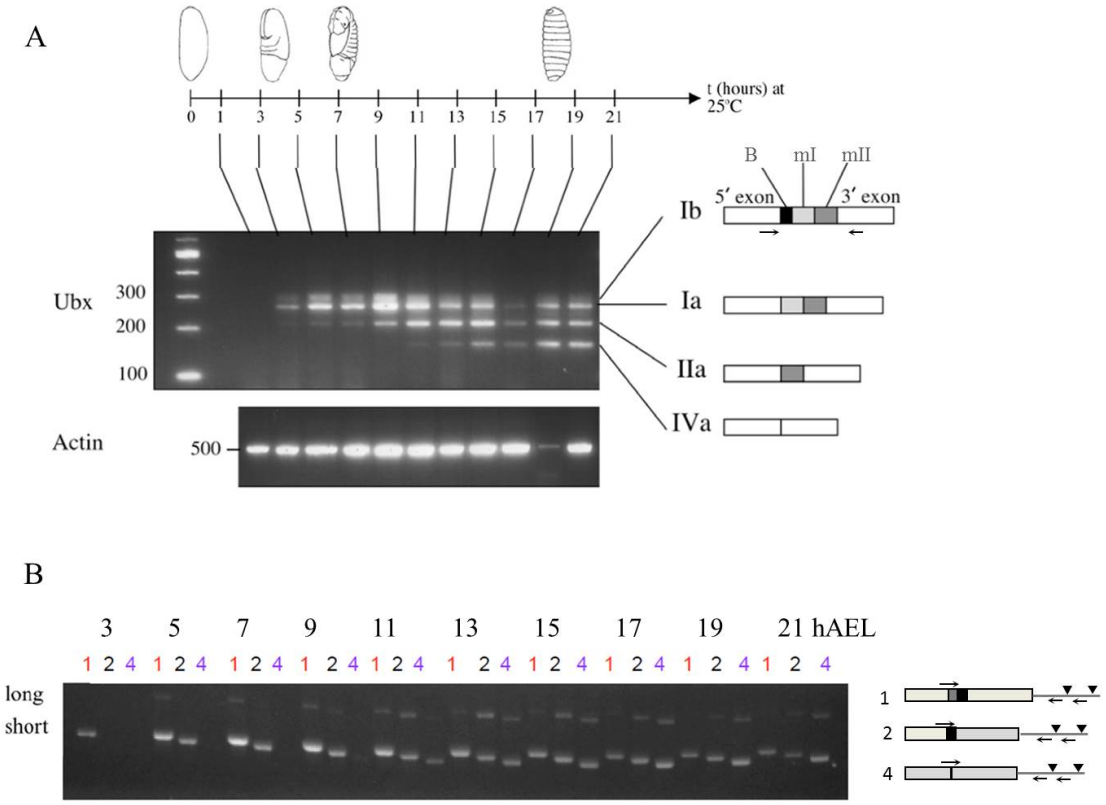


Figure 3.2: *Ubx* isoforms processing during *Drosophila* embryonic development. A) RT-PCR using primers on the 5' and 3' exons spaning the variable region and RNA samples collected at the time points shown above. The scheme of the isoforms is shown to the right also shows the position of the primers. The RT-PCR shows a progressive enrichment of all the splicing isoforms during embryonic development. B) Splicing and polyadenylation relationship, explored by RT-PCR using a forward primer in the exon-exon junction for each splicing isoform and a reverse before or after the proximal PAS (termed short and long respectively) as shown in the diagram at the right. The progression of development shows a gradual increase of the use of the distal PAS, generating more *Ubx* mRNA with "long" 3' UTR. Work done by Lucy Williams, Figure modified from supplementary Figure 1 in (Thomsen et al., 2010).

## 3.2 Results

As described in the introduction, *Ubx* has two alternative micro exons and two described polyadenylation sites that produce six splicing isoforms and two polyadenylation isoforms (Figure 3.1). The splicing and polyadenylation patterns during embryonic development have been determined before (O'Connor et al., 1988), (Kornfeld et al., 1989), (Lopez and Hogness, 1991), (Artero et al., 1992), (Thomsen et al., 2010). In this chapter, the effect of Elav in the establishment of such patterns was determined using an *elav* null mutant to explore by RT-PCR the changes in splicing and polyadenylation isoform abundance in comparison with Oregon Red (OR) wild type flies. The stock used in this study was kindly donated by Matthias Soller and its full genotype is: w[*] sn *elav*[elav5]/FM6/y[2]Y[611].

This mutant is a deletion of the open reading frame and does not produce Elav protein (Campos, A.R., Grossman, D., White, 1985). The balancer was changed for FM7c, GAL4-Kr.C,UAS-GFP to allow me to sort mutant homozygotes by the absence of a fluorescent marker in late embryonic stages (Figure 3.3). Embryos of late embryonic stages (15 onwards) were collected and RNA was isolated using Trizol. cDNA was prepared using oligo dT under standard conditions (see details in the Methods section).

### 3.2.1 Elav affects the alternative splicing pattern of *Ubx*

To explore the *Ubx* splicing patterns, primers designed around the variable microexons, in the E5' and E3', were used (Burnette et al., 1999), (Figure 3.1 C). Semiquantitative PCR of late stage embryos using agarose/ethidium bromide quantification showed clear differences in the splicing isoforms produced by wild type flies compared to the $elav^5$ mutant (Figure 3.4A).

Initial quantification of the PCR products taking the profile of the bands of the gel showed that in the wild type, the most abundant splicing isoform is IVa, followed by IIa and then by IVa (Figure 3.4 A and dashed line in B). In contrast, in the $elav^5$ mutant, isoform Ia is more abundant than isoform IVa (Figure 3.4 A and black line in B). Two types of analysis were performed: Ia/IIa and Ia/IVa isoform ratio and percentage of each isoform to total *Ubx*.

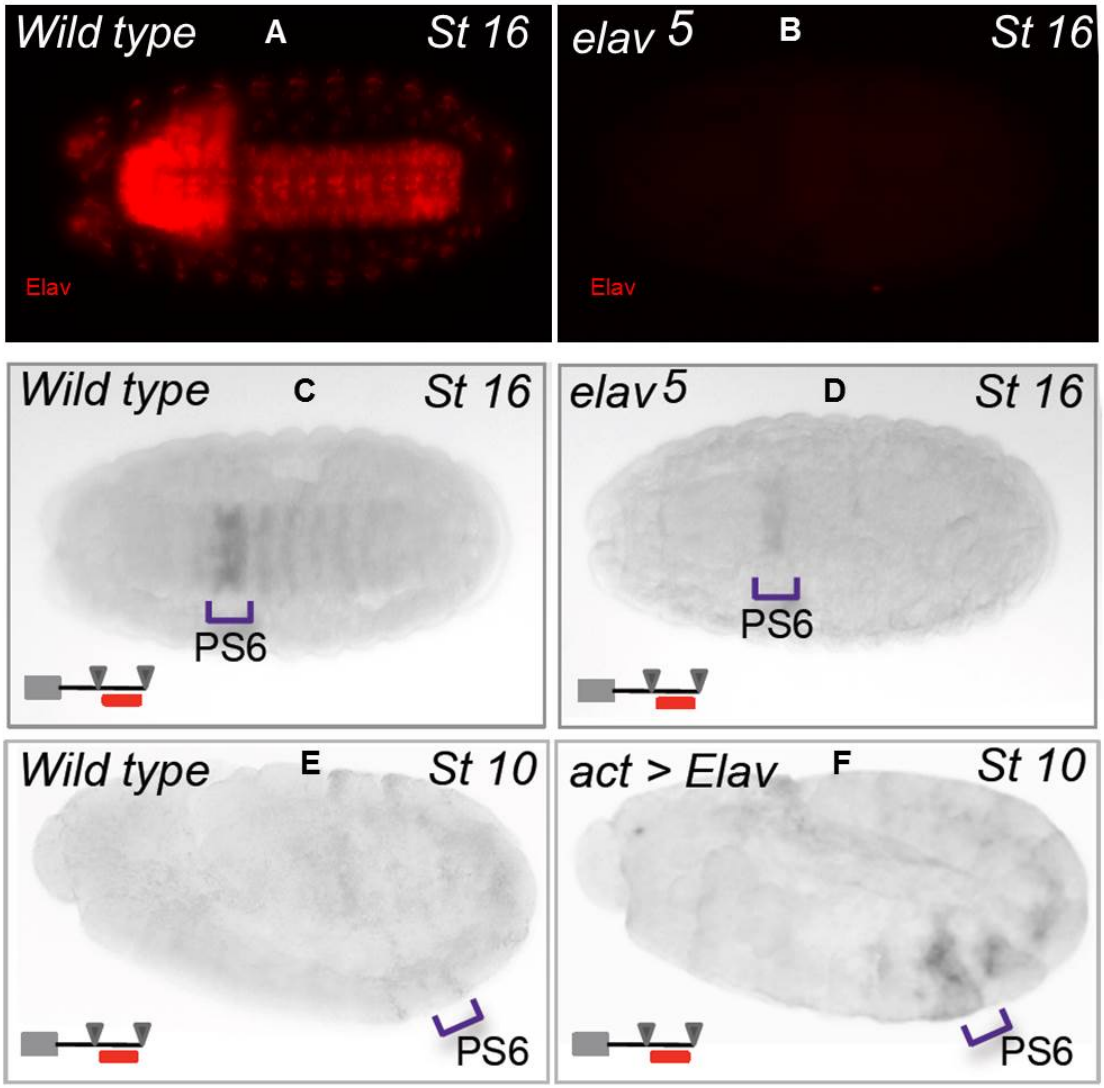


Figure 3.3: Elav affects *Ubx* PAS in late embryogenesis stage early 16. A) and B) Antibody staining against Elav protein in wild type and $elav^5$ homozygote mutant respectively. C) and D) *in situ* hybridisation against a region in the distal 3' UTR at stage 16, (scheme at the bottom) shows the normal use of the distal PAS in wild type embryos, generating the long 3' UTR which is reduced in the $elav^5$ homozygote mutant embryos. E) Wild type embryos at the earlier stage 10 expresses very little long *Ubx* 3' UTR. F) Ectopic expression of ELAV at stage 10 shows an increase in the expression of long 3' UTR of *Ubx*, indicating that Elav is sufficient to induce a change in *Ubx* polyadenylation site choice. Work by Elvira LaFuente, Figure taken from (Rogulja-Ortmann et al., 2014).

The proportions of IIa and IVa to Ia isoforms are shown in Figure 3.4 C and D. In wild type, the Ia/IIa isoform is around 0.9 which means that both isoforms have very similar abundance. In contrast, the Ia/IIa proportion in the *elav*$^5$ mutant is 1.2: the IIa isoform is less abundant than Ia. In relation to the IVa isoform, the Ia/IVa ratio in wild type is 0.7, reflecting a higher abundance of IVa isoform compared to Ia which is reversed in the *elav*$^5$ mutant: the ratio is around 2.5, reflecting much less IVa isoform.

The analysis by isoform percentage related to total Ubx showed clearly an increase of isoform Ia abundance in the *elav*$^5$ mutant from 30% to 45% whereas isoform IVa decreases from 35% to 16% (Figure 3.4E). In contrast, isoform IIa does not seem to be affected, shifting only from 33% to 35% and no significant difference.

In these RT-PCR experiments for *Ubx* splicing isoforms, the *elav*$^5$ mutant background shows changes in the balance of the *Ubx* isoforms produced, affecting the exclusion of both microexons. In late stages, where there is an approximate equal abundance of the three isoforms (Ia, IIa, IVa), the lack of Elav protein shifts the balance towards isoforms including one or two microexons. This pattern is more characteristic of earlier stages. This indicates that Elav is involved in microexon splicing and that it affects both microexons. This suggests that Elav could have a general role in alternative exon splicing, or that the elements needed for Elav function are present in both microexons.

### 3.2.2 Elav changes the polyadenylation pattern of *Ubx*

To explore changes in polyadenylation site use, two sets of primers were designed, one pair within the proximal region of the 3' UTR (before the proximal polyadenylation site) and another pair within de distal portion, after the proximal polyadenylation site (Figure 3.1 C). The RT-PCR experiments show that at late stages of embryonic development, there is a high production of the distal fraction of the 3' UTR in the wild type control, compared to the proximal UTR amplicon. In contrast, with the *elav*$^5$ mutant, the distal band produced has a reduced intensity (Figure 3.5A). Quantification of the intensity of the bands shows a reduction of about 30% in distal 3' UTR production (Figure 3.5B).

These experiments demonstrated that Elav also plays a role in polyadenylation

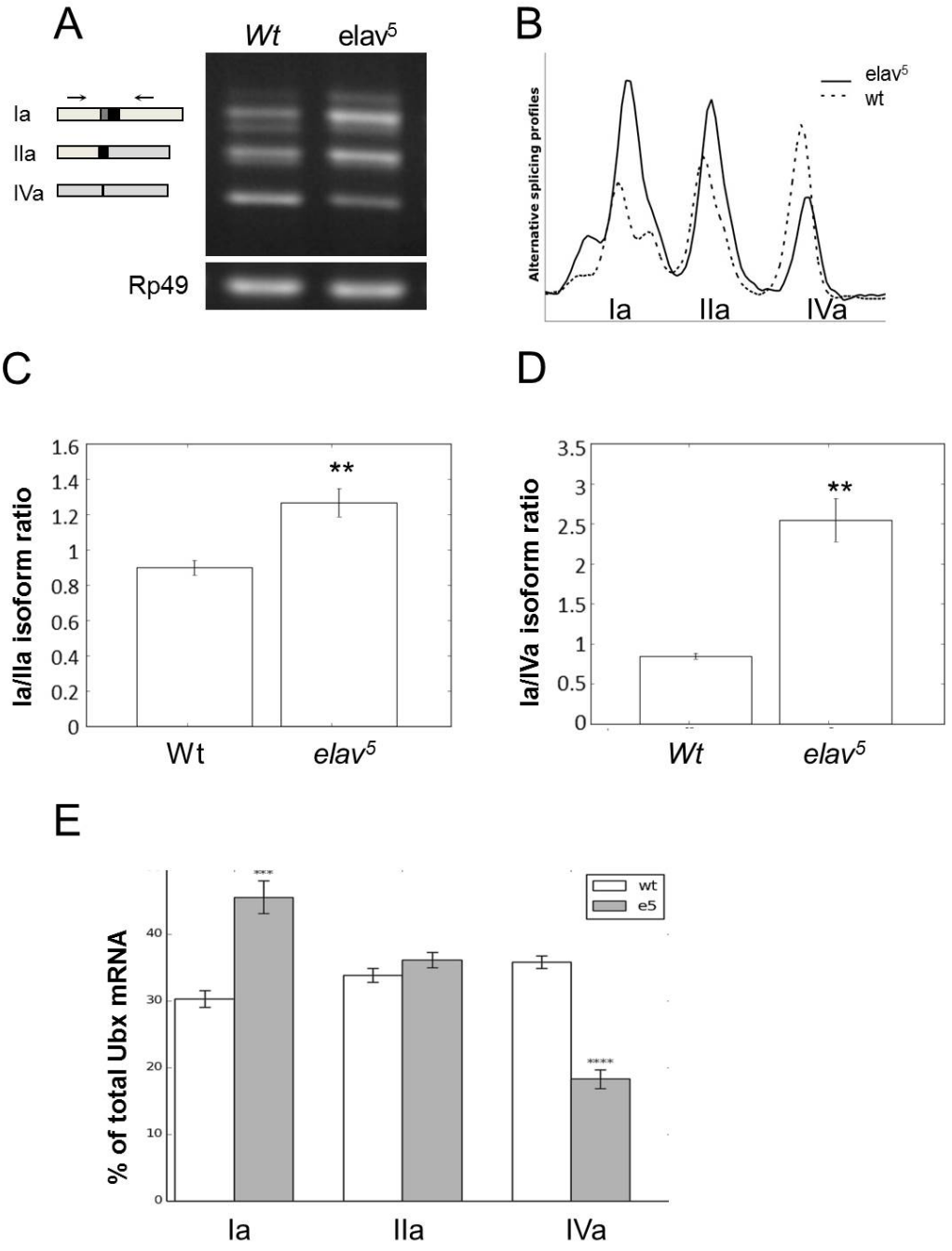


Figure 3.4: Elav affects *Ubx* splicing at late stages of embryogenesis. A) RT-PCR using primers outside the variable region as shown in the scheme to the left. *rp49* amplicon was used as loading control. B) Profile of the bands obtained showing the top band to the left and the bottom band to the right; wild type in continuous line and $elav^5$ mutant in dashed line. C) Proportion of Ia/IIa isoforms relative quantification from the agarose gel showing a higher proportion of Ia in the $elav^5$ mutant. D) Proportion of Ia/IVa isoforms relative quantification from the agarose gel showing a very high proportion of Ia in the $elav^5$ mutant, reflecting more abundance of Ia isoform compared to wild type. E) Amplicons of each isoform measured as a percentage of total *Ubx* mRNA. $elav^5$ mutant produces significantly more Ia isoform and less IVa isoform. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to wild type.

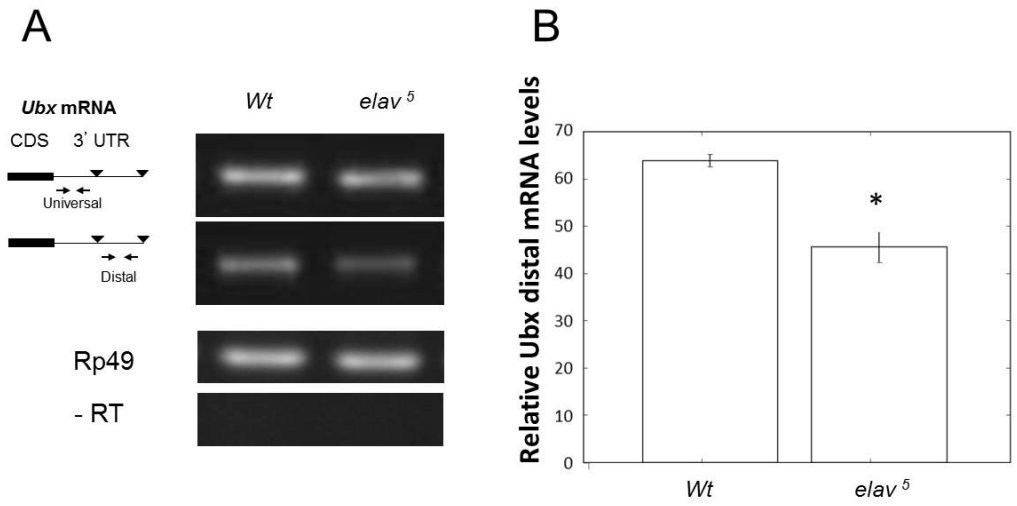


Figure 3.5: Elav affects polyadenylation site choice. A) RT-PCR with primers for the proximal or distal 3' UTR as shown in the scheme to the left for wild type or *elav*$^5$ mutant late embryos. B) Relative quantification of the bands in the agarose gel showing a reduction of the distal UTR of around 30% in the *elav*$^5$ mutant. Bars represent standard error and * p$\leq$ 0.05 of the t-test in respect to wild type.

site choice at late stages of embryonic development. The absence of Elav promotes the use of the proximal polyadenylation site, generating a shorter 3' UTR. These experiments confirm by RT-PCR, the effect previously observed in an *in situ* approach to study the role of Elav in alternative polyadenylation. This agrees with previous reports of Elav regulating the use of alternative polyadenylation signals (Soller and White, 2003).

### 3.2.3 Elav affects *Ubx* mRNA abundance

With the aim to establish the biological role of Elav-dependent *Ubx* regulation, the previous experiments showed that Elav has a role in two *Ubx* RNA processing events, splicing and polyadenylation. However, a role for Elav in other events regulating *Ubx* cannot be discarded. Next, I tested the effect of Elav in *Ubx* expression levels.

To this end, I asked if the *Ubx* mRNA levels were affected in an *elav*$^5$ mutant background. For this, primers within the E3' exon were used in RT-PCR experiments using RNA collected from wild type and *elav*$^5$ homozygote mutants at stage 16 of embryonic development. The PCR shows that the total levels of *Ubx* mRNA were reduced, relative to *rp49* mRNA (Figure 3.6A). The quantification of the bands

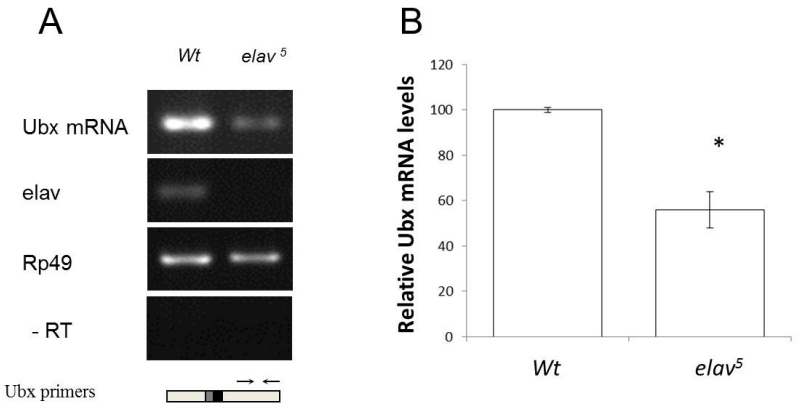


Figure 3.6: *Ubx* RNA levels are reduced in $elav^5$ mutant. A) RT-PCR was carried out as described in the Methods chapter using primers in the E3' exon for total mRNA of *Ubx*. PCR for *elav* mRNA as a control for the $elav^5$ mutant and *rp49* as RNA input control. B) Relative quantification of the bands in the agarose gel. Bars represent standard error and * $p \leq 0.05$ of the t-test in respect to wild type.

normalised to *rp49* showed a reduction of around 30% of total *Ubx* mRNA in the $elav^5$ mutant (Figure 3.6B).

This indicates that Elav could play a role in transcription itself or that a defect in mRNA processing is somehow leading to mRNA decay. Work in the lab by Joao Osorio using the 35UZ Ubx-lacZ promoter fusion showed no difference in *Ubx* expression in $elav^5$ mutant embryos compared to wild type (Rogulja-Ortmann et al., 2014). However, fluorescent *in situ* hybridization (FISH) experiments using intronic probes show that in the $elav^5$ mutants, *Ubx* transcripts get stalled in the chromosome and accumulate there (Figure 3.7A and B) (Rogulja-Ortmann et al., 2014). Figure 3.7 C and D represent the quantification of *Ubx* transcription foci and intensity, showing an increase in the $elav^5$ mutant.

Probably deficient splicing or polyadenylation prevents the transcript from being released from the transcription site leading to an accumulation in the chromosome and a decrease in the free cytoplasmic processed mRNA (Rogulja-Ortmann et al., 2014). This is in line with previous work showing that beta-globin transcripts with defects in splicing and polyadenylation are retain in the chromosome (Carmo-Fonseca et al., 1999).

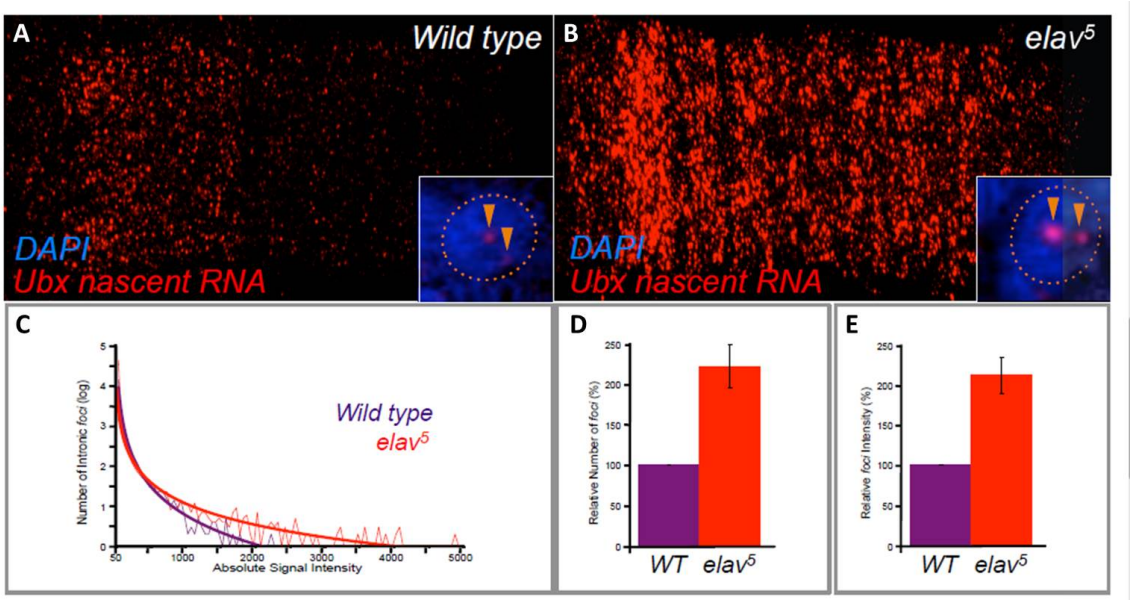


Figure 3.7: *elav*$^5$ mutant leads to nascent *Ubx* RNA transcripts accumulation within the CNS. A) *Ubx* nascent transcript expression detected by FISH using *Ubx* intronic probes against intron 3 indicates that *elav*$^5$ mutant embryos show an overall higher signal in *Ubx* nascent transcript foci. Inset shows *Ubx* signal detected in two discrete nuclear foci (orange arrowheads) per nucleus (blue, DAPI). (B) Quantification of *Ubx* nascent transcript expression in wild type and *elav*$^5$ mutant shows that the distributions of *Ubx* nascent transcript foci [log(10)] versus voxel signal intensity are substantially different among genotypes. Best fit curves are shown. Plotting either the relative number of *Ubx* foci (middle) or relative intensity of *Ubx* foci (right) detected in wild-type and *elav*$^5$ mutant embryos further confirms that mutants show an overall higher level of expression of *Ubx* nascent transcripts. n=7 per genotype; error bars indicate s.e.m.; *P¡0.05 [Wilcoxon matched- pairs signed rank test (non-parametric t-test)]. Figure taken from (Rogulja-Ortmann et al., 2014).

### 3.2.4 Ubx protein levels are reduced in the *elav*$^5$ mutant

The previous experiments showed that the overall levels of *Ubx* RNA were reduced in the *elav*$^5$ mutant, leading to the question of whether the protein levels were also reduced. To this end, a Western blot was performed using whole embryos of late developmental stages, from both wild type and *elav*$^5$ homozygote mutants. The embryos were collected, dechorionated and homogenised in protein loading buffer. The antibody FP3.38 against Ubx was used to detect all the protein isoforms and the SDS-PAGE gel was able to resolve all the Ubx protein isoforms. The wild type sample shows four distinctive bands corresponding to isoforms Ib, Ia, IIa and IVa (Figure 3.8 A) whereas the *elav*$^5$ mutant sample shows three main bands (Ib, Ia and IIb) and a reduced band that corresponds to the IVa isoform. A quantification of the isoforms together reveals a small non significant reduction in total Ubx protein levels (Figure 3.8B) which might only correspond to the missing IVa isoform. The reduction in total protein is not as evident as the reduction observed in total RNA. However, analysing the individual isoforms, the changes are more evident (Figure 3.8C). The increase of isoform Ia and decrease of IIa and IVa is small but significant. Comparing the RNA and protein products of Ubx, we observe that the Ia, IIa and IVa RNA isoforms show a similar level of expression, around 30% each. However, at protein level there is a decreased expression of the IVa isoform (less than 20%) (Figure 3.8C). In other words, isoform IVa shows higher expression at the RNA levels compared to the protein levels. This discrepancy could be explained by a translational lag, in other words, the protein isoforms reflect the RNA isoforms pattern of a previous developmental stage.

I also observed that in an *elav*$^5$ mutant background, the proportion of Ubx protein isoforms also changed compared to those seen in the wild type background. The isoforms containing one (isoform IIa) or two microexons (isoform Ia) are more abundant in the *elav*$^5$ mutants than in the wild type condition. Comparing RNA and protein levels, the effect of the *elav*$^5$ mutant, although in the same direction, is more evident at the RNA level, showing a 30% increase in isoform Ia and a 50% decrease in the IVa isoform. The magnitude of the change at the protein level is only of 15% for isoform Ia and less than 10% for IVa isoform (Figure 3.8C).

Complementary work done by Joao Osorio using antibody staining in dissected

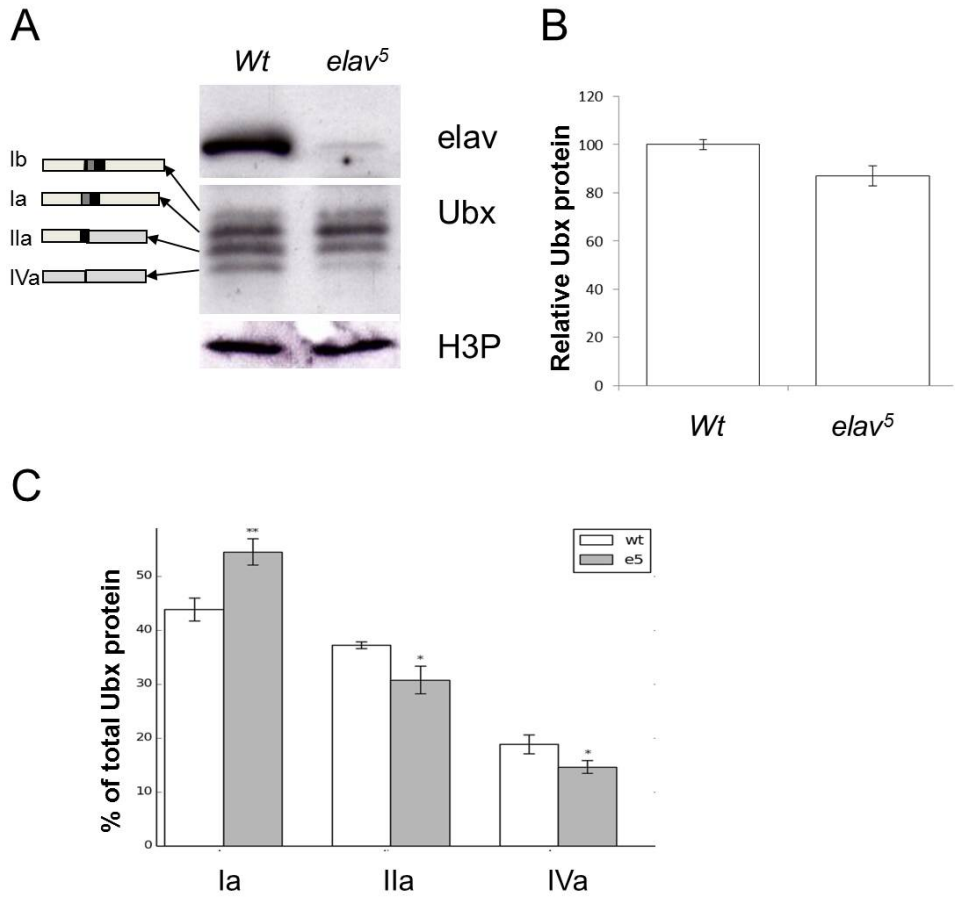


Figure 3.8: Ubx protein levels are also reduced in *elav*$^5$ mutant. A) Western blot using antibodies against Elav as a control for the *elav*$^5$ mutant, H3P as loading control and Ubx. The identity of the protein isoforms is shown in the scheme at the left. B) Quantification of relative levels of total Ubx protein in wild type and *elav*$^5$ embryos. C) Protein isoforms each measured as a percentage of total Ubx protein. *elav*$^5$ mutant produces significantly more Ia isoform and less IIa and IVa isoforms. Bars represent standard error and * $p \leq 0.05$ of the t-test in respect to wild type.

embryonic ventral nerve cord shows a more marked decrease in the Ubx protein levels in the *elav*$^5$ mutant (Rogulja-Ortmann et al., 2014). This could be explained by the fact that the Western blots were performed using whole embryos, containing *Ubx* expressing in non-neuronal tissue (epidermis, muscle and gut) that would not be affected by the *elav*$^5$ mutant.

### 3.2.5 *Ubx* isoforms during embryonic development

*Ubx* RNA processing changes during normal embryonic development, as described in the introduction (Figure 3.2) and Elav is partly responsible for the establishment of such patterns. We wondered if Elav plays a role in the establishment of *Ubx* RNA

patterns at other stages of development, so we focused in discrete late embryonic stages. The *elav* transcript appears at stage 9 of development, on the onset of neuron differentiation. Elav protein starts accumulating at this stage so we tested later stages the effect of its absence. We tested stage 15, early 16, late 16 and 17 for RNA splicing isoforms by RT-PCR and also protein isoforms by Western blot. RT-PCR of the embryonic stages sampled show a gradual increase in isoform IVa abundance in wild type samples that is observed at RNA and protein levels (Figure 3.9 A and B). At the first stage sampled, stage 15, the isoform IVa is absent and starts to appear at early 16 stage (E16). It continues to increase in stage late 16 (L16) and at stage 17, compared to the other isoforms. Isoform IIa at stage 15 is equally abundant to isoform Ia and increases slightly in abundance as development progresses, so at stage 17 it is sloghtly more abundant that isoform Ia. These trends can be observed at the RNA and protein levels (Figure 3.8 A and B). However, in the $elav^5$ samples, the isoform IVa is absent form stages 15 and early 16, and a very weak band appears until late 16 and is visible only until stage 17. This effect is also evident at both RNA and protein levels (Figure 3.8 A and B). The isoforms Ia is more abundant in the $elav^5$ mutant at all the samples tested and at both RNA and protein levels.

The difference of *Ubx* isoform pattern between wild type and $elav^5$ might be explained by a delay on the development of the central nervous system, to test this, a neuronal marker can be used at the developmental stages used for *Ubx* RNA and protein analysis.

## 3.3 Discussion

In this chapter I describe the role of Elav protein in some aspects of *Ubx* RNA processing during embryonic development. Using a null mutant, I demonstrated that Elav affects splicing and polyadenylation decisions, total *Ubx* RNA and protein levels are also affected.

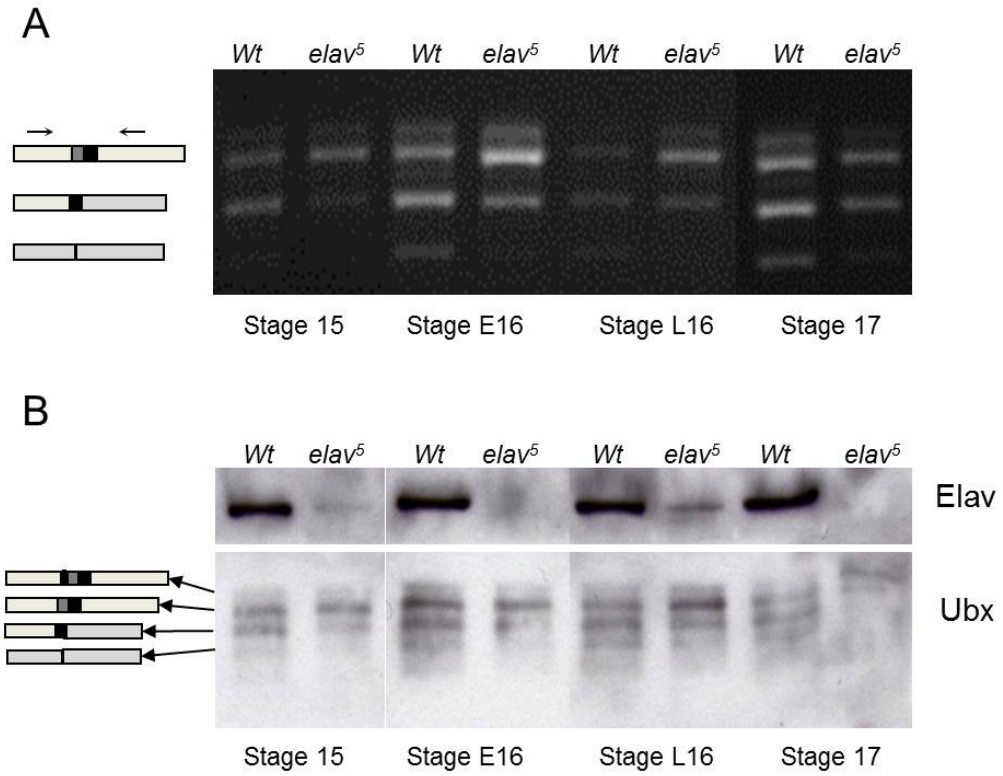


Figure 3.9: Elav affects *Ubx* splicing during development. A) RT-PCR using primers around the variable region showing the splicing isoforms produced in the control wild type and homozygote $elav^5$ embryos at stage 15, early 16, late 16 and 17. Primer position as shown at the left. B) Western blot using whole embryos at stage 15, early 16, late 16 and 17 of wild type and $elav^5$ homozygote embryos. Anti-Elav antibody was used to confirm the reduction of Elav protein in the mutants. The anti-Ubx antibody is able to detect all the Ubx isoforms, showing that in the $elav^5$mutant the isoform IVa is less produced.

### 3.3.1 Elav affects *Ubx* splicing decisions

Elav showed a role in splicing of *Ubx*, promoting the exclusion of both microexons. This results confirm the role of Elav in exon exclusion, already observed in the *Fas* in human and *arm* in Drosophila RNAs, in both transcripts, the presence of Elav promotes the exclusion of the alternative exon whereas its absence leads to exon inclusion. In the human *Fas* transcript, a sequence within the alternative exon was demonstrated to be needed for Elav regulation of splicing (Izquierdo, 2010). In our system, however, no U rich sequence that could be a candidate for Elav binding site exists within the microexons, but an informatics approach carried out by Pedro Patraquim in the Alonso Lab has revealed some candidates sequences within the introns. The functional analysis of such sequences is the subject of the next chapter.

### 3.3.2 Elav affects *Ubx* polyadenylation decisions

In the same way, at earlier stages of embryonic development, the polyadenylation site prefered is the proximal one. Later on, the distal PAS is also used. The presence of Elav promotes the use of the distal polyadenylation site, generating a longer 3' UTR, characteristic of the neural *Ubx* mRNA (Smibert et al., 2012). It has been demonstrated that a longer 3' UTR allows a more fine tuning of the regulation of the mRNA molecule, by various mechanisms, like miRNAs, ARE binding proteins, among others (Flynt and Lai, 2008), (Mayr and Bartel, 2009), (Thomsen et al., 2010), (Bava et al., 2013), (Li and Lu, 2013). Elav also demonstrated to play a role in PAS choice of a subset of transcripts in the CNS (Hilgers et al., 2012). The difference observed in the production of a long 3' UTR by usage of the distal polyadenylation site between the wild type and the $elav^5$ mutant is about 30%. This suggests that Elav is not the only factor involved in this process as observed in other cases, such as Dm *ewg*.

### 3.3.3 Elav affects total *Ubx* RNA and protein levels

An observation made during the course of these experiments is the reduction of total *Ubx* mRNA in the $elav^5$mutant. *In situ* hybridisation experiments with intronic probes done in the Alonso Lab by Joao Picao show an accumulation of non released

RNAs from the site of transcription in the chromosome which can explain the reduction of mature *Ubx* mRNA in the $elav^5$mutant. This suggests that the role of Elav in the correct splicing of *Ubx* is very important for the expression of this gene, such that in its absence the transcript is not properly released from the transcription site. This raises the question of what happens then in non neuronal tissues that do not express Elav normally. Are other proteins covering the role of Elav in non neuronal cells? If so, what is the identity of such proteins? What are the molecular mechanism that allows proper release of the spliced and polyadenylated transcript? At the protein level, the Ubx isoform pattern observed in the $elav^5$mutant compared to the wild type is changed in the same direction as the observed at the RNA level but is not at the same extent. The differences between $elav^5$mutant and wild type patterns are more evident at the RNA level. This might be derived by differential translation efficiency of each isoform contrary to previous reports (Subramaniam et al., 1994). Alternatively, this discrepancy could be explained by translational delay: at earlier stages, Ia isoform is more abundant and IVa is less abundant.

### 3.3.4 Elav modifies *Ubx* RNA patterns during embryonic development

The change in the pattern of *Ubx* isoforms observed between the wild type and the *elav* mutants is also observed in earlier stages: the isoform IVa appears later in development (stage 11) as observed in Figure 3.2. However, the effects at earlier stages are more difficult to study because the Elav protein is less abundant in the control stock.

### 3.3.5 The relationship between splicing and polyadelylation

The experiments performed in this chapter show that Elav effects both splicing and polyadenylation processes in *Ubx* RNA processing. Previous studies reported Elav/HuR involvement in splicing independently of polyadenylation in human RNAs (Izquierdo, 2010), but no previous reports have demonstrated Elav/HuR involved in both processes in the same mRNA molecule. This leads to the question if Elav could be the factor involved in the coupling of these two events, or if these are two separated

and independent events. Given the long size of the *Ubx* transcript, it is possible to consider that the effect on splicing and polyadenylation occur independently. However, previous studies demonstrate that the mRNA molecule can be folded and sequences far away can be brought together. So there is a possibility that Elav protein serves as a direct link for these two regulatory events. Moreover, the role of the RNApol-II CTD as a platform for regulatory proteins could provide a anchoring point for Elav.

Our results show that Elav has a partial effect on splicing and polyadenylation of *Ubx* RNA. These results can be explained by the use of whole embryos instead of dissected CNS in the experiments done here. Another explanation could be that the alternative splicing and alternative polyadenylation decisions in *Ubx* transcripts are regulated by more factors, not Elav only. It is possible that the combination of a number of factors is needed in certain proportions to change the use of splicing and polyadenylation sites. The exploration of other potential candidates for *Ubx* RNA processing is the subject of the 5th chapter.

# Chapter 4

# Elav binds *Ubx* mRNA directly

## 4.1 Introduction

The HuR/Elav family of proteins have a crucial role in mRNA processing (Lisbin et al., 2001), (Soller and White, 2003), (Zhu et al., 2007), (Dai et al., 2012); RNA stability (Ratti et al., 2006) and even transcription and translation (Antic and Keene, 1997), (Meng et al., 2005), (Rivas-Aravena et al., 2009). How can an RBP play so many diverse roles? The answer might lie in its biochemical properties, in the structure of the protein and its ability to recognise and bind its targets (Figure 4.1).

The HuR/Elav family of proteins contains three RNA recognition motifs (RRM), two of them close together and the third one separated by a hinge that contains a Nuclear localisation signal (Samson, 2008). The RRM domain is a common feature of RNA binding proteins and it is composed by around 80 aa containing two ribonucleoprotein motifs of eight (RNP-1) and six (RNP-2) amino acids. This domain is arranged in four beta-sheets packed against two alfa helices (Gamberi et al., 2006), (Benoit et al., 2010) (Figure 4.1C). Crystallography studies in the human proteins have shown that the RNPs form a cleft where the RNA molecule fits (Benoit et al., 2010) (Figure 4.1C). In mammals, the first RRM has the main RNA binding activity and RRM2 has a supporting role. The third domain is separated by a 60-87 amino acids hinge. In*Drosophila* the main role of the third domain is the multimerisation of the protein (Toba and White, 2008).

The mammalian homolog has an stimulatory effect in translation, by enhancing RNA circularisation. The human HuD interacts with the polysomes, the RRM3

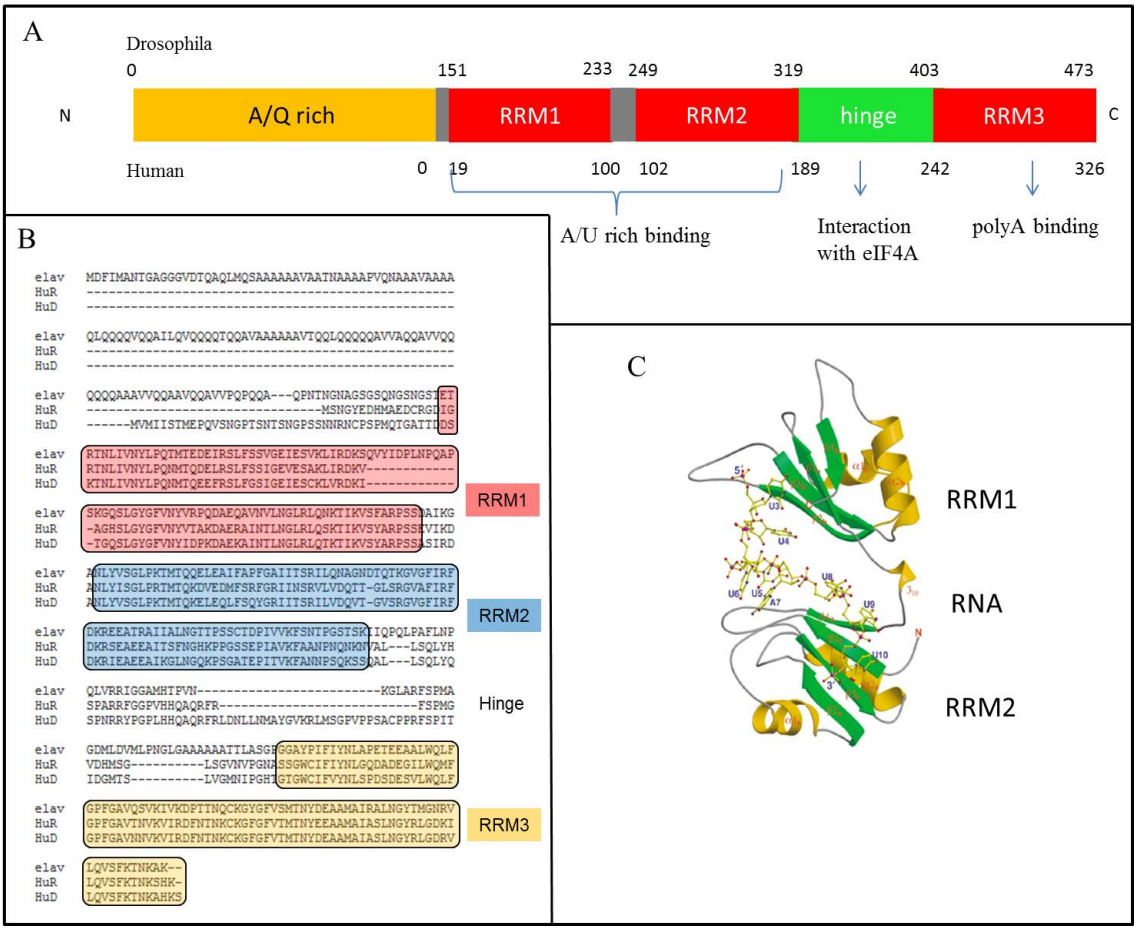


Figure 4.1: Elav protein structure. A) Scheme of the Elav protein domains. The numbers mark aminoacid number position of the domains at the top for *Drosophila* and below for the human homolog. Some of the described roles for the domains in humans are listed below (Toba and White, 2008), (Fukao et al., 2009). B) Aminoacid sequence alignment of *Drosophila* elav and human HuR and HuD proteins showing a very high conservation particularly in the RRM domains. C) Ribbon representation of the structure of RRM 1 and 2 making contacts with human *c-fos* ARE RNA molecule (Figure taken from (Wang and Hall, 2001a).

domain interacts directly with eIF4A and the hinge binds polyA sequences (Fukao et al., 2009) (Figure 4.1). The RRM domains of the Drosophila and human Elav proteins are very well conserved (Fig 4.1B), opening the possibility of Drosophila homolog interacting with the translation machinery. The hinge, however, shows less conservation between the Drosophila homolog and HuD, reducing the probability of any role of Drosophila Elav in translation.

In the interaction between RNA and proteins, secondary structure of the RNA is crucial. Studies using computational RNA secondary predictions have been able to find common sequences and structures used by proteins to identify their targets (Kazan et al., 2010), (Li et al., 2010), (Mauger et al., 2013), (Dieterich and Stadler, 2013). The RBP HuR is no exception, and extensive studies have proposed a putative consensus sequence for this protein. Unlike other RNA binding proteins, the Elav/HuR family has a very loose recognition sequence, the only condition is the presence of A/U or U stretches (López de Silanes et al., 2004),(Wang and Hall, 2001b).

Crystallographic experiments using the first two RRM domains of human HuD and 11nt RNA fragments from human *c-fos* ARE reveal that the minimal sequence requirements are Uracils in positions 3, 4, 5 and 8, 9 and 10 (Wang and Hall, 2001b), (Benoit et al., 2010) (Figure 4.1C). Biochemical experiments have determined that human HuR has similar minimal requirements. These results are supported by RNA-immunoprecipitation experiments followed by cDNA array hybridisations (López de Silanes et al., 2004). RNA/protein *in vitro* assays (EMSAs) have demonstrated that the HuR protein requires at least partly single stranded RNA with two stretches of 8 or 9 Uracils (Barker et al., 2012), (Benoit et al., 2010) and that substitutions for Gs are the most disruptive mutations for interactions of HuR with target RNAs.

In previous experiments, I determined that the Elav protein has a role in the *Ubx* splicing and polyadenylation. To determine the mechanism used by this protein on *Ubx* RNA processing, I started a series of experiments with the aim to determine if Elav binds directly to the *Ubx* RNA. The strategy used was *in vitro* RNA/protein binding assays: UV crosslinking and Electrophoretic Mobility Shift Assay (EMSA) as reported previously by Soller Lab (Soller and White, 2003). Later, I analysed the secondary structure of the sequences tested by *in vitro* binding experiments to

identify any common features. In addition, the sites were amplified and cloned using two different lengths for each site, to test wether different nucleotide context affects Elav binding. The Elav used was a recombinant protein produced and purified in the Alonso Lab by Ali Mumtaz using a plasmid kindly provided by Matthias Soller and following the procedures described in the Methods section, taken from (Soller and White, 2003).

We used the previously described *ewg* sequence as a positive control for binding and its antisense as a negative control (Soller and White, 2003). The RNA for each sequence to test was labelled by *in vitro* transcription using $^{32}$P-UTP. The probes were gel purified and quantified by scintillation method by Ambion (Methods).

## 4.2 Results

### 4.2.1 Identification of putative Elav binding sites within the *Ubx* locus

The first step was to search using bioinformatics tools, potential sites within the *Ubx* transcript for Elav binding sites. This work was carried out by Pedro Patraquim and took into consideration the previously described Elav binding sites in *erect wing* (Soller and White, 2003) and *neuroglian* transcripts (Lisbin et al., 2001), as well as AU-rich sequences. Using this strategy, many sites were identified and subjected to an evolutionary conservation analysis in the understanding that conservation can indicate functionality (Figure 4.2).

The sequences used as reference from *ewg* are AATTTTTT and CATTTTTT (Soller and White, 2003) and from *ng* are TTTTTGTTGT, TTGTTTTTTT, TTTG TTTTT, TTTTATTTAT and TTTTTTTT (Lisbin et al., 2001). AU rich sequences in general were also considered. The first search resulted in seventeen possible binding sites. Phylogenetic analysis including *Ubx* sequences conserved in distant drosophilids species (ultraconserved) resulted in only five ultra-conserved sites: Sites 3, 5, 8, 13 and 16 (Figure 4.2B). The *Drosophila* species used are: *D. simulans, D. sechelia, D. yakuba, D. erecta, D. ananassae, D. pseudooscura, D. persimilis, D. willistoni, D. mojavensis, D. virilis and D. grimshawi.*

Site 3 is located in the first intron in position 9276, 45 nt upstream the microexon

I. This makes it a very strong candidate for an Elav binding site to control alternative splicing. The sequence of this site is related to the *erect wing* site: ATTTTTT and is perfectly conserved in all the species tested except for 3 nucleotides in *D. willistoni* (Figure 4.2A and B).

Site 5 is also an interesting candidate, it is located in the third intron, at position 24134 and 68 nt downstream the microexon II. The sequence of this site is more *neuroglian* like: TTTGTTTTT and it is perfeclty conserved until *D. ananassae*. *D. mojavensis, D. virilis and D. grimshawi* show an 8 bp insertion (Figure 4.2A and B).

Site 8 is located in the third intron, in position 32460, 8352 nt downstream mII. Interestingly, 40 nt upstream, there is an identical sequence that is not conserved. The sequence for both sites is TTTGTTTTT, more *neuroglian*-like. The conserved site is perfectly conserved except for two nucleotides in *D. willistoni* and *D. virilis* (Figure 4.2A and B).

Site 13 is a *erect wing* related sequence (AATTTTTT), perfectly conserved among the Drosophilids except in *D. willistoni*. It is located in exon III in position 53583; 21,958 nt upstream the 3' Exon, 21,166 nt downstream the Site 8. Its position within the largest intron of *Ubx* (51,476 nt), around the middle, makes it an unlikely site for the control of the alternative splicing of mII. However, it is very conserved and is relatively close (3,547 nt downstream) to a resplicing site, making it an interesting site for splicing control in general (Figure 4.3).

The resplicing model of *Ubx* (Burnette et al., 2005) is the splicing of large introns in two sequential steps, where an upstream resplicing site is used as an intermediary. In the *Ubx* resplicing model implies that the splicing of the intron 1 and joining of E5' with mI regenerates a splicing donor (Figure 4.3 step 1), that can be used or not. If it is used, then mI is spliced together with intron 2 (Figure 4.3 step 2a). If is not used, then intron is spliced alone (Figure 4.3 step 2b). The regeneration of the splicing site also occurs in the removal of intron 2 and joining of E5' and mII (Figure 4.3 step 4a and b). For the last intron, the splicing mechanism is different. Intron 3 is extraordinarily large (51476 bp) and its splicing is divided in two steps: an upstream splicing (or resplicing) site is used, called RP3. This site allows the splicing of half of intron 3 and regenerates a donor site used in the second splicing

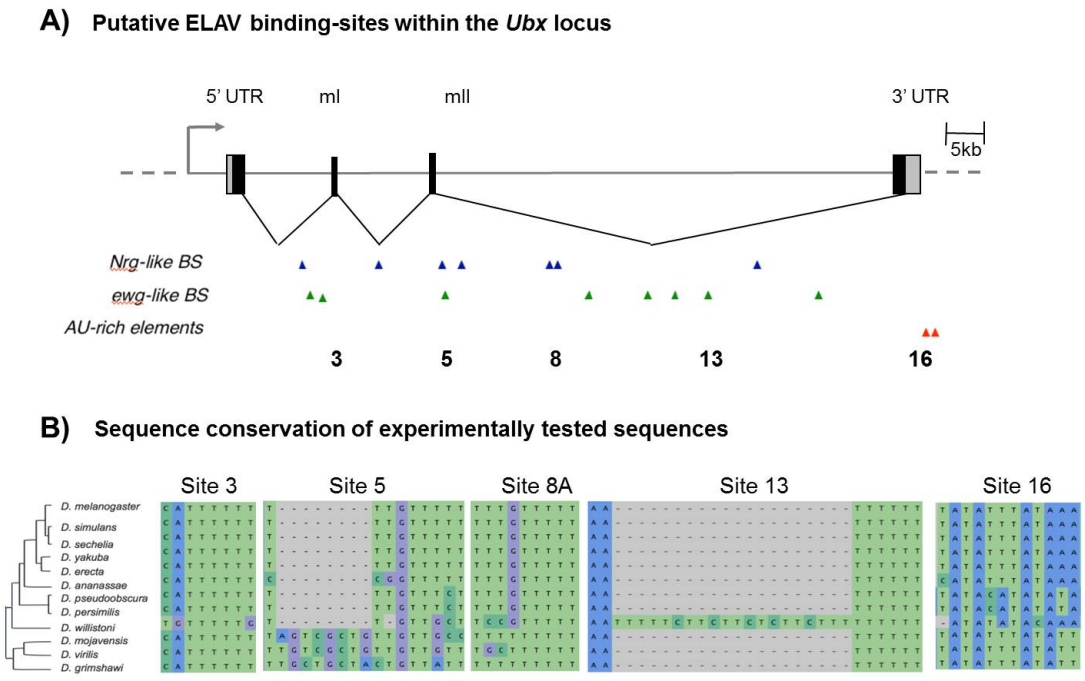


Figure 4.2: Putative Elav binding sites in the *Ubx* locus. A) Representation of the *Ultrabithorax* transcript, showing the exons in black boxes, the UTRs in grey boxes and the line represents the introns. The arrow represents the start of transcription. The arrowheads represent the positions of putative Elav binding sites and the numbers represents the ones being tested. *ng*: *neuroglian*; *ewg*: *erect wing*. B) Phylogenetic analysis of the putative sites showing ultra conservation in sites 3, 5, 8, 13 and 16.

reaction removing the rest of the exon (Figure 4.3 step 5 and 6.)

The last sequence (Site 16) found to be highly conserved is an AU rich sequence in the 3' UTR of *Ubx*. It is located in position 77261, 1418 nt downstream the stop codon and 457 nt downstream the proximal polyadenylation signal. This makes it a very interesting candidate site for polyadenylation control, following the model of *ewg* and *ng*. This sequence has two possible Elav binding sites: TATATTTAT and ATATATTTATCTA separated by 10 nucleotides. Secondary structure analysis of this sequence reveals a high similarity to a sequence demonstrated to be recognised by Elav in the Msi1 3' UTR (Ratti et al., 2006) (Figure 4.2A and B).

Analysis of sequences suggests that the most likely factor to be involved in splicing is Site 3, which is only 45 bp upstream microexon I whereas site 8 is more than 8 kb downstream microexon II and site 13 is 24 kb downstream microexon II. This means that site 3 has a very high potential by its position, to have a role in the

splicing decision. Site 8, being so far into the third intron, has less possibilities of playing a role, however, we cannot exclude this possibility. Site 16 is very close to the proximal polyadenylation site, making it a very strong candidate for an Elav binding site involved in polyadenylation site choice (Figure 4.2).

### 4.2.2 Elav binds directly to *Ubx* RNA in EMSA experiments

To test wether Elav binds *Ubx* directly, increasing amounts of the purified protein (0, 200, 400 and 600 nM) were incubated with 100 pM of radiolabeled Ubx probes, following the protocol described in (Soller and White, 2003). The probes included the sequences of sites 3, 5, 8, 13 and 16 (relative position described in the scheme in Figure 4.4A), also including labelled *ewg* as positive control and labelled *ewg* antisense (*ewg* as) as negative control. The positive control experiment shows a rapid shift as the concentration of Elav is increased whereas the negative control shows a retarded shift with Elav increments. The quantifications were done using Image J software measuring the integrated density of the free RNA signal and substracting it to the input (as percentage) (Figure 4.4B and C). The EMSA shows a high affinity of Elav to the *ewg* sequence as reported before (Soller and White, 2003). The Elav protein clearly binds to the *ewg* sequences at 200 nM input (Figure 4.3 B, lane 2) and reaches 100% saturation at the highest concentrations tested (400 and 600 nM) (Figure 4.4B lanes 3 and 4). The *ewg* antisense sequence shows little RNA bound to the Elav protein at 200 and 400 nM Elav, and only at 600 nM there is significant RNA binding (Figure 4.4B, lane 8). These two probes behave as previously reported (Soller and White, 2003).

Site 3 shows high affinity, with most of the RNA bound to Elav at 200 nM protein, showing a profile similar to the *ewg* control (Figure 4.4B lanes 9-12). In contrast, site 5 shows little binding at all concentrations tested, and never reaches 100% of RNA bound to Elav (Figure 4.4B lanes 13-16).

Site 8 also shows a profile similar to *ewg*, having most of the RNA bound to Elav about 200 nM Elav (Figure 4.4B lanes 17-20). Site 13 shows an intermediary binding profile, having significant % of RNA bound at 400 nM Elav, and reaching 100% RNA bound at 600 nM (Figure 4.4B lanes 21-24). Site 16 showed the lowest

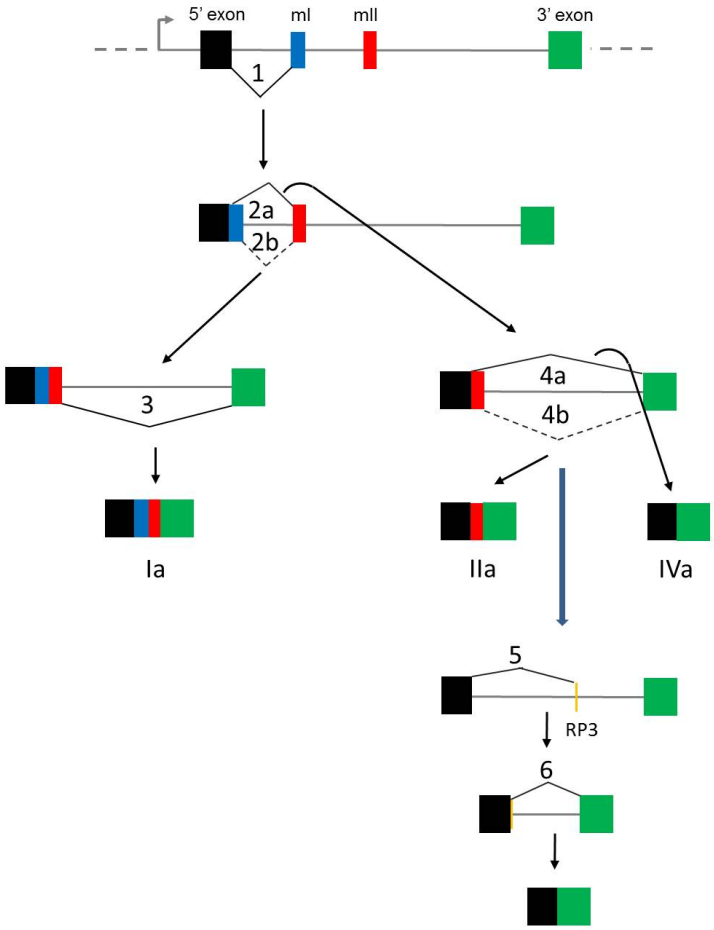


Figure 4.3: Ubx resplicing model. Exons are represented in colored boxes, black for the 5 exon, blue for microexon I, red for microexon II and green for 3 exon. The resplicing model is a sequential splicing model where a 5 splicing donor site is regenerated after the previous splicing event. In this model, the first event is the splicing of Intron 1 (1). The joining of 5 exon and mI regenerates a splicing donor site that can be used (2a) or not (2b). The splicing of intron 3, in turn, regenerates a splicing donor that can be used (4a) or not (4b). There is evidence that suggests that Intron 3 itself undergoes a resplicing process: the splicing of the last intron is achieved by two steps, the first splicing event (5) uses a recursive splicing site (RP3) located within the intron. The use of this site (5) regenerates a splicing donor site allowing the splicing of the second half of the intron (6). Models taken and modified from (Hatton et al., 1998) and (Burnette et al., 2005).

affinity, having almost no RNA binding at all concentrations tested (Figure 4.4B lanes 25-28). The quantification of the % of RNA bound for each probe is shown in Figure 4.4 C.

Measuring and plotting the % of bound RNA to the protein shows 100% binding at 600 nM of Elav and shows a high percentage of binding at low Elav concentrations; this is observed in of *ewg*, site 3 and site 8 probes (Figure 4.4C. The rest of the probes do not always reach 100% binding of RNA to Elav and the percentage of binding remains low at all the Elav concentrations tested compared to the positive control (site 5, 13, 16 and ewg as in Figure 4.4C).

The binding curve for the antisense of *ewg* behaves as previously reported, showing low affinity at lower Elav concentrations but reaching near 100% at the highest Elav concentration tested (600 nM) (Figure 4.4 B and C). The EMSA experiment allows for the observation of protein/RNA complexes of different size and we observe complexes of different sizes in the *ewg*, site 3 and site 8 probes. This suggests that the Elav protein is able to assemble RNA/protein complexes with a different number of multimers.

The probes containing the putative Elav binding sites used in these experiments are of different size. To test if the length of the probe affects the binding properties of the sites, a set of probes of around 150 nt were constructed and tested as described above using recombinant Elav (Figure 4.5).

The EMSA experiments show similar results, the binding curve did not change significantly when shorter sequences were used in the binding assays. However, using shorter sequences we can observe some intermediate protein/RNA complexes formed (arrows in Figure 4.5A). *ewg* shows at least 6 different size complexes, site 3 shows 5 and site 8 shows 7 different size complexes.

As in the previous experiments, *ewg* probe shows binding from 200 nM Elav onwards, (Figure 4.5A lanes 1-4) and the *ewg* antisense showed binding only at 600 nM Elav (Figure 4.5A lanes 5-8). Site 3 shows RNA/protein complex forming from 200 nM Elav and an intermediate complex is observed (Figure 4.5A lanes 10-11). Site 8 shows also a high affinity as in the experiments with the longer probes, but now two intermediate RNA/protein intermediates are observed (arrows in Figure 4.5A, lanes 13-16). The previous experiments even with the shorter probes as with

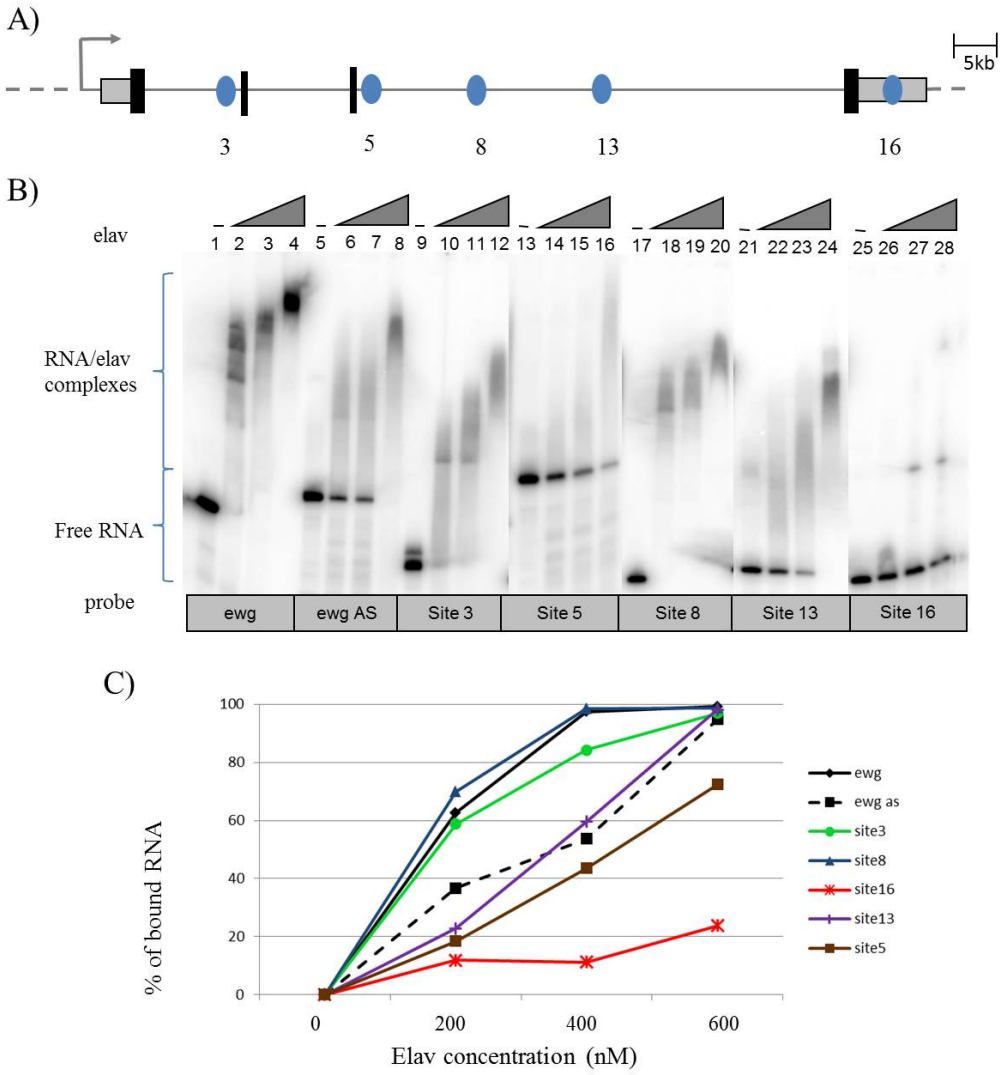


Figure 4.4: Elav interacts directly with *Ubx* RNA: A) Scheme representing *Ubx* transcript. Black bars represent coding exons, gray boxes, 5' and 3' UTRs, lines represent the introns. Blue ovals mark the position of the predicted Elav binding sites. B) EMSA experiments were carried out as described in the Methods section using 100 pM of radiolabeled RNA probes for *ewg*, *ewg* antisense, sites 3, 5, 8, 13 and 16 and 0, 200, 400 and 600 nM of recombinant Elav. Representative experiment from three replicates. C) Quantification of the % of input RNA bound by Elav protein.

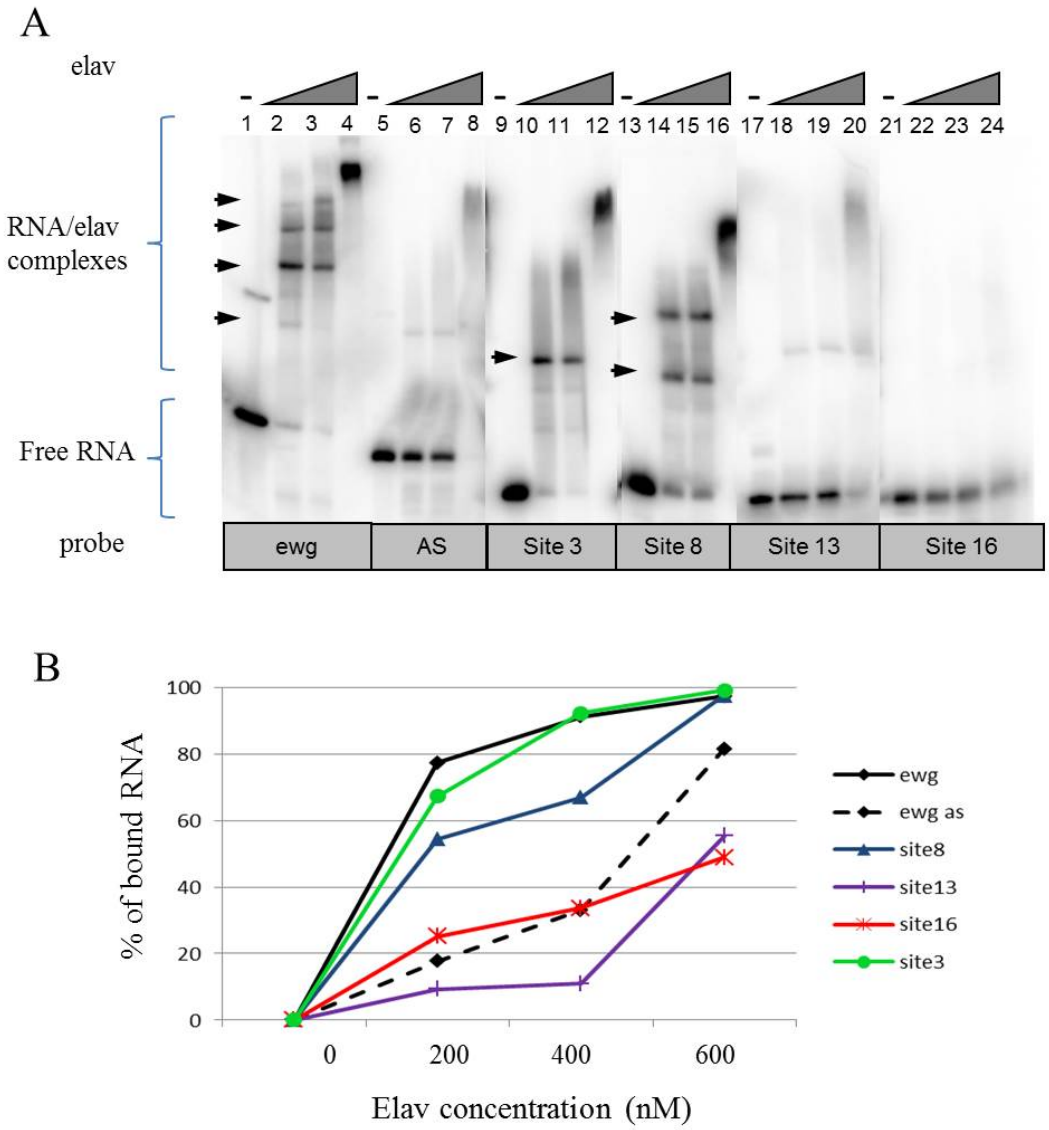


Figure 4.5: Probe size does not affect RNA/Elav interaction. A) EMSA testing 100 pM of radiolabeled RNA probes for *ewg*, *ewg* antisense, sites 3, 8 13 and 16 at 0, 200, 400 and 600 nM of Elav, showing that the ewg, site 3 and site 8 have high affinity for Elav. Sites 13 and 16 show even less affinity that the ewg antisense probe. Representative experiment form three independent replicates. B) Quantification of the % of RNA bound by recombinant Elav protein. Arrows show intermediate RNA/protein complexes.

sites 13 and 16 show low affinity with partial RNA binding observed at 600 nM Elav (Figure 4.5A lanes 17-24).

In order to demonstrate that the suspected sequences are responsible for the binding, I performed directed site mutagenesis to the probes that showed high affinity for Elav binding; alternate T residues were substituted by G residues (Figure 4.6).

The mutated sites were prepared in the 150 bp probes and tested by EMSA experiments as described. Site 3 mutated in the shorter probes shows a reduction of affinity for Elav and loss of the intermediate RNA/Elav complex formed at 200 and 400 nM Elav observed with the site 3 wt (Figure 4.7A lanes 9-16). This suggests that ATTTTT only site in this sequence is responsible for the formation of the intermediate complexes. Sites 13 and 16 serve as well as the antisense of *ewg*, as negative controls here. The quantification of the free RNA shows a binding curve similar to the ewg antisense probe (Figure 4.7B).

For the site 8, mutations in site A, site B and mutants in both A and B were tested. The mutations are described in Figure 4.6. When site 8A is mutated, there is less RNA bound by Elav at 200 and 400 nM when compared to the wild type probe (Figure 4.8A lanes 9-12 and 13-16, and quantification in panel B). The formation of one of the RNA/Elav complexes is also disrupted (higher arrow in Figure 4.8A). In contrast, when site 8B is mutated, there is also a reduction on the total RNA bound by Elav (Figure 4.8B) and the formation of the intermediate complexes is also disrupted. The higher molecular intermediate complex is almost absent and the lower complex is reduced (both complexes marked with arrows). In addition, both minor complexes marked with arrowheads are also absent when site 8B is mutated. (Figure 4.8A).

The formation of several complexes of different molecular weight were observed in sites 3 and 8. Figure 4.9 compares the intermediate complexes formed with the *ewg* probe in previous experiments from the Soller Lab and the ones observed in the experiments carried out in this project (Soller and White, 2005). Site 3 forms an intermediate complex and two less abundant ones, marked with arrow and arrowheads respectively in Figure 4.9. In contrast, two intermediate complexes are observed in site 8, and three minor ones (Figure 4.9).

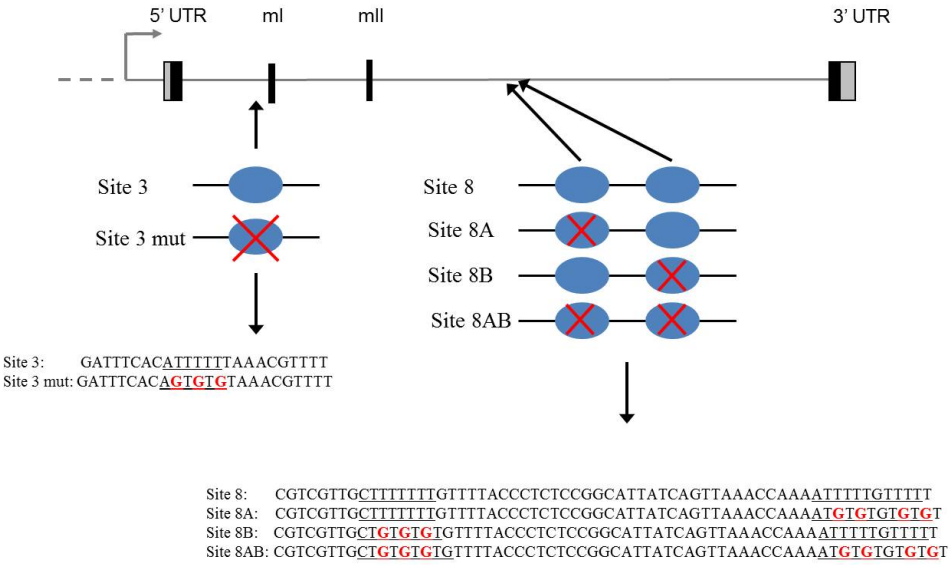


Figure 4.6: Mutagenesis of sites 3 and 8. Representation of the *Ubx* transcript. Arrows mark the position of the putative Elav binding sites 3 and 8. Blue ovals represent the binding sites and the red cross represent the binding site mutation. Wild type sequence of sites 3 and 8 and the mutagenesis is shown below. The putative Elav binding sites sites are underlined and the changed nucleotides are in red.

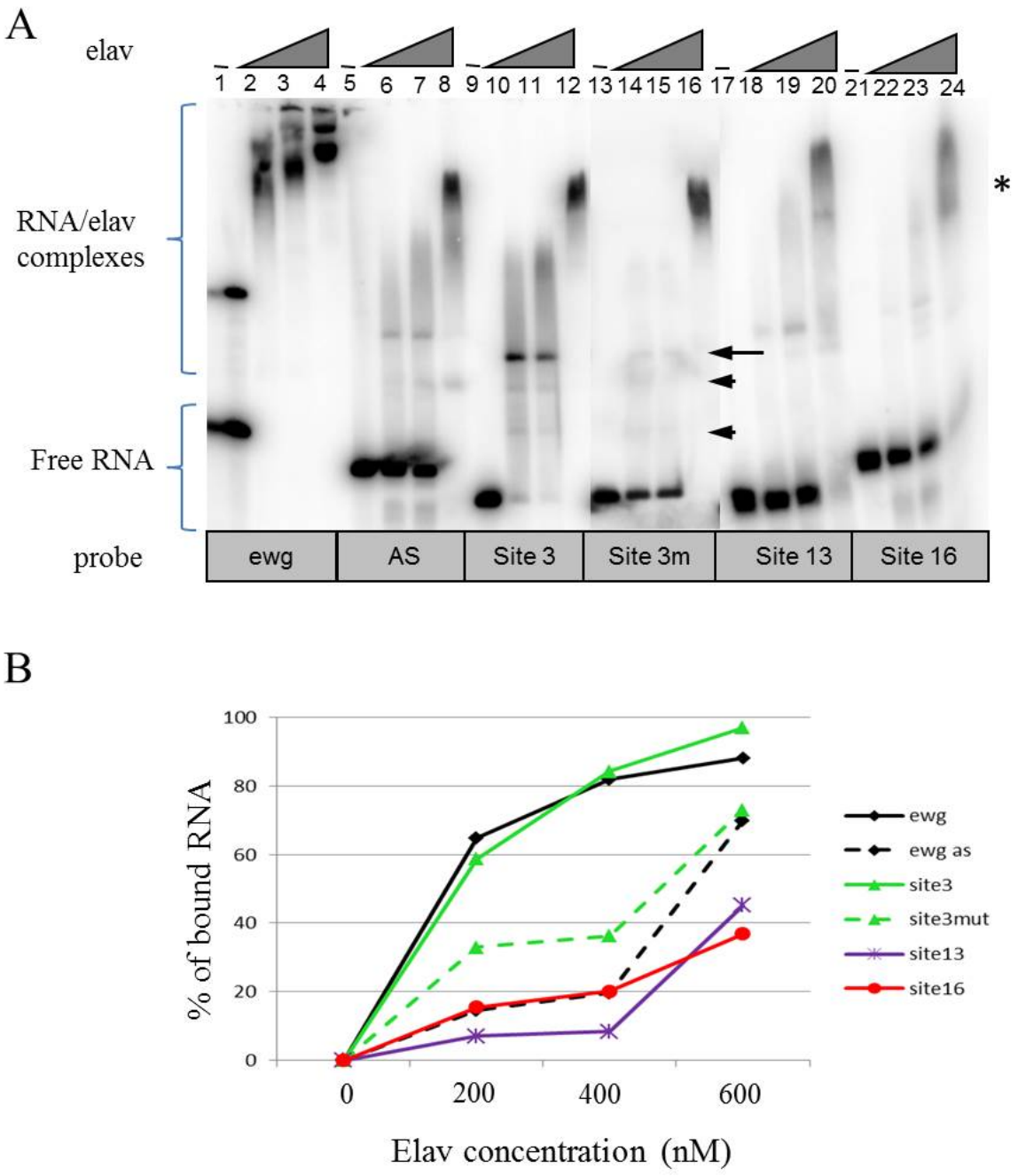


Figure 4.7: Mutation of potential binding site 3 partially disrupts interaction with Elav protein. A) EMSA testing 100 pM of radiolabeled RNA probes for *ewg*, *ewg* antisense, sites 3, 3mut, 13 and 16 for binding to Elav protein at 0, 200, 400 and 600 nM of Elav. The mutant shows an important reduction in affinity to Elav from 200 nM. Sites 13 and 16 serve as negative control together with the AS of ewg. Representative experiment from three independent replicates. B) Quantification of the % of RNA bound by Elav protein. The asterisk marks the complexes of higher molecular weight.

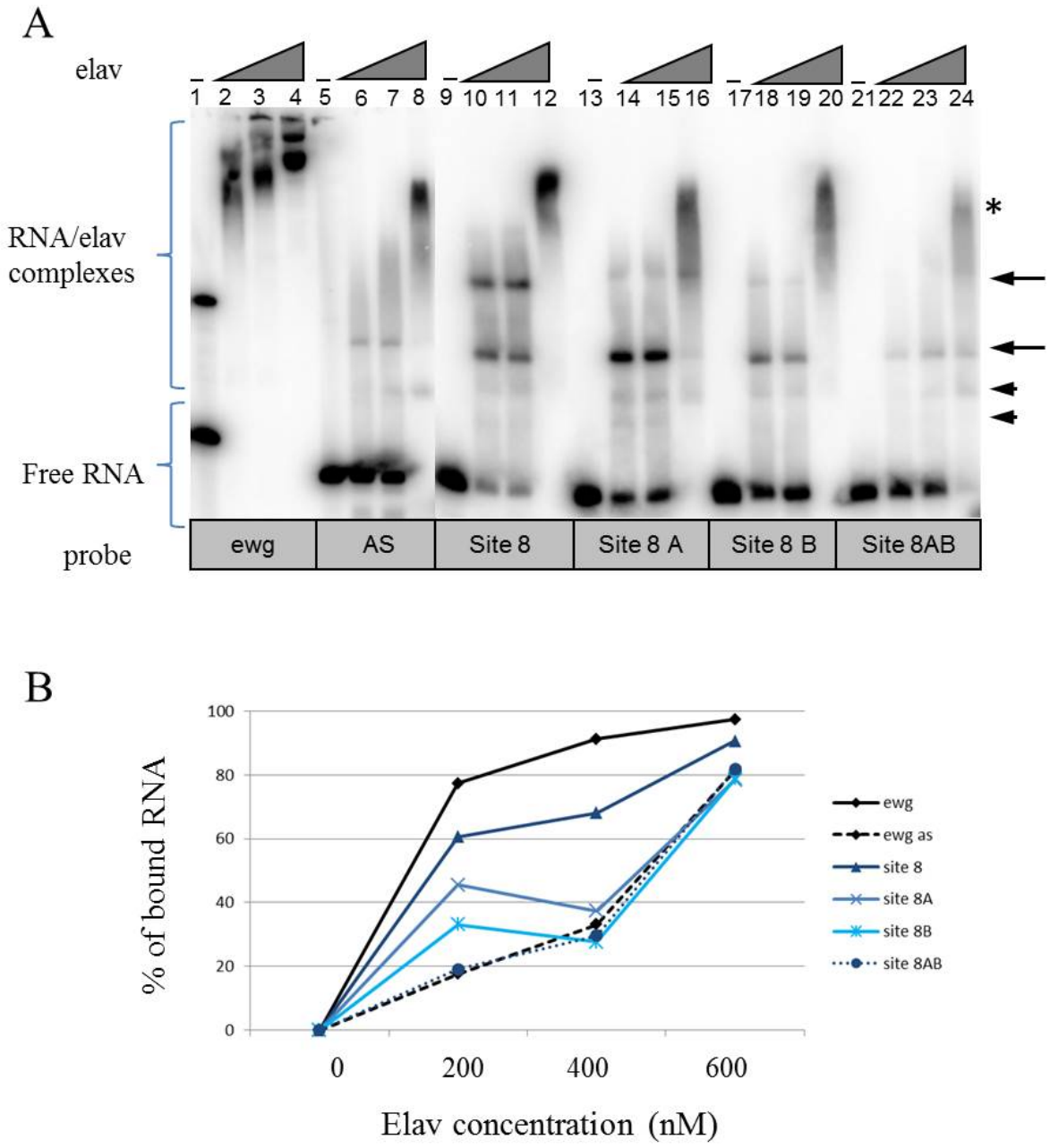


Figure 4.8: Site 8 mutation partially disrupts Elav interaction. A) EMSA testing 100 pM of radiolabeled RNA probes for *ewg*, *ewg* antisense, site 8, 8A, 8B and 8AB with recombinant Elav at 0, 200, 400 and 600 nM of Elav. Individual mutations show an important reduction of affinity from 200 nM of Elav and when both sites are mutated in the same probe, the affinity is reduced even more. Representative experiment from three independent replicates. B) Quantification of the % of RNA bound by Elav protein. The arrows mark the RNA/Elav complexes and the asterisk marks the complexes of higher molecular weight.

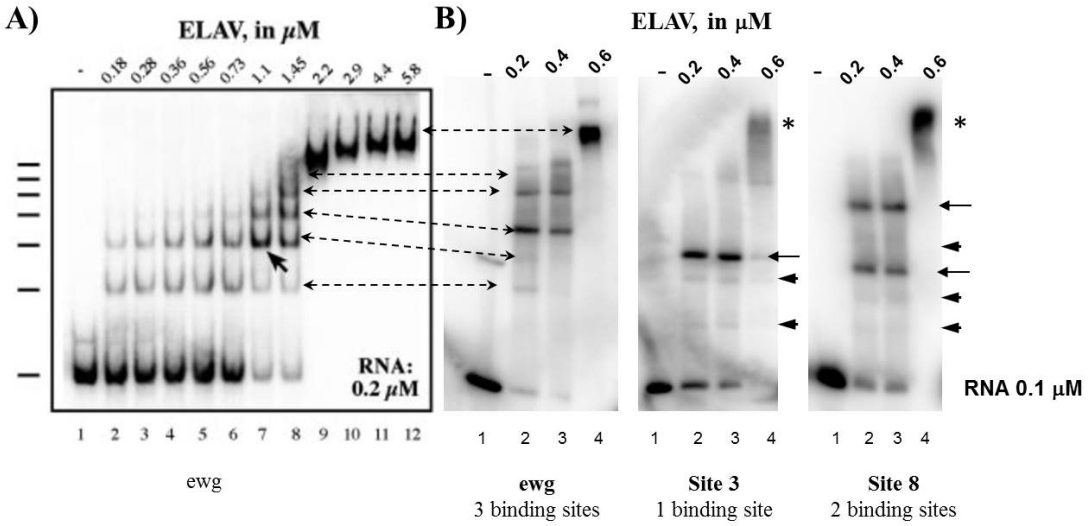


Figure 4.9: Elav and *Ubx* RNA forms complexes of different sizes. A) EMSA taken from (Soller and White, 2003), Figure 3A showing the different Elav/RNA complexes formed by Elav multimerisation. B) Representative EMSA experiments for *ewg*, site 3 and site 8 showing the major intermediate RNA/protein complexes marked with arrows and less prominent complexes marked with arrowheads.

### 4.2.3 Elav binds directly to *Ubx* RNA - UV crosslink/RNAse protection assay

To corroborate the results obtained by EMSAs, I decided to test RNA/Elav binding using an independent method. To this end, the UV crosslink/RNAse protection assay was performed, including the positive and negative controls , utilising probes incorporating sites 3, 8, 13 and 16. For these experiments, radiolabeled RNA was incubated with Elav protein *in vitro*, crosslinked with UV light and then RNA was digested with RNAse A. The samples then were resolved using a standard 12% polyacrylamide gel and exposed to a phospho-imager plate overnight. The probes that were protected by Elav binding were not digested and therefore could be detected as a band in the gel. A non UV croslinked control was included for each probe. First the long probes were tested and the experiments show a very evident protected band for the *ewg* probe, site 3 and site 8, but no band was observed in *ewg* antisense. In addition, no protected fragments were observed for sites 13 and 16 (Figure 4.10B and C). The UV crosslinks showed consistent binding of *ewg* and lack of binding for its antisense sequence. Sites 3 and 8 showed also very strong binding supporting the presence of a functional Elav binding site. Sites 13 and 16

did not show any binding to Elav in this experiment confirming the observations above using the EMSA approach.

The 150 bp probes were also tested also by UV crosslink and the results were similar to the experiments carried out with the longer probes. *ewg*, site 3 and 8 showed a significant binding to Elav protein whereas *ewg* antisense and site 16 did not (Figure 4.11). The UV crosslink experiments showed a much more marked difference between the high affinity group of sites and the low affinity group described in the EMSA experiments. In the UV crosslinking experiments, sites 13 and 16 shows a similar binding activity to the *ewg* antisense, of about 10% whereas site 3 and 8 showed a very high affinity, of around 60 and 70% respectively.

Next it was tested if the mutations in sites 3 and 8 showed a decrease in Elav protein binding as observed by EMSAs (Figure 4.12).

The site 3 mutant showed a large reduction in affinity for Elav, with only a weak band observed compared to the wild type probe, where a strong interaction is observed. In the case of the site 8, mutation of site A has only a minor impact on its affinity. Mutation of site B has a bigger impact on Elav binding but still this probe shows high affinity for Elav. Only the double mutant shows a marked reduction of Elav affinity (Figure 4.12).

The UV crosslink of Elav with the long and short probes demonstrated that the affinity of Elav for the RNAs offered did not change significantly with probe length. This suggests that under these assay conditions, the length of the probe does not influence the binding activity, and that the sequence responsible of the binding was not lost in the shorter sequences. These results were consistent with those obtained previously with *ewg*/Elav (Soller and White, 2003). In addition, the work described here shows that sites 3 and 8 are able to interact with Elav protein and that they also form intermediate complexes of different sizes.

We observed that the effect of the mutation in site 3 and Elav binding is very pronounced, decreasing the binding to Elav from 75% of the wt to 25% in the mutant (Figure 4.12C). Interestingly, for site 8, the individual mutations did not have a very strong effect in Elav binding, going from 60% of the wt to 55% and 50% in mutant 8A and B respectively. However, the double mutation decreases the binding to 20% of the wild type probe. These data suggest that the Elav protein uses the two sites

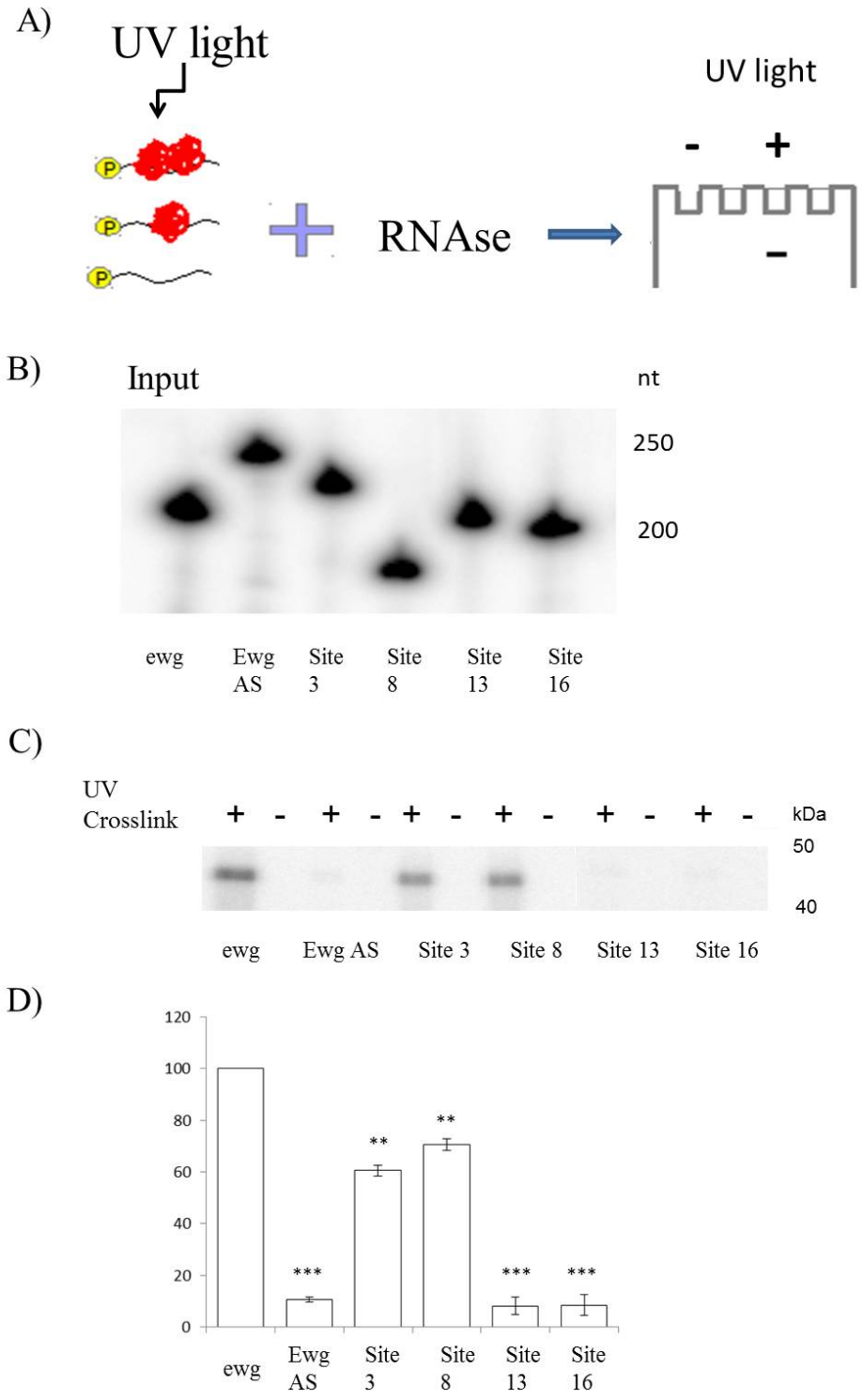


Figure 4.10: UV crosslink confirms interaction of Elav protein with *Ubx* mRNA. A) Scheme of the UV crosslink strategy. The line and the yellow P represent the radiolabelled RNA. The red figure represent Elav protein. B) Radiolabeled RNA probes input for UV crosslink of *ewg*, *ewg* antisense, site 3, 8, 13 and 16. C) UV crosslink showing Elav-dependent protected RNA fragments of *ewg*, site 3 and site 8 probes. No UV crosslink serves as a control for RNAse digestion. D) Quantification of the protected RNA plotted as % of input showing high affinity of ewg probe and sites 3 and 8. ewg AS, and sites 13 and 16 show a very reduced binding to Elav. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to *ewg*.

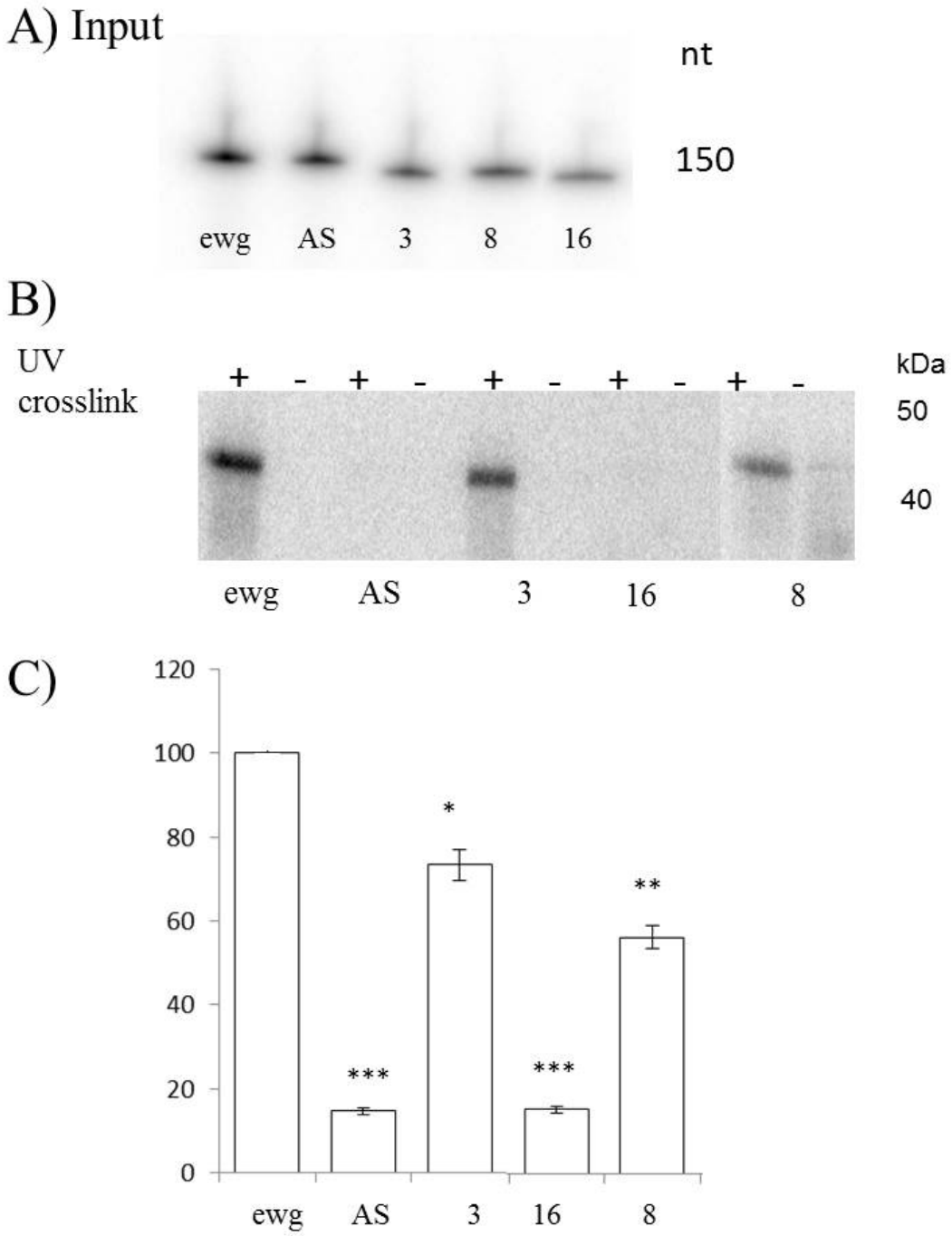


Figure 4.11: Probe size is not important for Elav interaction with putative Elav binding sites. A) Input radiolabeled RNA probes for UV crosslink of *ewg*, *ewg* antisense, site 3, 8 and 16. B) UV crosslink showing protected RNA fragments of *ewg*, site 3 and site 8 probes. No UV crosslink serves as a control for RNAse digestion. C) Quantification of the protected RNA plotted as % of input showing Elav protection of ewg probe and sites 3 and 8 whereas ewg AS and sote 16 show very little binding. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to *ewg*.

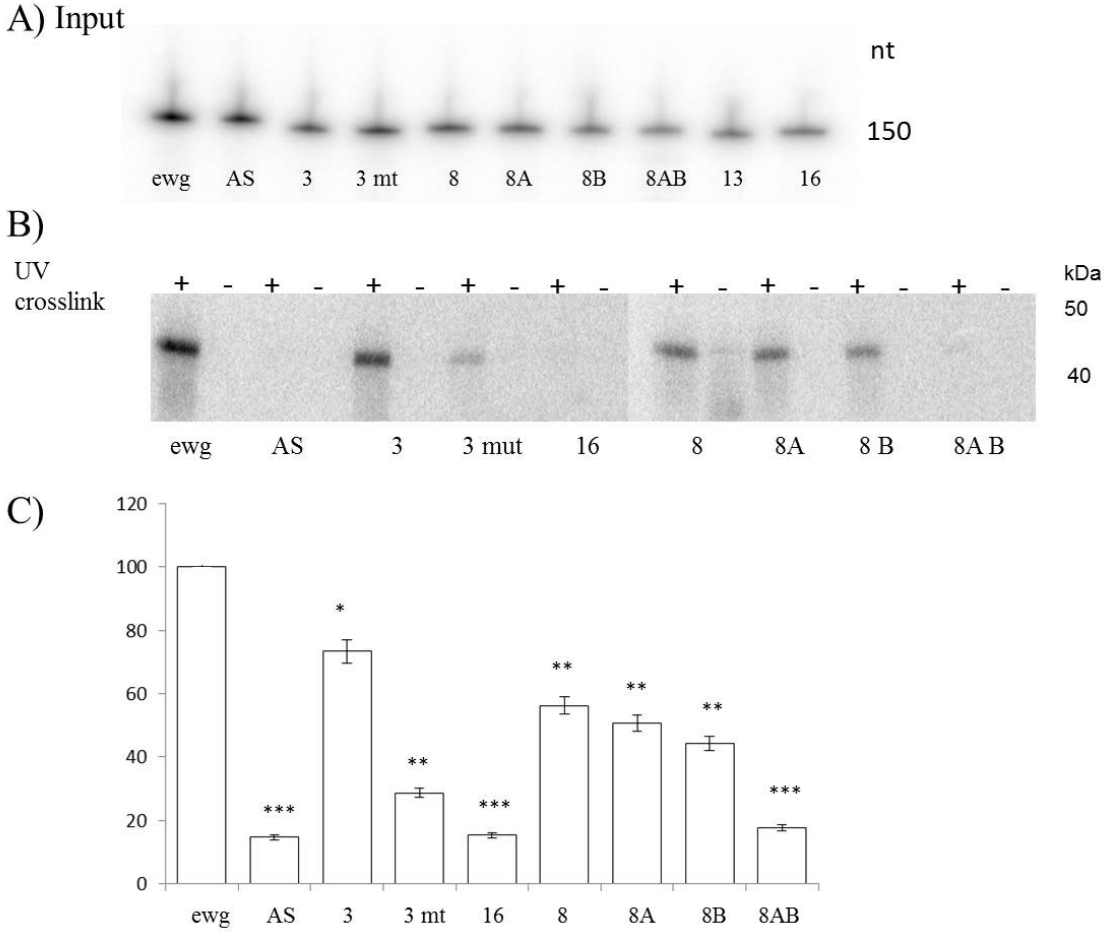


Figure 4.12: Mutations of sites 3 and 8 partially decreases RNA/Elav interaction. A) Input radiolabeled RNA probes for UV crosslink of *ewg*, *ewg* antisense, site 3, 3mutated, 16, 8, 8A, 8B, 8AB. B) UV crosslink showing Elav-dependent protected RNA fragments of *ewg*, site 3 and site 8 probes. No UV crosslink serves as a control for RNAse digestion. C) Quantification of the protected RNA plotted as % of input. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to *ewg*.

for an effective recognition of the site. This is as expected as it is the case for the Elav binding site in *ewg* (Soller and White, 2003).

The other sites tested show a very low affinity for Elav. However, site 13 shows the formation of an intermediate complex, suggesting that it might need another factor to be able to be recognised by Elav. Site 16 showed the lowest affinity among the sites tested. However, since recombinant purified protein was used here, we cannot discard the possibility of the need of a third partner for this interaction to occur. Supporting this possibility, the two putative Elav binding sites in the site 16 sequence is separated by a predicted polypyrimidine tract binding protein (PTB) binding site (Shen et al., 2004).

### 4.2.4 Role of RNA secondary structure in Elav binding

Previous experiments demonstrated that the accessibility of the putative binding site for protein has a very important role in the RNA binding protein interaction to its target (Li et al., 2010). Moreover, clustering of the putative binding sites within the secondary structure of a transcript increases the affinity of RNA binding proteins.

*Ubx* is not an exception here and the secondary structure of its transcript has been analysed by other members of the Alonso Lab (Patraquim et al., 2011). The RNAfold tool was used in order to identify potential secondary structure features within the sequences studied here by biochemical experiments. The *ewg* control sequence forms a predicted secondary structure with loops exposing the putative Elav binding sequence, marked with a red square in Figure 4.13. This particular array is very efficient in the recognition of Elav, as already demonstrated above by UV crosslink and EMSA analysis. This array presents three loops, two of them close together and the third a bit more distant. However, the predicted structure produced by the *ewg* antisense also presents three loops exposing AU rich sequences. The AU rich stretches do not have exactly the same sequence but they still form the loops possibly explaining why the antisense probe is able to bind Elav when used at high concentrations. In this array, there are two loops close together and a third one further away. This particular structure is demonstrated to have a very reduced affinity for Elav as demonstrated by the UV crosslink and EMSA experiments. This

is probably due to the longer distance of the third loop in respect to the other two loops, compared to what is predicted for the wt.

A table with the secondary structure observations is shown below:

| Site | Sequence identity | WT structure | Affinity | Mutant structure | Affinity |
|---|---|---|---|---|---|
| 3 | *ewg* like | loop | High | stem/loop | Low |
| 5 | *neuroglian* like | stem | Low | nt | nt |
| 8A | *neuroglian* like | stem/loop | High | loop | High |
| 8B | *neuroglian* like | stem/loop | High | stem/loop | High |
| 13 | *ewg* like | loop | Low | nt | nt |
| 16 | AU rich | stem | Low | nt | nt |

The site 3 structure is predicted to have a very big loop in the middle of the sequence completely exposing the putative Elav binding site that has been demonstrated to have high affinity for Elav. This structure is similar to the one reported in (Ratti et al., 2006) suggesting that the structure is conserved. The mutation of the putative binding site 3 would be predicted to disrupt the big loop and separate the putative binding site in the loop from the stem next to it. This rearrangement demonstrated to partially disrupt Elav recognition to this putative sequence (Figure 4.7).

The predicted structure of site 8 RNA shows a big central loop and a two long stem loop structures. The conserved putative binding site (A) is located at the end of the second long stem loop structure whereas the non-conserved site (B) is located between the central loop and the first stem. Mutation of the conserved putative binding site is predicted to disrupt the stem loop into a bigger terminal loop. This change in structure was demonstrated to have little effect in the affinity for Elav, probably because the putative binding site is still exposed. Intriguingly, mutation of the non conserved site (B) is predicted to have a substantial effect on secondary structure, yet this does not affect Elav binding to any great degree. This might be due to the fact that the putative Elav binding sites remain exposed in loops. Mutation of both sites prediction demonstrated to have a very big effect on the affinity to Elav. The secondary structure prediction shows that the mutations allow non-conserved site (B) to remain in a stem loop structure but it disrupts the secondary structure of the conserved site (A), which might explain the dramatic

reduction of affinity for Elav.

Predictions indicate that site 13 secondary structure also presents the putative Elav binding site in a loop. However, this site showed a low affinity for Elav. This result is intriguing because the loop structure where the site 13 lies is very similar to the one in site 3 that showed high affinity to Elav. It is probable that the vicinity of the small stem loop structure next to the putative site 13 blocks Elav access. Another possibility is that this particular site needs additional sequences for an effective binding or an Elav ligand is required. Site 16 contains two putative Elav binding sites separated only by 10 nt and in the predicted secondary structure, they are located in two adjacent stem loop structures. The sequences themselves are located within the stems, not the loops. This might explain the extremely low affinity for Elav that they showed in the biochemical experiments shown in this chapter. This secondary structure is observed to form independently of the size of the sequence used for RNA secondary structure predictions: from 150 nt as in this case or the whole *Ubx* transcript (Patraquim et al., 2011).

## 4.3 Discussion

With the experiments described in this chapter, it has been possible to identify two sequences that are specifically recognised and bound by Elav protein *in vitro*. The sites are located within intronic sequences and could potentially affect decisions on splicing site use *in vivo*. The experiments demonstrated that among the sites tested, site 3 and 8 (Figure 4.5) have a very high affinity for Elav and are able to form RNA/protein complexes of different size. Site 3 has only one putative Elav binding site that under the conditions of the EMSA and UV crosslink experiments employed here, is sufficient to mediate Elav recognition and binding with high affinity. This site shows a decrease of affinity for Elav when mutated, but does not abolish binding altogether. It is possible that other sequences around it help to stabilise the complex. There are a few A(T)4-5 sequences upstream of the conserved and tested site that could play a role, as demonstrated in the case of site 8. It would be interesting to test these other sites and mutate them to see if the interaction with Elav is completely abolished. In the EMSA experiments using 150

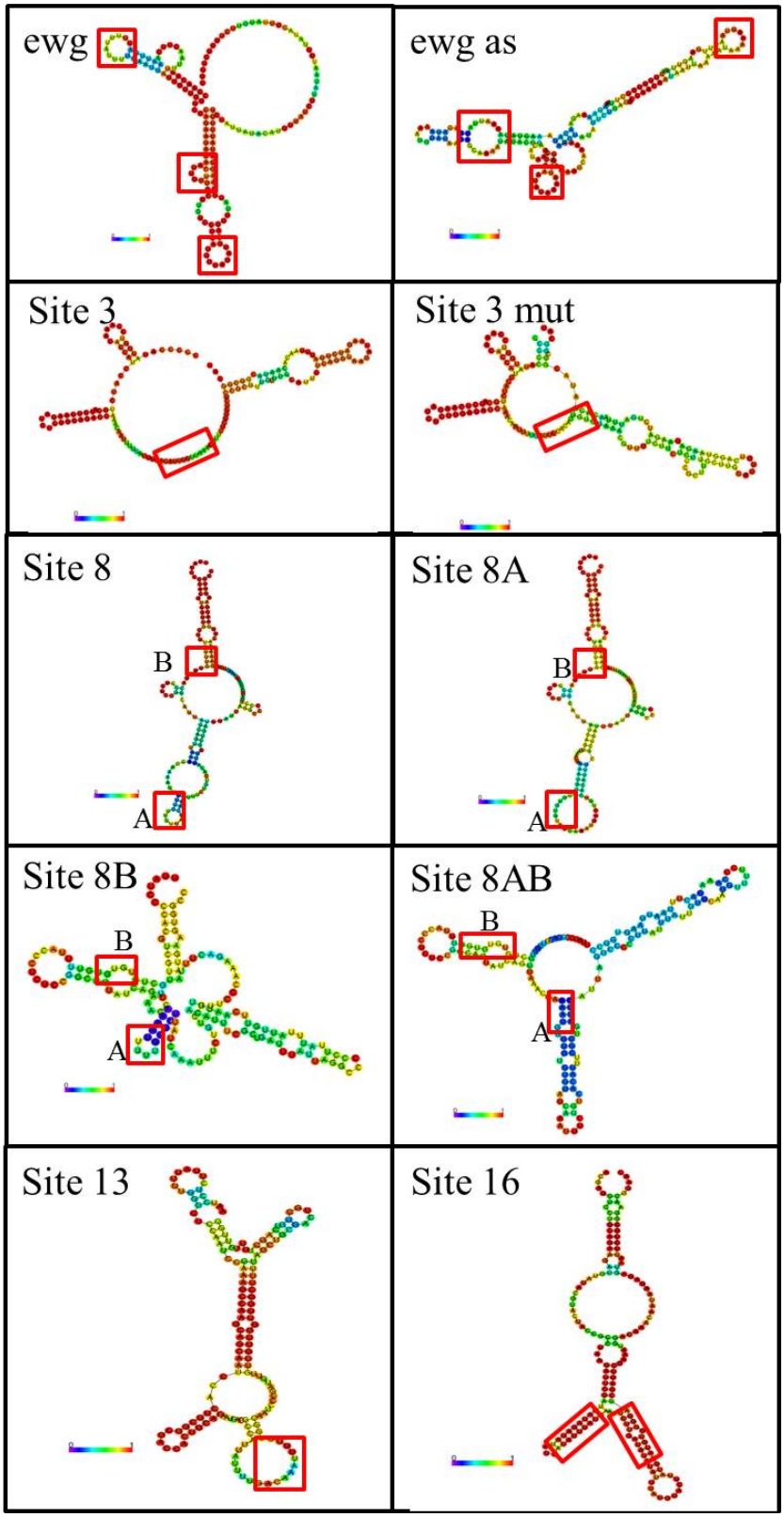


Figure 4.13: RNA secondary structure is important for Elav recognition. Predicted secondary structures generated using the RNAfold tool for the putative Elav binding sites and their respective mutations used in this work. The putative binding sites are marked with a red square.

bp probes, an intermediate RNA/Elav complex was observed at 200 and 400 nM Elav which disappear when this site is mutated. This suggests that the site might be important for Elav multimerisation. Previous studies have demonstrated that Elav multimerisation by RRM3 is mediated partially by RNA (Toba and White, 2008).

The other site found to have a strong affinity to Elav using EMSA and UV crosslink experiments is site 8. This sequence has two putative Elav binding sites, one identified by its ultra conservation among Drosophilids and another one identified by having the exact same sequence. The two sites are relatively close, only 35 nt apart. Both sites demonstrate similar affinity for Elav; when only one of the sites is mutated, the affinity is reduced around 5-10% when the sites are mutated individually. However, this affinity is greatly reduced about 75% when the two sites are mutated (Figure 4.12C).

When we look more detailed at the affinity of binding of site 8 for Elav we observed the formation of multiple intermediary complexes of different size. These are still present in the individual mutations but are less stable in the double mutant. Site 8 produces two abundant intermediates at 200 and 400 nM (arrows in Figure 4.8A), which are reduced at 600 nM when all the RNA probe is sequestered in the highest weight complex (asterisk). In the individual mutants, the lower molecular weight complex remans equally (8B) or is even more abundant (8A) compared to the wild type. In contrast, the higher molecular weight complex decreases in both individual mutants and almost disappears in the double mutant.

The other three sites tested, site 5, 13 and 16 (Figure 4.4 and 4.10) showed low affinity in both UV crosslink and EMSA experiments, and site 5 was not tested further because it showed the lowest affinity among the sites within introns. Predicted secondary structure of this site shows the putative binding site within a stem (data not shown). Site 16 showed the lowest binding ability among all the sites tested but we further tested as negative control for the rest of the experiments, with site 13. Interestingly, site 13 showed the formation of a very weak intermediate complex at 400 nM Elav in all experiments, suggesting that the site might be able to interact very weakly with Elav or it might need the presence of another factor. This hypothesis could be tested using nuclear extracts instead of purified recombinant Elav alone, allowing its association with potential RNA binding facilitators. This site

could be very interesting considering the resplicing model of *Ubx* (Burnette et al., 2005). Resplicing or recursive splicing occurs in the splicing of very long introns and whose splicing is more likely to occur alongside transcription. It uses intermediary splice sites that regenerate a 5' splice site that will be use in the next splicing event (Hatton et al., 1998), (Black, 2003), (Kandul and Noor, 2009).

The *Ubx* has very long introns and a resplicing mechanism has been described where the splicing of the first intron and joining of 5' exon and mI regenerates a 5' splice site that can be used for mI exclusion in a subsequent splicing event. With this model, the formation of all the splicing isoforms are explained (Hatton et al., 1998) (Figure 4.3). However, the third intron is particularly long and around the middle, there is one resplicing site named RP3. This forms intermediate lariats in *in vitro* experiments, suggesting that it is used for the complete removal of the third intron (Hatton et al., 1998). In this context, the site 13 could be an important site for the recognition and function of this resplicing site together with members of the splicing machinery or other RNA binding proteins. Very early evidence of the possible *in vivo* use of this site was provided when sequencing cDNAs from different sources first identified the splicing isoforms of *Ubx*. Welcome Bender's group identified five splicing isoforms of *Ubx*: Ia, Ib, IIa, IIb and IVa plus another one termed "class 4" which consisted in the 5' exon, mI, mII and a new 62 bp exon that maps at around the middle of intron III. The position of this new 3' exon maps around the RP3 site, probably resulting from a failed resplicing event (O'Connor et al., 1988).

Site 16 showed the lowest affinity in all the experiments described in this chapter. However, the ten nucleotides that separate the two potential binding sites contains a predicted mammalian PTB binding site. It is possible that PTB, or another factor, is necessary for the proper recognition of this site by Elav, and that it is actually required for polyadenylation site choice. Previous experiments demonstrated that the putative binding sequence for an RBP is preferentially recognised if it is accessible (Li et al., 2010). In the experiments described here, this is confirmed. The sites that showed the highest affinity to Elav contain the putative binding site completely or partially within a loop. However, previous studies have revealed that the HuR protein in mammals also recognises binding sites in both loop and stem (López de Silanes et al., 2004). This might seem contradictory and the results in this work

show high affinity binding sites in both circumstances: within a loop and partially paired in a stem (*ewg*, Site 3 and 8A). In contrast, site 8B is partially paired in a stem and partially accessible in a loop. However, the mutation in site 8B that maintains most of the binding ability of this sequence might be due to the fact that the binding site A remains in an open loop, highly accessible to Elav and that site 8B itself remains partially accessible.

Analysing the common features of the sequences analysed by biochemistry experiments and secondary structure predictions, it is difficult to assign a particular feature that predicts, Elav recognition with high affinity. However, it is clear that the sites with higher affinity are located within loops (*ewg*, site 3, site 8) and the disruption of the loop disrupts RNA/Elav affinity, as demonstrated in the site 3 mutant and particularly in the double mutant of site 8. However, being located in a loop does not guarantee high affinity for Elav (*ewg* antisense, site 13). It might be that the structure within the close vicinity is also important for the accessibility of the Elav protein, or other factors needed to stabilise the interaction.

The initial observation shows that Elav is necessary but it is not the only factor involved as we still see all of the *Ubx* splicing isoforms formed in the mutant. This suggests that there are more factors interacting and giving the final splicing output. Which are those factors and how they interact remain unknown. However, some lines of evidence suggest that many RBPs provide a mechanism of fine tuning of alternative splicing or polyadenylation, providing with a certain balance of isoforms produced. Maybe a small change on this balance of isoforms is enough for particular responses (Kalsotra et al., 2010).

# Chapter 5

# Other factors affecting *Ultrabithorax* mRNA processing

## 5.1 Introduction

In the previous chapters I have described the role of Elav in the processing of *Ubx* mRNA, specifically in exon inclusion and polyadenylation site selection during embryonic development. I demonstrated a direct interaction between Elav protein and intronic sequence elements *in vitro* that could mediate splice site choice in the *Ubx* mRNA. The results obtained showed that the absence of Elav protein results in a change in the proportion of *Ubx* isoforms produced but does not abolish the production of specific isoforms. A possible explanation is that the experiments were carried out using whole embryos containing RNA from epidermal and endodermal tissue, adding noise to the samples. Another possibility is that there are more post-transcriptional regulators acting in parallel to Elav controlling *Ubx* alternative splicing and polyadenylation. The objective of the present chapter is to test this second hypothesis.

The final splicing and polyadenylation patterns of many transcripts are determined by a combination of protein factors (Burnette et al., 1999), (Smith and Valcárcel, 2000). The relative levels of factors in the cell acts as a "cellular code" to establish specific splicing and polyadenylation patterns (Smith and Valcárcel, 2000). To start elucidating the "cellular code" that controls *Ubx* splicing and polyadenylation during development, I looked for candidate protein regulators of *Ubx* RNA

processing suggested from previous work. The aim was to investigate if other factors play a role in the regulation of *Ubx* mRNA splicing and polyadenylation during embryonic development. I initially focussed on factors with evidence to suggest that they affect some aspect of RNA processing.

### 5.1.1 Other factors reported to affect *Ubx* RNA processing

*Ubx* has been an excellent model to study how alternative splicing is controlled because it produces a number of splicing and polyadenylation isoforms during development; it offers a simple and sensitive system to screen for regulatory factors. The various Ubx isoforms have different abilities to rescue a *Ubx* phenotype in the haltere, so a sensitised genetic background can be used to screen for factors affecting the balance of isoforms, which will therefore impact the degree of the phenotype rescue (Burnette et al., 1999).

A deficiency screen for factors that allow the rescue of *Ubx* phenotype (haltere to wing transformation in adults) has previously identified Virilizer (Vir), Female lethal (2)d (Fl(2)d), Croocked neck (Crn) and (Heterogeneous nuclear ribonucleoprotein at 27C)(Hrb27c) genes. These four factors affected *Ubx* splicing during larval and adult stages (Burnette et al., 1999). As I have seen for *elav* mutants, *virilizer*, *fl(2)d*, *crn* and *hrb27c* mutants were found to affect the relative levels of isoforms; however, all the isoforms are still present. This suggests that a single factor is insufficient to make a splicing decision rather, this decision is made by a combination of factors. How many more factors are involved in the *Ubx* splicing decision? What is the role of each of the factors involved? Do the factors affect both microexons the same way? What about other RNA processing events? In the previous chapters, I described how a single RBP is able to affect splicing and polyadenylation events. Do *virilizer*, *fl(2)d*, *crn* and *hrb27c* mutants affect polyadenylation site choice in *Ubx* too? Do other factors affect splicing and polyadenylation? Do factors affecting transcription events also play a role in splicing and polyadenylation of the *Ubx* RNA molecule? To address these questions, I have explored the effect on splicing and polyadenylation of *Ubx* by a number *Drosophila* factors, including: Crn, Hrb27c and Fl(2)d. I also wanted to look at the relationship between transcription and post-transcriptional RNA processing, so I also looked at Brahma and Mi-2 chromatin remodelling factors.

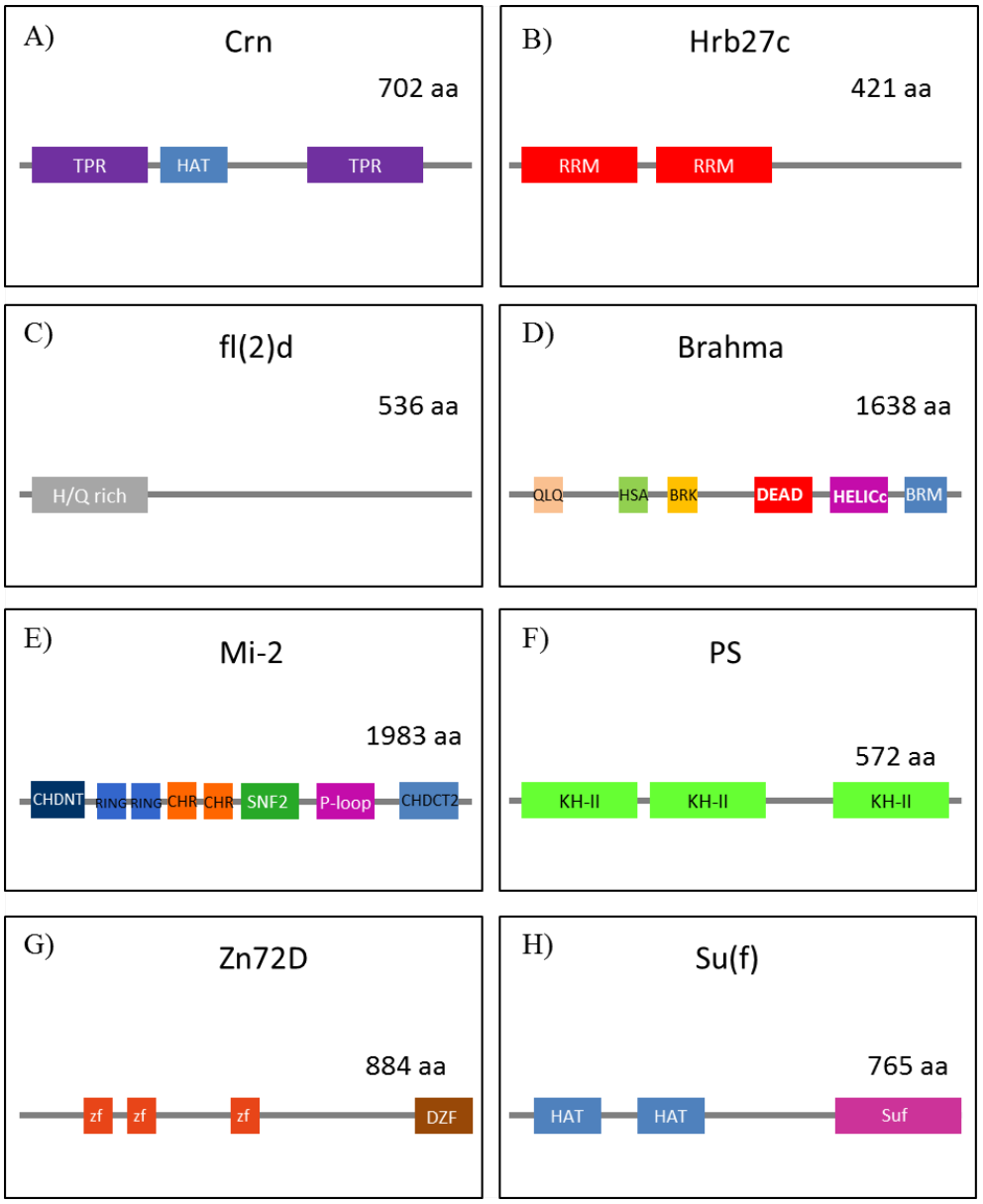


Figure 5.1: Panel of candidate *Ubx* mRNA processing regulators. A-H) Schemes showing the domain organisation of the candidate factors in boxes. A) Crn protein domain organisation: TPR, Tetratricopeptide repeat domain; HAT, Half a TPR domain. B) Hrb27c protein domain organisation: RRM, RNA recognition motif. C) fi(2)d protein domain organisation: H/Q rich domain, histidine and glutamine rich domain. D) Brahma protein domain organisation: QLQ, glutamine and leucine rich domain involved in protein/protein interaction; HSA, helicase and SANT domain related domain; BRK, associated to Brm domain; DEAD box helicase domain; HELICc is a conserved C-terminal domain of two super families of helicases; Brm, Bromodomain. E) Mi-2 protein domain organisation: CHDNT and CHDCT2 are PHD/RING finger and chromodomain associated domains; RING, Really Interesting New Gene zinc finger domain; CHR or CHROMO, Chromatin Organisation MOdifier domain; SNF2, domain from SNF2 protein. F) PS protein domain organisation: KH-II, heterogeneous nuclear ribonucleoprotein (hnRNP) K homology domain, involved in RNA recognition. G) Zn72D protein domain organisation: zf, ZfC2H2-zinc finger double stranded RNA binding domain; DZF, Nucleotidyl transferase superfamily domain. H) Su(f) protein domain orgnanisation: HAT, half a TPR; Suf, domain from the Su(f) family.

Three other RNA processing factors were also tested: Pasilla, Zn72D (Zinc-finger protein at 72D) and Su(f) (Suppressor of forked).

### Crn

*crooked-neck* gene is located in the X chromosome (CG3193). It was identified in a screen for genes important for development that produced embryos that were twisted along the longitudinal axis (Perrimon et al., 1984). Crn protein is 702 aa containing a tetratrico peptide repeat (TPR) motif, involved in cell cycle progression in yeast (Zhang et al., 1991) (Figure 5.1A). However, its function in the cell cycle has not been described in *Drosophila*. More recent studies have revealed that Crn is expressed ubiquitously in *Drosophila* embryos, in cell nuclei and enriched in "speckles". These foci have been described to serve as storage for splicing factors,(Spector and Lamond, 2011) suggesting that Crn could have a direct role in splicing. Crn fractionates with snRNPs in sucrose gradients, supporting the existence of a splicing multiprotein complex containing Crn (Raisin-Tani and Léopold, 2002), moreover, Crn protein promotes the assembly of the spliceosome (Wang et al., 2003).

Its human homolog (Hcrn) has been demonstrated to associate with splicing complexes and *in vitro* splicing is compromised if Hcrn is immunodepleted from the nuclear extract (Chung et al., 2002). Crn in *Drosophila* is required for germ cell viability as well (Perrimon et al., 1984). Interestingly, despite the ubiquitous distribution of Crn, *Drosophila* deletion mutants have strong phenotypes in the CNS and gut (Eberl and Hilliker, 1988). In agreement with this, Crn is present in the cytoplasm and affects the development of glial cells. Crn interacts with HOW (held out wings) to mediate migration and ensheathment of peripheral axons (Salecker, 2006). Crn is not predicted to bind RNA directly, but HOW does. Together Crn and HOW regulate the alternative splicing of *nervana2 (nrv2)* (Na/K ATPase beta subunit) and *neurexinIV (nrxIV)*, which is transmembrane cell-surface protein. The resulting proteins are components of the glial septate junctions (Salecker, 2006), (Edenfeld et al., 2006). This suggests an important role of Crn in the CNS, modulating the splicing of a variety of targets, maybe including *Ubx*. If Crn affects *Ubx* mRNA processing within the CNS, the determination of cell lineages might be compromised since *Ubx* is important in the determination of some neuronal lineages, such as NB-1

(Rogulja-ortmann and Technau, 2008). *crn* mutants have been shown to affect *Ubx* by modulating the inclusion of mI exon, generating less isoform I and more isoform II (Burnette et al., 1999).

### Hrb27c

Heterogeneous nuclear ribonucleoprotein at 27C (CG10377) is located in the left arm of chromosome 2. It is also known as Hrp48, p50, RRM7, Hrb27-c. It was first described in the tissue-specific regulation of P- element transposase third-IVS3 intron splicing. It binds a pseudo-5' splice site within the exon regulatory element to reduce the accessibility of real 5' SS to splicing machinery (Siebel et al., 1994).

The *Hrb27c* protein contains two N- terminal RBD-type (RNA Binding Domain) RNA-binding domains and a glycine-rich carboxyl-terminal domain (Matunis et al., 1992) (Figure 5.1B). It is one of the most abundant hnRNP proteins in *Drosophila* (Blanchette et al., 2005), (Blanchette et al., 2009). These proteins have been shown to associate with the majority of nascent transcripts (Hammond et al., 1997). More recently, it has been shown that *Hrb27c* promotes grk localization and affects embryonic DV patterning, through direct interaction (RNA-independent) with otu protein (Ovarian tumor) to aid RNA localization. *Hrb27c* interacts with Squid in an RNA dependent manner resulting in the proper localization of grk mRNA (Goodrich et al., 2004). Some evidence suggests that *Hrb27c* plays a role in translational regulation of osk RNA, in a similar way to grk regulation (Goodrich et al., 2004). *Hrb27c* appears to have a role in chromosome dispersion, as polytene chromosomes in nurse cell fail to disperse in a *Hrb27c* mutant (Goodrich et al., 2004).

Acting on alternative splicing, it has been previously shown that *Hrb27c* affects the inclusion of both *Ubx* microexons in *Drosophila* larvae and adults. In *Hrb27c* mutants the level of *Ubx* isoforms II and IV are increased whilst I is reduced (Burnette et al., 1999).

### *fl(2)d*

Female lethal d (CG6315) is located in the right arm of chromosome 2. It was first described as a factor that interacts with Sxl. It is involved in sex determination and X chromosome dosage compensation (Penalva et al., 2000). Fl(2)d is a 539 aa

protein, with no structural domain characteristic of proteins involved in splicing in its primary sequence. The amino- terminal region is a HQ-rich domain (Figure 5.1C). It also has a glutamine-rich region and a putative dimerization domain (Penalva et al., 2000), (Ortega et al., 2003). Fl(2)d is also needed for a vital process which is independent of Sxl (Granadino et al., 1990). The homozygote mutant is lethal in females and males, but the precise role of Fl(2)d is unknown. It is unlikely that it is a member of the core splicing machinery because the mutant is not cell lethal, when it is embryonic lethal. It could have a role in directing the correct splicing of a factor needed for development, independently of the sex determination pathway (Penalva et al., 2000). Interestingly, in a screen looking for interactors with *Ubx*, significant reductions of mI and mII inclusion were observed in larvae and adults with mutations in this factor (Burnette et al., 1999).

### 5.1.2 Chromatin structure in mRNA processing

Two of the proteins that can be associated to the Swi/SNF complex have been chosen to test the role of chromatin regulation in splicing and polyadenylation regulation: Brahma and Mi-2. Brahma protein has been shown to affect alternative splicing by modulating the transcription rate of RNApolII. Mi-2 protein has been shown to affect *Ubx* expression by interacting with Polycomb. I sought to investigate if *Ubx* splicing is affected in the absence of these two factors.

#### Brahma

Located in the left arm of the 3rd chromosome, (CG5942) Brahma (Brn) was first described in a screen looking for factors that suppress or enhance mutations of homeotic alleles of the Antenapaedia and Polycomb groups (Tamkun et al., 1992). Mutations in the *brm* locus suppressed the production of extra sex combs (Polycomb phenotype) and antenna to leg transformation (Antenapaedia phenotype). It is a 1638 amino acid protein, with a structure related to the SNF2/SWI2 (mating-type switch/sucrose nonfermenting) yeast transcription activator. It contains a conserved bromodomain at its C-terminus (Tamkun et al., 1992) (Figure 5.1D). It belongs to the trithorax group (Kennison and Tamkun, 1988) and is part of a chromatin remodelling complex, interacts with U1 and U5 snRNPs and with the RNA polymerase II,

slowing its rate of transcription affecting exon inclusion in general. The presence of Brahma in the transcriptional complex leads to a slower transcription, allowing the coexistence and competition of the splicing sites, therefore, promoting exon inclusion (Batsché et al., 2006). It also interacts with splicing factors to regulate alternative splicing. In larval stages and S2 cells, depletion of Brahma affects the splicing of lola and Gpdh transcripts (Batsché et al., 2006), (Waldholm et al., 2011). h*Brm* regulates the alternative splicing of the CD44 pre-mRNA in human cells by decreasing the elongation rate of RNAP II (Waldholm et al., 2011). *Drosophila Brm* is a nuclear protein, expressed ubiquitously, in early embryogenesis and is maternally encoded. Further to its interaction with Polycomb and Antenapaedia, it has been shown that it interacts with *Ubx* in combination with trithorax. A *brm* mutant will increase a *Ubx* mutant phenotype, suggesting a role for Brahma activating *Ubx* gene (Tamkun et al., 1992). However, no effect on Ubx RNA processing has been described.

A more detailed study of Brahma function, has found that this factor associates with open chromatin, which is actively transcribed and moreover, that Brm is essential for global transcription (Armstrong et al., 2002).

The interaction with chromatin depends on the bromodomain and is direct. However, direct interaction of Brm with the RNAPol-II has not been reported. hBrm ATPase activities are stimulated to the same extent by nucleosomes and naked DNA.

### *Mi-2*

Gene located in the left arm of the 3rd chromosome (CG8103), *Mi-2* was originally described in mammals as a dermatomyositis specific auto antigen (Seelig et al., 1995). The *Mi-2* is a 1983 aa protein that contains two PHD-zinc-finger domains, two chromo-domains and a a SWI2/SNF2-type helicase/ATPase domain (Figure 5.1E) (Seelig et al., 1995), (Khattak et al., 2002). It has been described as a ATP-dependent chromatin remodelling factor and participates in the repression of Hunchback and Polycomb (Kehle, 1998), (Fasulo et al., 2012). *Mi-2* interacts with Polycomb to repress the expression of *Ubx*. Mutations in *Mi-2* alone have no evident *Ubx* phenotype, and the expression of *Ubx* and Abdominal B is normal in

*Mi-2* mutant embryos, probably due to the maternal contribution of *Mi-2*. *Mi-2* is expressed strongly in early embryogenesis, less strongly in late embryos and weakly in larvae; it is also expressed strongly in ovaries (Kehle, 1998).

*Mi-2* is a component of a chromatin remodelling complex whose ATPase is preferentially stimulated by nucleosomes and modifies chromatin through histone deacetylation to repress transcription and (Brehm et al., 2000). The sites of deacetylation are targeted but the mechanism remains unknown.

### *Pasilla*

Located in the right arm of the 3rd chromosome (CG42670), *Pasilla* encodes the *Drosophila* orthologue of the mammalian Nova (Neuro-oncological ventral antigen) protein, which was first described as a target protein of the human autoimmune disorder POMA (paraneo- plastic opsoclonusmyoclonus ataxia) (Seshaiah et al., 2001). It was one of the first neuronal specific splicing factors studied and it is involved in neuronal survival, migration and splicing of transcripts with dendritic functions (Norris and Calarco, 2012). *Pasilla* protein contains three conserved KH RNA binding domains (Figure 5.1F). The human homologue, Nova, promotes the inclusion of the exon e24a and represses the inclusion of exon e31a in the P-type Calcium channel mRNA in the brain, demonstrating that the same factor has both enhancing and repressing capabilities depending on the position of the binding site (Allen and Lipscombe, 2010). The binding site of this factor has been well characterised (YCAY) (Licatalosi et al., 2008). Studies using high - throughput sequencing of crosslinked immuno-precipitation with Nova antibody have revealed a large number of targets, confirming that the position of binding sites determines the effect on splicing, generating a sort of RNA "map". Nova binding elements located in regions upstream of exons act as repressor sites whereas elements located downstream the exons act as activators of splicing.

The data is so consistent that it provides the grounds to predict outcome of Nova action de novo. Interestingly, the study found a large number of interactions located at the 3' UTR of transcripts, revealing a role of Nova in alternative polyadenylation in the mouse brain (Licatalosi et al., 2008). Nova could provide a direct link between alternative splicing and polyadenylation in the brain. In *Drosophila*, *Pasilla* is highly

expressed in larval salivary glands and has been implicated in salivary gland secretion (Seshaiah et al., 2001). Interestingly, it has been demonstrated that the RNA map described in mammals is very well conserved to *Pasilla* in *Drosophila* (Brooks et al., 2011), suggesting that the mechanism used in both systems is similar. *Pasilla* is not maternally encoded but embryonic expression starts at stage 6 at the endoderm and at later stages it is expressed in the hindgut and midgut, salivary glands and fat body (Tomancak et al., 2007).

### *Zn72D*

Zinc-finger protein at 72D (CG5215), is located in the left arm of chromosome 3. It has three C2H2 zinc fingers, a DZF domain (Figure 5.1G) *Zn72D* is a protein involved in the sex determination pathway in *Drosophila*, needed for the splicing of the *maleless (mle)* transcript, which encodes an essential subunit of the MSL (Male Specific Lethal) complex in coordination with Belle DEAD box helicase(Worringer and Panning, 2007), (Worringer et al., 2009). However, *Zn72D* is also required for the viability of both sexes. Consistent with this, *Zn72D* co-localises with elongating RNA polymerase, potentially controlling splicing indirectly by modulating transcription speed (Worringer and Panning, 2007). Additional to this role, it has been suggested that recruitment of *Zn72D* to the *mle* transcript has an effect on translation levels, (Worringer et al., 2009). *Zn72D* protein is localised to the nucleus and cytoplasm in humans. ZFR (homolog of *Zn72D*) is localised in neuronal granules where it interacts with and localises *staufen262* to the cytoplasm, presumably affecting its translation rates (Worringer et al., 2009). *Zn72D* is maternally encoded and is expressed ubiquitously at early embryonic stages. At later stages its expression is restricted to the CNS, PNS, hindgut and midgut (Tomancak et al., 2007).

### *Su(f)*

Suppressor of forked, (CG17170) is located in the X chromosome. It was first described as a protein that suppresses the bristle phenotype produced by some of the alleles at the forked locus (Dudickz et al., 1974). It is the *Drosophila* homologue of *CstF-77*, a subunit of human cleavage stimulation factor (CstF). It is also con-

served to yeast, suggesting that its role in 3' UTR processing is conserved (Takagaki, Yoshio ; Manley, 1994). The *Su(f)* protein has a proline-rich domain a HAT, hald TPR domain (Figure 5.1H). *Su(f)* is required for the cleavage of pre-mRNA during mRNA 3' end formation. It also interacts *in vitro* with *Drosophila CstF-64*, *CstF-50*, and itself. *Su(f)* activity must be optimal for regulated utilisation of poly(A) sites (Benoit et al., 2002).

Because *Su(f)* is a factor of the general cleavage and polyadenylation machinery, it is expressed ubiquitously during embryonic stages, and is maternally encoded (Tomancak et al., 2007). *Su(f)* is required for larval and pupal stages; in the adults is needed for oogenesis and clones in imaginal discs lacking *Su(f)* show apoptosis and disruption of pattern determination (Mitchelson et al., 1993). Early evidence suggests that *Su(f)* also plays a role in RNA stability, as its mutant leads to accumulation of *f (forked)* RNAs in a gypsy insertion in f locus, and also of an heterologous RNA (Mitchelson et al., 1993).

## 5.2 Results

### 5.2.1 Normal expression of candidate factors during *Drosophila* embryonic development

To test the effect of mutating the candidate factors in *Ubx* mRNA processing I first investigated their normal expression levels at embryonic stages 15 and 17. For this, I did a search in RNA-seq data from modENCODE available in FlyBase (http://flybase.org/). From this collection I recovered the RPKM (Reads Per Kilobase of transcript per Million mapped reads) count for each gene at stages 14-16 and 16-18 hAEL, which given in the table below:

| Candidate factor | Stage 14-16 hAEL | Stage 16-18 hAEL |
|---|---|---|
| *brm* | 53 | 36 |
| *mi-2* | 101 | 71 |
| *ps* | 26 | 78 |
| *crn* | 23 | 19 |
| *zn72D* | 23 | 16 |
| *fl(2)d* | 41 | 36 |
| *hrb27c* | 62 | 59 |
| *su(f)* | 18 | 14 |

The RNAseq data shows that all of the factors I planned to test are expressed at the stages I am interested in this study. In general, earlier stages show higher levels of expression that later stages, except for *ps*, which shows much higher levels at stage 16-18 hAEL.

*brahma*, *su(f)*, *zn72D*, *hrb27c*, *fl(2)d* and *mi-2* are maternally provided and show higher expression levels at earlier stages, being *mi-2* the factor with highest expression. *crn* and *ps* are not maternally encoded but *crn* expression is slightly higher at stage 14-16 hAEL.

### 5.2.2 Checking of mutants for the candidate factors

Next I confirmed the mutants for the selected factors. The collection of mutant stocks for the candidate factors is a mixture of mutation types. Four of them (*zn72D*, *ps*, *fl(2)d* and *hrb27c*) are P element insertions (Figure 5.2 and 5.3) , *crn* and *brm* are point mutations whose exact positions are unknown (Figure 5.4), *su(f)* is a deletion and *mi-2* is an insertion (Figure 5.5). The P-element insertion mutants were genotypified by PCR using primers that map outside the insertion and inside the insertion (Figure 5.2 and 5.3). Genomic DNA was extracted from whole adults (two males and two females) and wild type flies were used as control. The *ribosomal protein L32* (also known as *rp49*) was used as a control for RNA input for PCR.

The four mutations by insertion produced an amplicon when a forward primer before the insertion and a reverse primer inside the inserted element were used in genomic PCR, and no band was observed in the wild type control (Figure 5.2 and 5.3).

The same strategy was used for the *su(f)* mutant, using a forward primer before the deletion and a reverse primer inside the deletion. The heterozygote showed a reduction of near 50% relative to *rp49* signal (Figure 5.4A). *mi-2*, *crn* and *brahma* mutations were not possible to characterise by genomic PCR (Figure 5.4B, 5.5).

***zn72D* mutant:** The mutant *zn72D*[BG02677] is a P element (PGT1) insertion at the beginning of the transcript, interrupting the ORF. The BDSC stock 21586 (y[1]w[67c23]; Pw[+mC]=GSV7GS21586/SM1) was balanced with a Cyo chromosome that expresses GFP under a Twist promoter (Figure 5.2A).

***pasilla* mutant:** The *pasilla* mutant ps [10615] is a P element (PPZ) insertion at approximate the center of the gene (Figure 5.2B). The BDSC 10236 (Pry[+t7.2]=PZps[10615] ry[506]/TM3, ry[RK] Sb[1] Ser[1]) stock was balanced with TM3 with a Twist-GFP (Figure 5.2B)

***hrb27c* mutant:** *hrb27c*[DG19705] is a P element (PwHy) insertion at the start of transcription, interrupting the ORF (Figure 5.2C). The BDSC stock 21702 (y[1] w[67c23]; Py[+t7.7] w[+mC]=wHy*hrb27c*[DG19705]/SM6a) was balanced with a CyO chromosome with GFP expression under a Twist promoter (Figure 5.3A).

***fl(2)d* mutant:** The mutant *fl(2)d*[DG37212] is a P element (PwHy) insertion located at the beginning of the transcript, interrupting the ORF (Figure 5.1D). The BDSC stock 21586 (y[1] w[67c23]; Py[+t7.7] w[+mC]=wHy*fl(2)d*[DG37212]/SM6a) was balanced with a Cyo chromosome expressing GFP under Twist promoter (Figure 5.3B).

***su(f)* mutant:** The mutant *su(f)*[6] is a deletion of 1kb (Figure 5.3A). The BDSC stock 4735 (y[1] v[1] *su(f)*[6]/FM7a) was balanced with a FM7c chromosome that expresses GFP under Kruppel promoter (Figure 5.4A).

***mi-2* mutant:** The *mi-2*[4] mutant is a 4 bp insertion caused by ethyl methanesulfonate (Seelig et al., 1995)that creates a frameshift producing an early stop codon (Figure 5.3B). The BDSC stock 26170 (*mi-2*[4] red[1] e[4]/TM6B, Sb[1] Tb[1] ca[1])

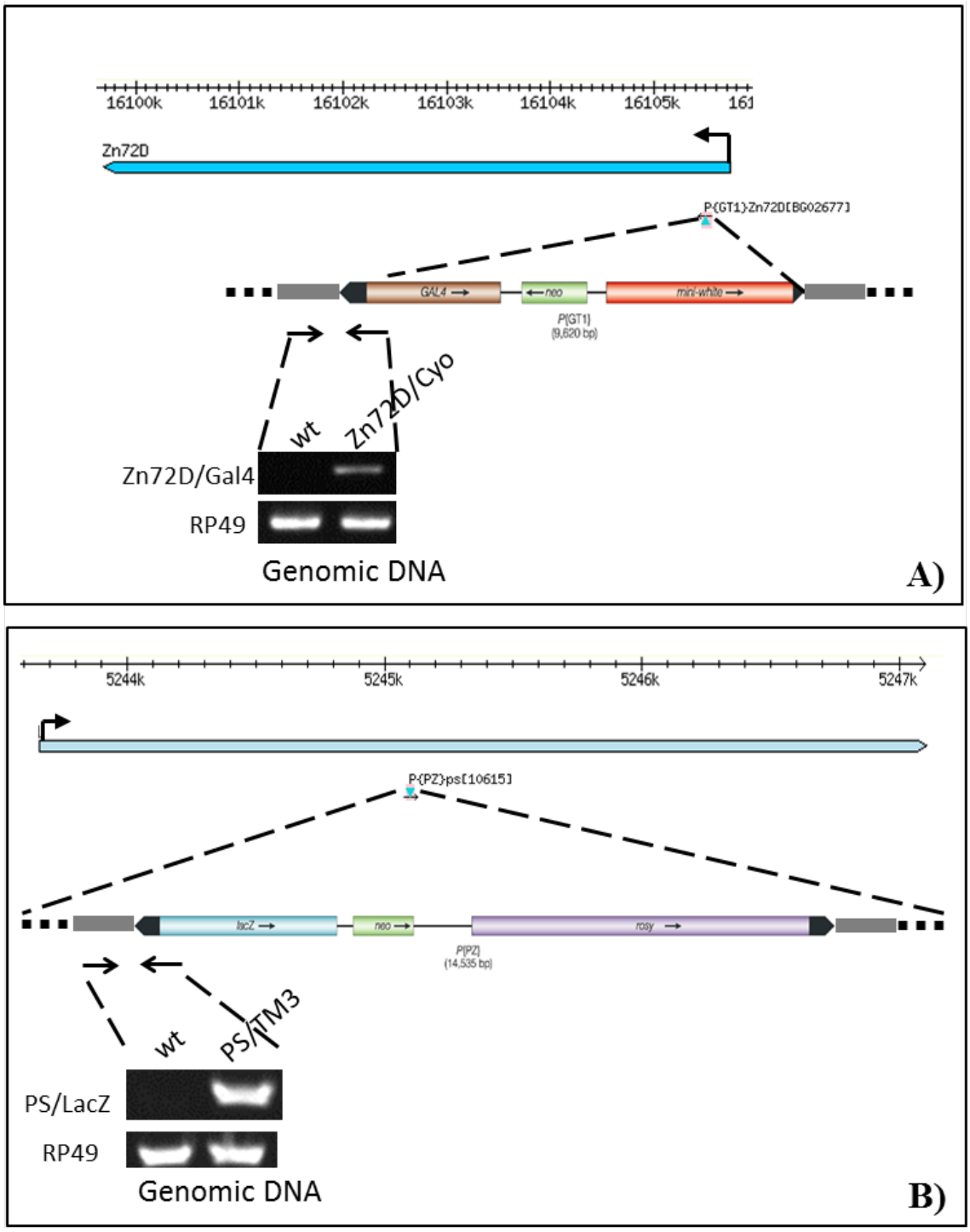


Figure 5.2: Geno-typification of P-element insertions in *zn72D* and *ps* mutants. A-B) The top scheme represents the position of the insertion of the P-element in each case. Just below is a scheme of the element inserted. Genomic DNA was extracted for PCR using the primers in the positions marked with arrows, forward primer before inserted element and reverse primer within the inserted element. Agarose gels were used to resolve the PCR, wild type genomic DNA was used as negative control and *rp49* as loading control. A) *zn72D* mutant. B) *ps* mutant. Gene representation taken from FlyBase and P-element scheme taken from (Adams and Sekelsky, 2002).

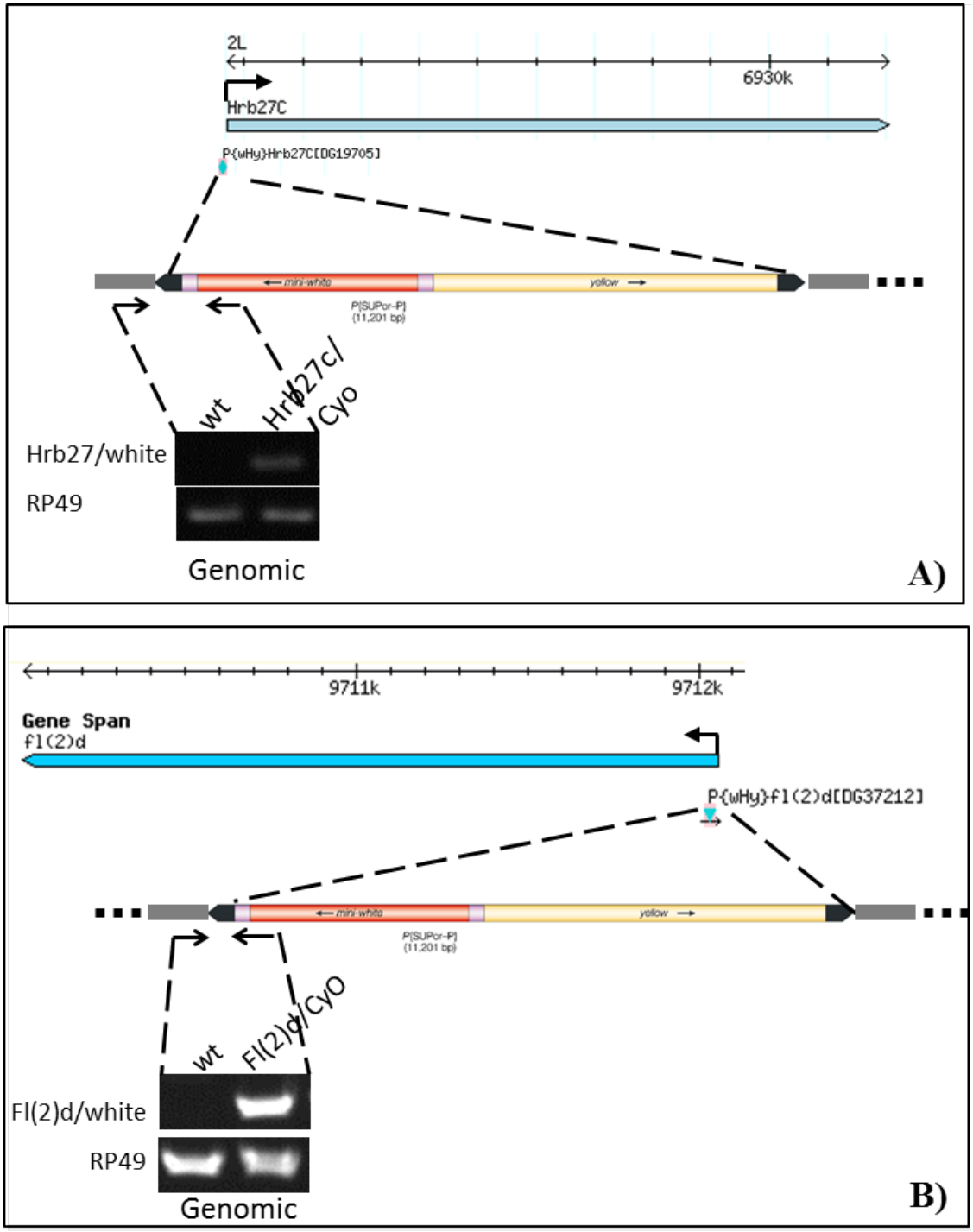


Figure 5.3: Geno-typification of P-element insertions in *hrb27c* and *fl(2)d* mutants. A-B) The top scheme represents the position of the insertion of the P-element in each case. Just below is a scheme of the element inserted. Genomic DNA was extracted for PCR using the primers in the positions marked with arrows, forward primer before inserted element and reverse primer within the inserted element. Agarose gels were used to resolve the PCR, wild type genomic DNA was used as negative control and *rp49* as loading control. A) *hrb27c* mutant. B) *fl(2)d* mutant. Gene representation taken from FlyBase and P-element scheme taken from (Adams and Sekelsky, 2002).

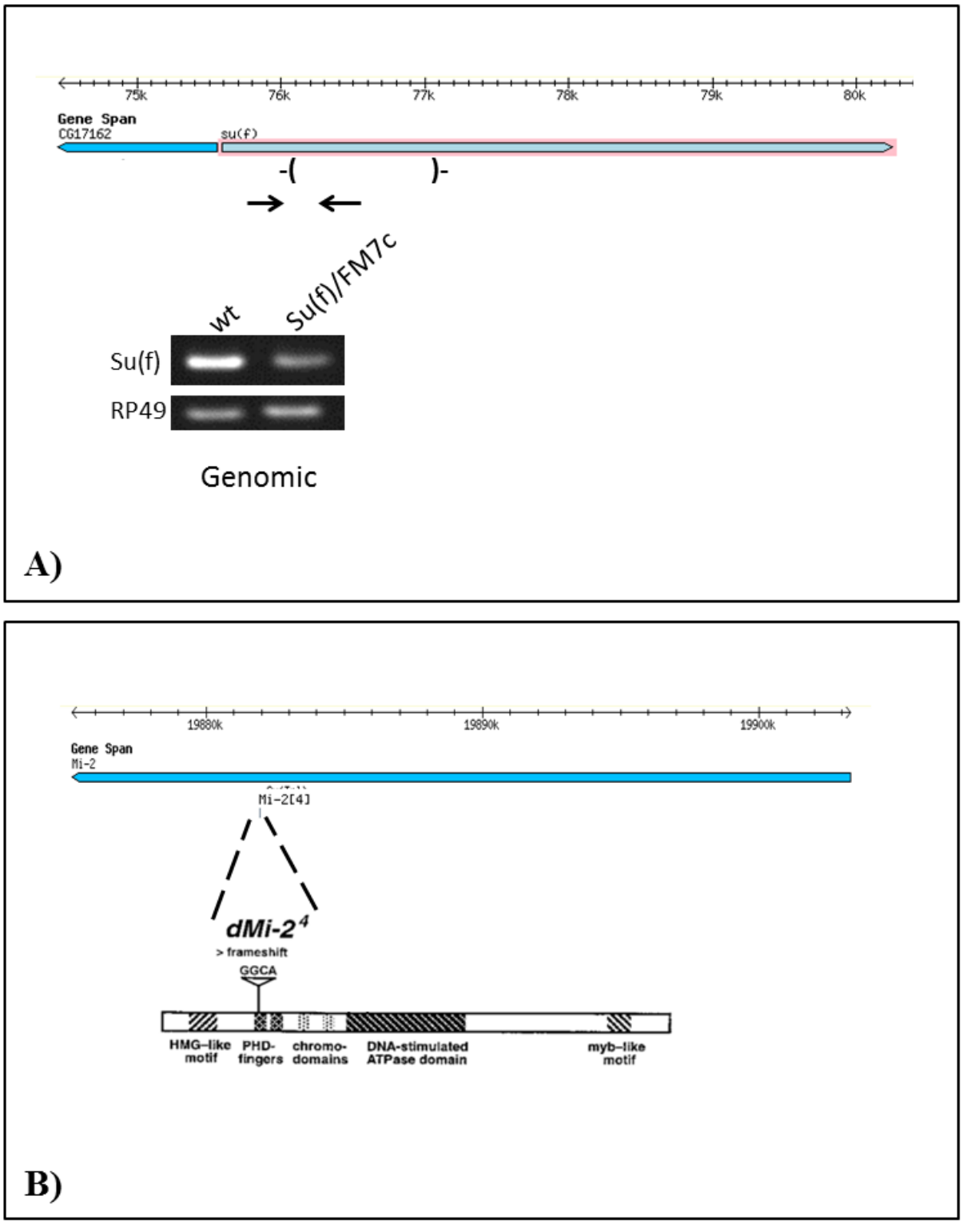


Figure 5.4: *su(f)* and *mi-2* mutant. A-B) The top scheme represents the gene and the position of the mutation. A) The *su(f)* deletion was the only one possible to detect by genomic PCR using primers outside and inside the deletion, marked with arrows. Agarose gel showing the amplicons produced on wild type and *su(f)*/FM7c adults. Wild type flies were used as control and *rp49* as loading control. B) *mi-2* mutant is an insertion of 4bp at the start of the PHD-fingers domain causing a frameshift. Scheme from (Seelig et al., 1995).

was balanced with a TM3 balancer with GFP expression under Twist driver (Figure 5.4B).

***crn* mutant:** The mutant used here is a point mutation caused by ethyl methanesulfonate (Eberl and Hilliker, 1988) (Figure 5.3C). The stock yok-EH352 from BDSC 4292 (y[1] *crn*[yok-EH352]/C(1)DX, y[1] f[1]/Dp(1;Y)y[2]sc) was not possible to balance with a GFP expressing chromosome.

***brahma* mutant** The *brahma* mutant *brm*[2] has been described as a point mutation at an unknown position that changes the activity of the protein but does not affect its expression (Tamkun et al., 1992) (Figure 5.3D). Mutant embryos die before hatching and loss of maternal protein affects oogenesis (Brizuela et al., 1994). The BDSC stock 3619 (*brm2 e5 Ca1*—/ TM6b Tb Sb DfdLacZ) was balanced with a TM3 balancer that expresses GFP under a Twist promoter.

### 5.2.3 Mapping the Brahma mutation

The brahma mutant was described using a mutant produced by ethyl methanesulfonate (brm[2]) reported by Kennison in 1988 (Kennison and Tamkun, 1988) displays normal protein expression (Tamkun et al., 1992), however, the study claims it is not functional suggesting a point mutation in a functional domain such as the bromodomain. The protein structure of the brm protein contains the bromodomain in the carboxyl end (blue in Figure 5.5), a conserved helicase domain of the HELICc family (purple in Figure 5.5) and a DEAD-like domain (red in Figure 5.5), also an helicase. This three domains are the most conserved and better characterised domains within the brahma protein so I designed primers to detect any mutation in those areas. The sequence of the primers are enlisted in the following table:

| Oligo name | sequence | domain |
|---|---|---|
| brm.1.fw | ATCAATTGCCCATGCCTTC | Bromodomain |
| brm.1.rv | TGAGTTACACTGCTCGCTGTT | Bromodomain |
| brm.2.fw | AGCGAGCAGTGTAACTCACCT | HELICc |
| brm.2.rv | AGGACCGGATCTGTATCGTG | HELICc |
| brm.3.fw | AAGATAAACTTACCGGATACAACTCC | DEAD-like helicase |
| brm.3.rv | GTGAAACAGCACAAGGACGA | DEAD-like helicase |

High Fidelity PCR was carried out and the PCR products were cloned in pGEMT easy vector and sequenced. Four independent clones for each amplicon were sent to sequence using both T7 and M13RV oligos.

The sequencing of the cloned fragments showed only a single point mutation with the brm.1 pair of oligos corresponding near to the carboxyl end of the bromodomain, in the BC loop: a T was substituted by a C changing the codon TCT to TCC. Both codons code for Serine (Figure 5.5). Structure of bromodomain was taken from (Mujtaba et al., 2007).

The silent point mutation found does not explain the lack of function of the brahma mutant. The rest of the domains were not analysed so the mutation might be in another region fundamental for brahma normal function.

**The mutations of the candidate factors show reduction in the expression of their proteins**

.

To confirm that the stocks for the candidate factors used in this study, RT-PCRs for each factor were carried out using cDNA from mutant embryos at stage 17 (Figure 5.6 and 5.7).

Most of the factors tested showed a marked reduction of their expression in their corresponding mutant embryos (Figure 5.6 and 5.7). *brahma* did not show significant reduction but this was not expected because the mutation still produce protein, although the reports of the mutant claims is not functional (Tamkun et al., 1992)(Figure 5.7C).

*ps* and *su(f)* showed a complete reduction of their expression in the mutants (Figure 5.6B and D) whereas the *crn* heterozygote shows a significant reduction in

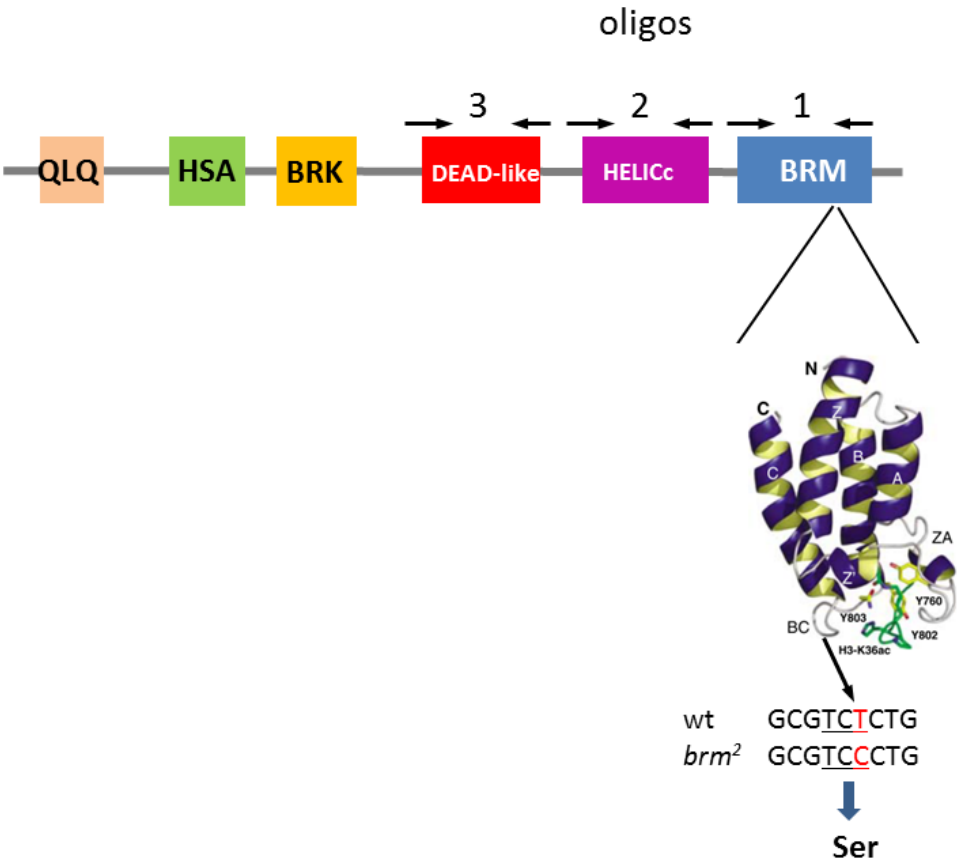


Figure 5.5: Mapping of the $brm^2$mutant. The protein domains are shown as coloured boxes: light pink for QLQ domain, green for HSA, yellow for BRK, red for DEAD-like domain, purple for HELICc domain and blue for the bromodomain. The arrows represent the approximate position of the primers designed for sequencing. The BRM domain structure is shown below and the position of the silent point mutation is indicated with an arrow (Mujtaba et al., 2007).

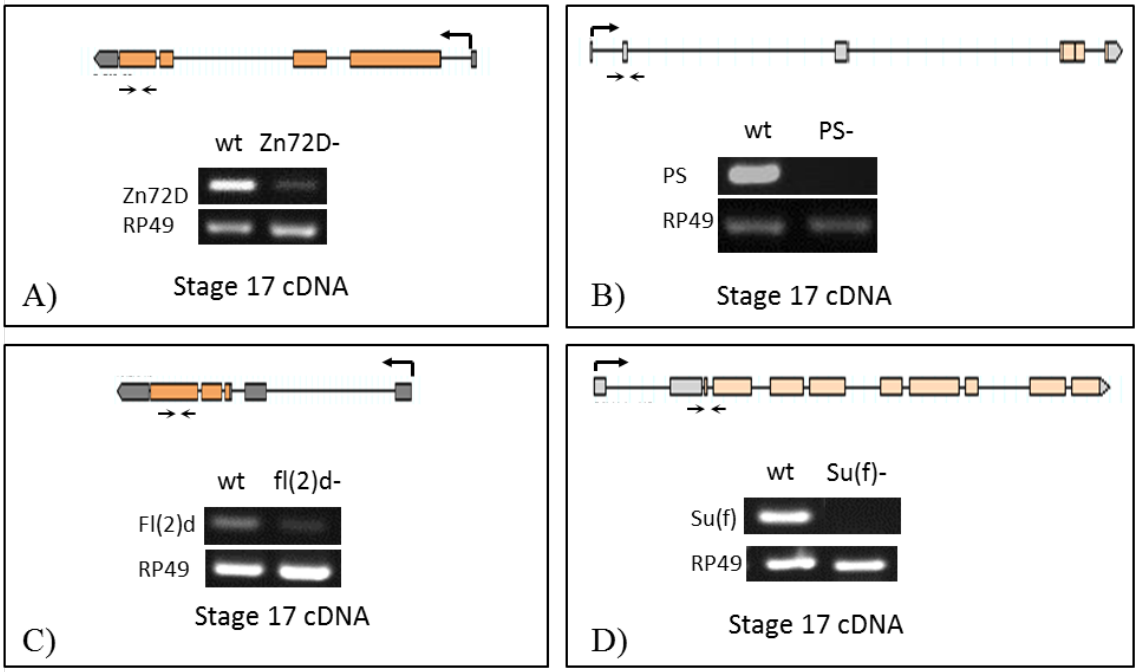


Figure 5.6: The mutants of the candidate factors reduce their expression. RT-PCR using RNA extracted from homocygote mutant embryos at stage 17. Top scheme is a representation of the transcript for each candidate factor and the primers used for RT-PCR are marked with arrows. Agarose gels were used to resolve the bands, wild type RNA was used as positive control and *rp49* was used as loading control. A) *zn72D*. B) *ps*. C) *fl(2)d*. D) *su(f)*.

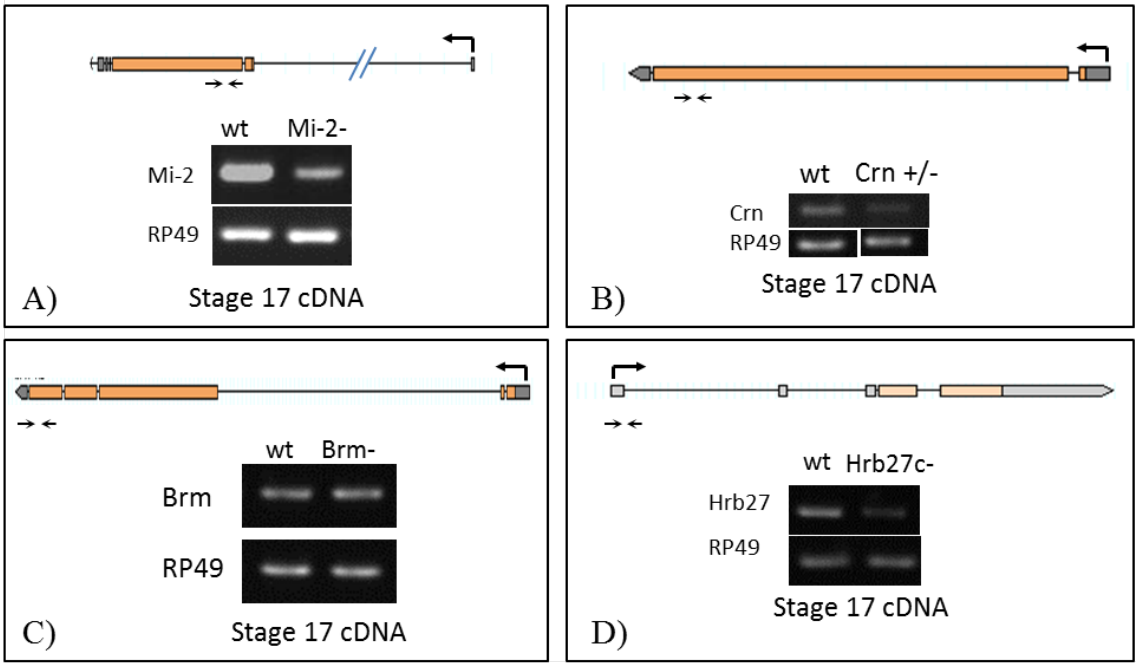


Figure 5.7: The mutants of the candidate factors reduce their expression. RT-PCR using RNA extracted from homocygote mutant embryos at stage 17. Top scheme is a representation of the transcript for each candidate factor and the primers used for RT-PCR are marked with arrows. Agarose gels were used to resolve the bands, wild type RNA was used as positive control and *rp49* was used as loading control. A) *mi-2*. B) *crn*. C) *brm*. D) *hrb27c*.

its expression (Figure 5.7B).

*zn72D*, *fl(2)d* and *hrb27c* show a substantial reduction of their expression but not complete, due to their maternal contribution (Figure 5.6A, C and 5.7D).

### 5.2.4 Effects on *Ubx* mRNA levels

RNA binding proteins are factors that usually affect more than one aspect of RNA metabolism, for example, transcription or degradation (Gamberi et al., 2006). To assess if the protein factors have a role in transcription, the abundance of the *Ubx* transcript was analysed by semiquantitative RT- PCR using primers that amplify a region in the proximal 3' UTR (Universal) (Thomsen et al., 2010) (Figure 5.8A). For each experiment, three independent biological samples of at least 50 embryos were used for each genotype tested, RNA was extracted and RT-PCR was carried out as described in the Methods section.

The preliminary results presented in this chapter suggest that the factors tested affect splicing, polyadenylation and RNA abundance of Ubx.

An interesting observation is that the effects of mutations were different at the two developmental time points sampled. At stage 15, in general there are not major differences between the wild type reference and the mutants for the different factors, except for *ps* mutant which showed an increase in Ubx RNA and *hrb27c*, which show a decrease in total *Ubx* levels. *crn* shows a statistically non significant increase of *Ubx* levels. The effect of *ps* and *crn* might be related to their role in splicing or may be an indirect effect via the processing of a factor involved in RNA stability.

The other factors show either no difference or a slight increase on mRNA levels (Figure 5.8B and C). However, at stage 17, the effect of the factor mutants on *Ubx* mRNA levels change dramatically and all of them show a marked decrease, except for *fl(2)d* mutant which shows no change (Figure 5.9). *brm* and *crn* show a statistically significant reduction of *Ubx* mRNA levels.

This result suggests that most of the protein factors tested here are not involved in transcription regulation or RNA stability at stage 15. *fl(2)* has an impact in earlier stages but not at later stages. This could be due to a lower expression of the factor at stage 17. Given that *fl(2)* is maternally encoded, it is more strongly expressed at earlier stages, making it possible that normally at stage 17 its role is

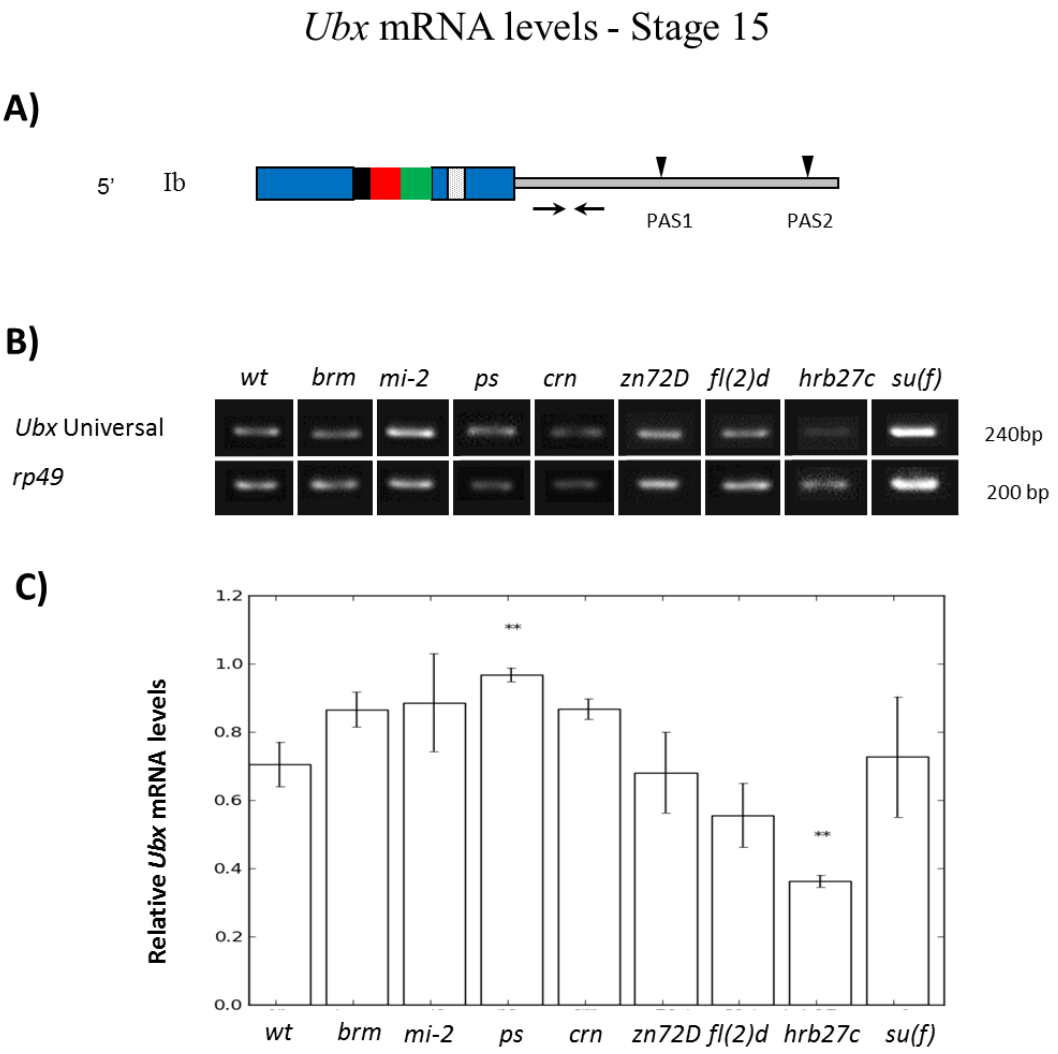


Figure 5.8: *Ubx* mRNA levels at stage 15: RT-PCR of three independent biological RNA samples. A) Diagram of *Ubx* mRNA showing the position of the primers used for RT-PCR. B) RT-PCR of *Ubx* Universal region and *rp49* as loading control in each of the mutants cDNAs and wild type as control. C) Quantification of the signal normalised with *rp49* levels showing a reduction of *Ubx* mRNA in the *hrb27c* mutant and an increase in *ps* mutant. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to the wild type.

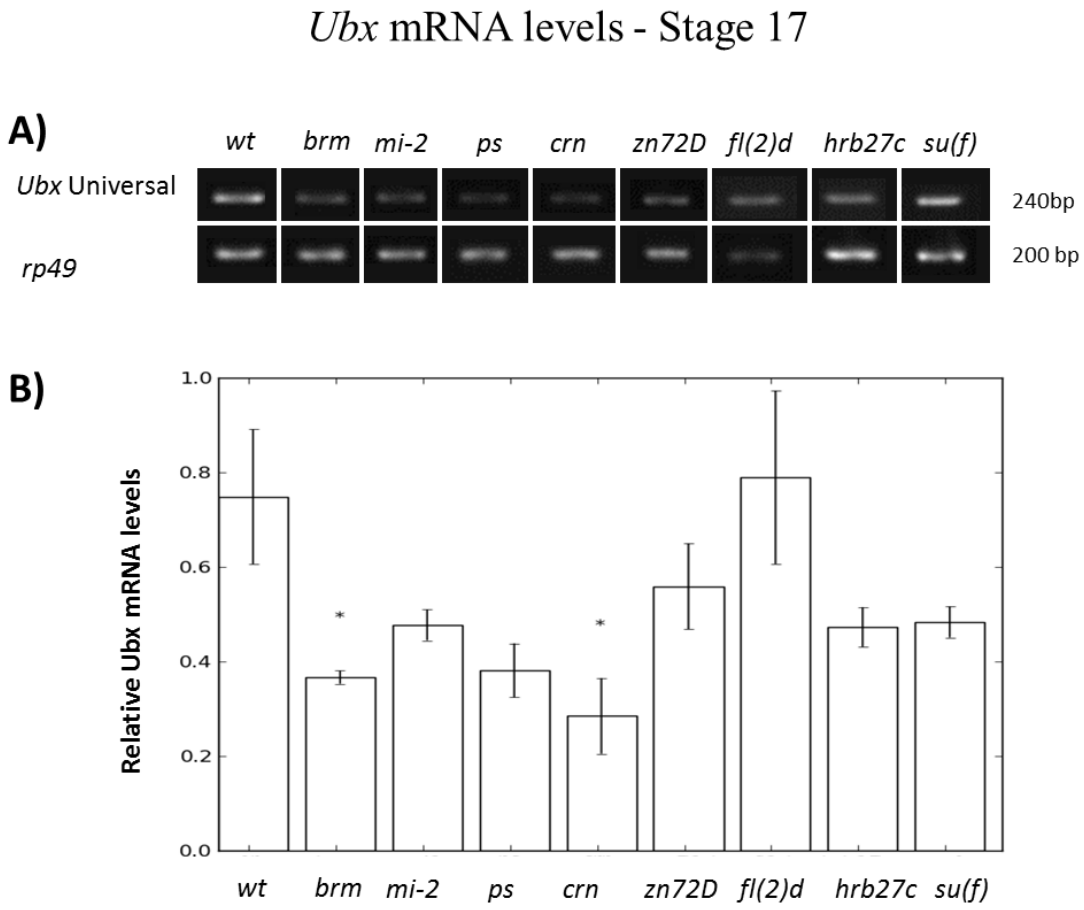


Figure 5.9: *Ubx* mRNA levels at stage 17: RT-PCR of three independent biological RNA samples A) RT-PCR of *Ubx* Universal region and *rp49* as loading control in each of the mutants cDNAs and wild type as control. B) Quantification of the signal normalised to *rp49* levels showing a reduction of *Ubx* mRNA in *brm, ps, crn* mutants. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to the wild type.

no longer important. Interestingly, like *fl(2)d*, *hrb27c* is also maternally provided and has a stong effect at earlier stages, but at later stages it seems to have a be less important effect, maybe explained by its lower levels of expression in the wild type. The decrease in mRNA levels in the other mutants could be explained by an RNA destabilizations due to defective RNA processing or indirect effect on transcription. On the contrary, *brm* and *crn* show no effect at stage 15 (Figure 5.8) but a statistically significant reduction of *Ubx* at stage 17 (Figure 5.9).

### 5.2.5 Effects on *Ubx* splicing patterns

The *Ubx* isoform patterns in a wild type fly change during normal development (O'Connor et al., 1988),(Kornfeld et al., 1989), (Thomsen et al., 2010). At earlier stages of development, the trend is to include both microexons, resulting in more production of isoform I and II and less production of isoform IV. At later stages, isoform IV starts to appear and by the end of embryonic development, the three isoforms have about the same abundance. To investigate whether the factors screened have a role in the splicing patterns of *Ubx*, RNA was extracted from three independent samples of at least 50 homozygote embryos. RT-PCR was performed using the primers previously used to assess the splicing of *Ubx* microexons by (Burnette et al., 1999) that are designed in the 5' exon and 3' exon spanning the two microexons (scheme in Figure 5.10A.).

In this experiment, RT-PCR of *Ubx* in selected mutants resulted in small variations relative to those seen with the wild type. This suggests that each factor contributes modestly to the final output.

As a general trend, the results show a decrease in mI inclusion at the two developmental stages tested, suggesting a role of the factors for this process. Microexon I inclusion at stage 15 is clearly reduced in the *brahma* mutant, and very little change was seen in *crn* and *fl(2)d* mutants. The other genotypes do not show any significant difference compared to the wild type control. mII inclusion at the same stage is in general increased in all mutants tested, particularly in the *Zn72D* mutant, which interestingly does not show an effect in mI inclusion. *su(f)* mutants also show the same trend as *zn72D*; no difference in mI exon inclusion but an increase in mII exon inclusion (Figure 5.10). This suggests that these factors do not affect alternative

splicing in general, but more likely work together with other factors present around microexon II to modulate its splicing. Interestingly, at this stage *brm* appears to affect only slightly the inclusion of microexon II, which is unexpected given its general role in splicing. *mi-2* shows a notable variation in the inclusion of microexon II, which makes it difficult to interpret. However, there is a small increase of mII inclusion, which contrasts with no difference in mI inclusion. *mi-2* is a chromatin remodelling factor and it has regulatory function in *Ubx* gene expression by making the chromatin less accessible to transcription factors. For this reason, *mi-2* was expected to affect the inclusion of both exons in the same way so its differentially effect in both *Ubx* microexons is unexpected. *pasilla* mutant shows a pattern very similar to *mi-2*, no change in mI exon inclusion and a small increase in mII exon inclusion (Figure 5.10).

Later in development at stage 17 (Figure 5.11), the role of the factors tested seem different, and there is a general trend to a reduction on mI and mII inclusion, leading to a slight increase in the production of II isoform. The *brm*, *mi-2*, *zn72D* and *hrb27c* mutants show a reduction of mI inclusion, whereas the others show a less marked effect. Interestingly, the *brm* mutant also slightly represses the inclusion of the mII microexon, suggesting that it has a general effect in exon inclusion, in agreement with previous studies in mammalian cells (Batsché et al., 2006). The *crn* mutant shows the same effect on splicing as the *brm* mutant, also consistently with its general role in the splicing machinery (Chung et al., 2002). However, we observe that the splicing of the constitutive exons (5' and 3' exons) is not affected.

The *mi-2* mutant shows a specific effect on mI inclusion, whereas mII seems not to be affected. This effect seems contrary to the one observed at stage 15, where mI inclusion was reduced the microexon affected in the mutant. The normal expression of *mi-2* at stage 15 is much higher than at stage 17, probably different levels of the protein results in different regulatory effects.

*pasilla* shows a very similar pattern to *mi-2*, but with a less marked reduction in mI inclusion, which is again contrary to the effect observed at stage 15. However, PS normal levels change considerably between stage 15 and 17. In contrast, the *zn72D* shows opposite effects in mI and II. Whereas it shows a reduction in mI inclusion, it is the only factor that shows a slight increase in mII inclusion. This suggests that

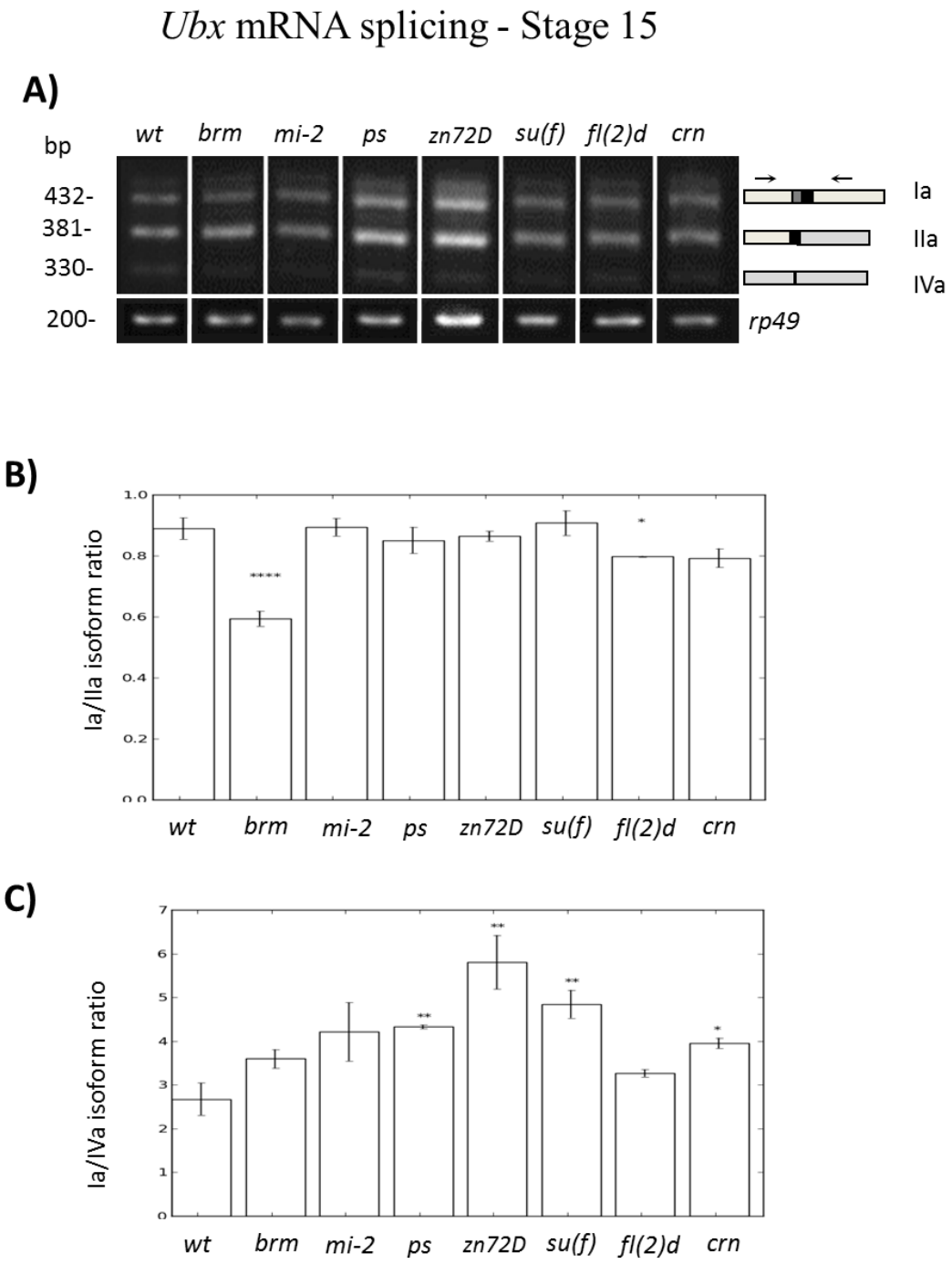


Figure 5.10: *Ubx* splicing patterns at stage 15: RT-PCR of three independent biological RNA samples. A) RT-PCR using oligos around the variable region showing the production of isoforms Ia, IIa or IVa in wild type or mutants for the factors tested. B) Quantification of the Ia and IIa bands intensities and represented as proportion of Ia/IIa isoforms. C) Quantification of the Ia and IVa bands intensities and represented as the proportion of Ia/IVa isoforms showing a reduction of Ia/IIa ratio in *brm* mutant and an increase in Ia/IVa ratio in *ps, zn72D, su(f) and crn* . Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to the wild type.

there may be a *cis* or *trans* factor around mII that is absent in mI context. *fl(2)d*, *su(f)* and *hrb27c* show a very similar pattern between them, a reduction in mI exon inclusion and a very mild reduction in mII inclusion. These data show that they affect both exons the same way at this developmental stage. However, *su(f)* shows a different trend that the one observed at earlier stages, where mII inclusion was promoted in the mutant background (Figure 5.11).

### 5.2.6 *Ubx* polyadenylation patterns

Polyadenylation site use in *Ubx* also changes during embryonic development. At early stages, the proximal site is prefered, generating a short 3' UTR and the distal site is used later in development, specifically in the developing CNS, resulting in long 3' UTR (O'Connor et al., 1988), (Kornfeld et al., 1989),(Thomsen et al., 2010). To assess the role of our candidate factors in polyadenylation site selection, primers complementary to the proximal region (Universal) or the distal region (Distal) were used in semiquantitative RT-PCR (Figure 5.9). The two pair of primers used in these experiments show different efficiency in the PCR reaction.

Preliminary results show little difference in the ratio of proximal to distal polyadelylation site selection in stage 15. *crn*, *hrb27c* and *su(f)* mutants show no difference in PAS choice, whereas mutations in *zn72D* and *fl(2)d* show higher levels of distal polyadenylation site use, generating longer 3' UTR. *su(f)* was expected to have a more significant effect on polyadenyation site selection however, at this stage, *su(f)* does not show any significant difference in long 3' UTR levels. In contrast, *mi-2* shows a slight decrease in the use of the distal PAS, generating shorter 3' UTRs (Figure 5.12). The presence of more mRNA with a long 3 'UTR in the *fl(2)d* mutant could explain the reduction of general *Ubx* mRNA levels seen in *Ubx* Universal RT-PCRs, as the distal portion of the UTR has been demonstrated to be involved in miRNA downregulation (Thomsen et al., 2010).

However, at the later stage of development, polyadenylation patterns change and the mutant for *crn* and *su(f)* show an increase in the use of the second polyadenylation site, generating more long 3' UTR (Figure 5.13). The effect in the *su(f)* mutant is consistent with its function as a member of the CstF (Benoit et al., 2002) in that it is needed for the first step of the 3' end formation reaction. However, the effect of

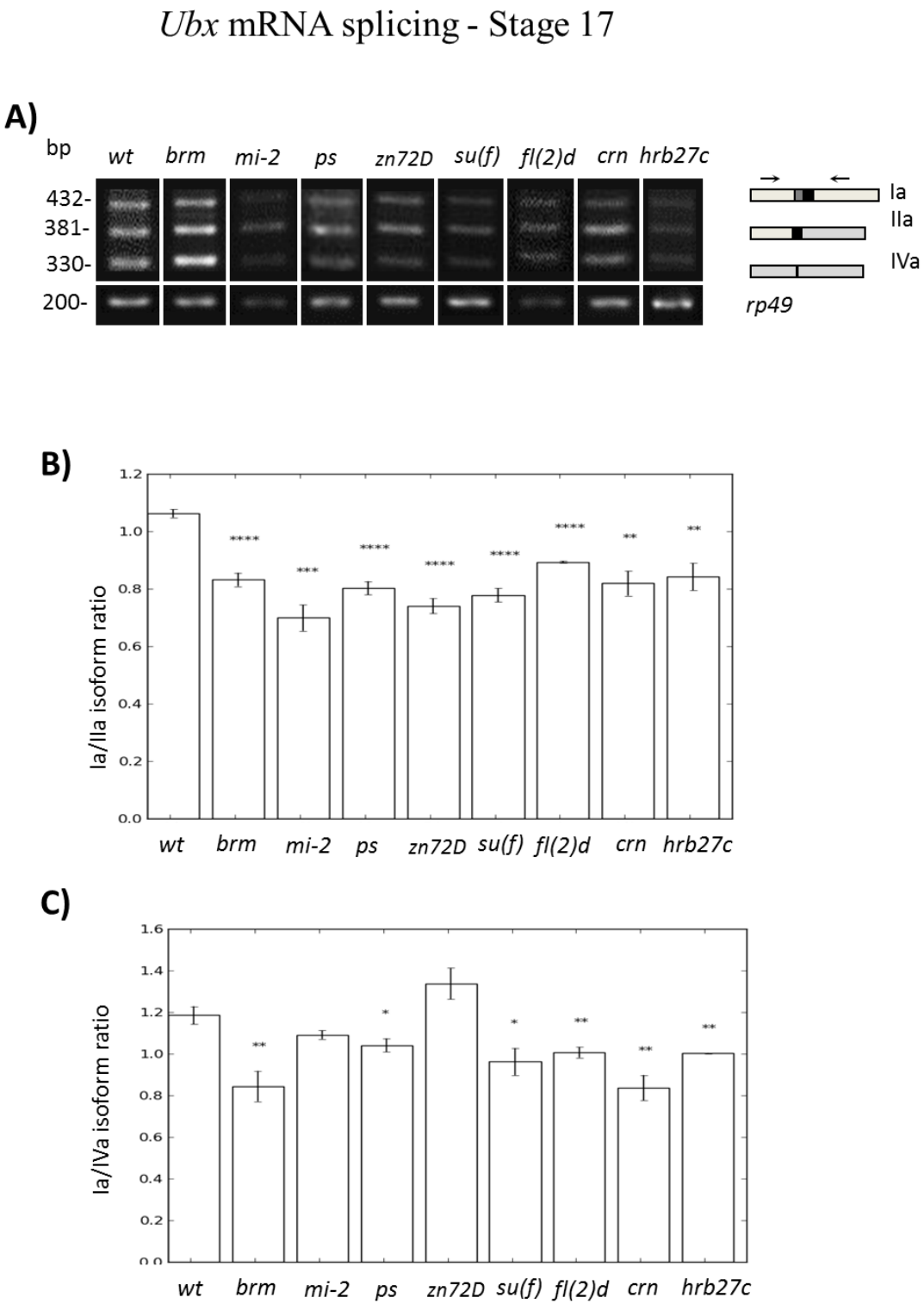


Figure 5.11: *Ubx* splicing patterns at stage 17: RT-PCR of three independent biological RNA samples. A) RT-PCR using oligos around the variable region showing the production of isoforms Ia, IIa or IVa in wild type or mutants for the factos tested. B) Quantification of the Ia and IIa bands intensities and represented as the ratio of Ia/IIa isoforms. C) Quantification of the Ia and IVa bands intensities and represented as the ratio of Ia/IVa isoforms showing that all of the mutants show a reduction in the Ia/IIa ratio and *brm,ps, su(f), fl(2)d, crn and hrb27c* show a reduction in the Ia/IVa ratio. Bars represent standard error and * p$\leq$ 0.05, ** p$\leq$ 0.01 of the t-test in respect to wild type

mut st15.png

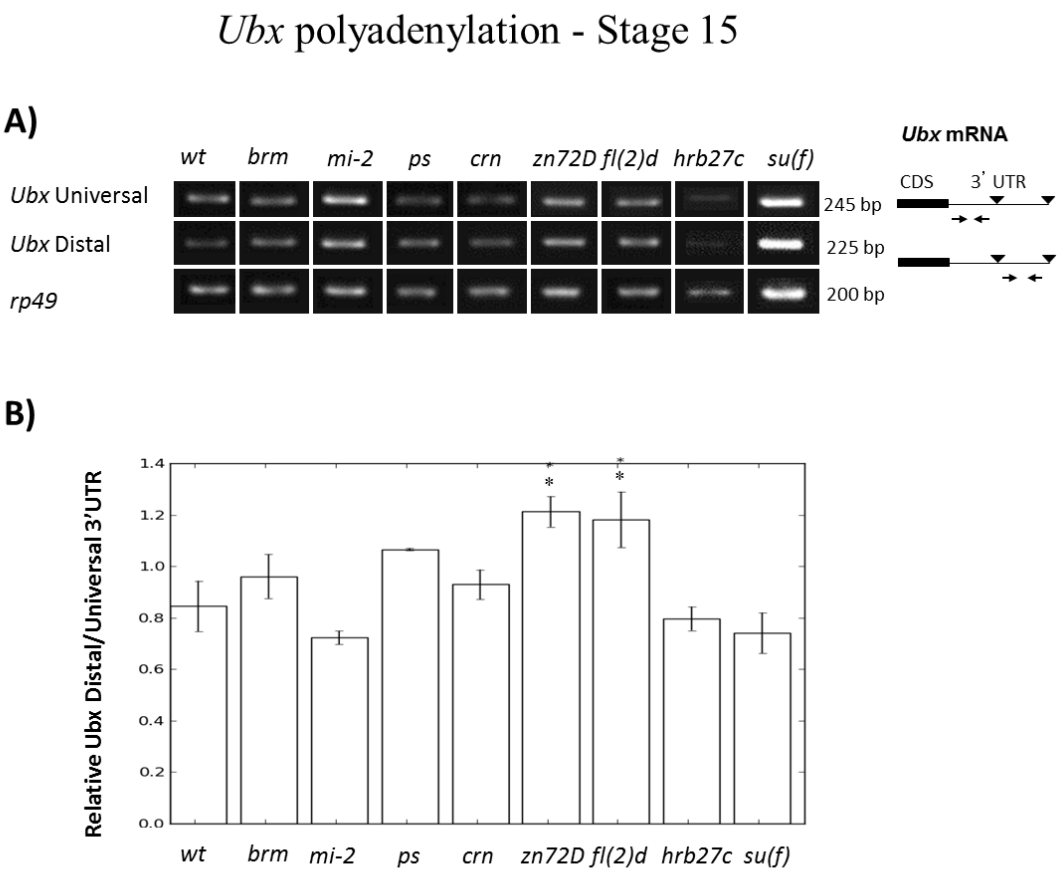


Figure 5.12: Polyadenylation site choice at stage 15: RT-PCR of three independent biological RNA samples. Primers designed complementary to the proximal region of distal region of the 3' UTR were used (see scheme). A) RT-PCR of *Ubx* Universal or Distal regions as shown in the right scheme and *rp49* as loading control. B) Quantification of the bands intensities and represented as *Ubx* Distal/*Ubx* Universal ratio, showing an increase in *zn72D and fl(2)d* mutants. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to the wild type.

the *crn* mutant is novel in that it has not been reported to have a role in polyadenylation. Nova is known to plays an important role in alternative polyadenylation in the mammalian brain (Licatalosi et al., 2008), so the mutant of the *Drosophila* homologue, *pasilla* was expected to show an impact in *Ubx* polyadenylation. However, I did not observed any significant effect.

This results indicate that the *Ubx* 3' end formation during *Drosophila* development is a dynamic process that depends on different factors at different time points. In line with these results, it has been documented that at later stages the production of a long UTR is predominant, compared to earlier stages (Hilgers et al., 2011). The presence of a long UTR allows a more strict control of the transport, translation and decay of this transcript and that one possible mechanism is downregulation by micro RNAs (Thomsen et al., 2010), (Patraquim et al., 2011), (Szostak and Gebauer, 2013).

These data points at a differential regulation of cassette exon inclusion between stages 15 and 17 (Figure 5.14). At stage 15, fewer factors have an impact in splicing than at stage 17 where almost all of the factors tested show an effect. Another interesting observation is that at stage 15 the factors act in opposite directions in microexon I and II inclusion whereas at stage 17 the effect on both microexons is to the same direction.

For polyadenylation decisions, the effect is to the same direction (factors repress the use of the distal polyadenylation site) but the factors involved are not the same at stage 15 and 17 (Figure 5.14).

### 5.2.7 Factors binding to *Ubx* probes *in vivo*-preliminary results

In the attempt to identify proteins bound to the *Ubx* RNA probes tested in this chapter, UV crosslinks followed by RNAse protection assays were performed as described in the Methods section using Nuclear Extracts form whole embryos at late stages. Preliminary results show several protection bands of different sizes in site 3 suggesting that there are several protein directly interacting with it *in vivo* Figure 5.15). The size of the bands suggest that the *Drosophila* homolog of *PTB*, haephesius, could be one of these factors (arrows in red in Figure 5.15). Further

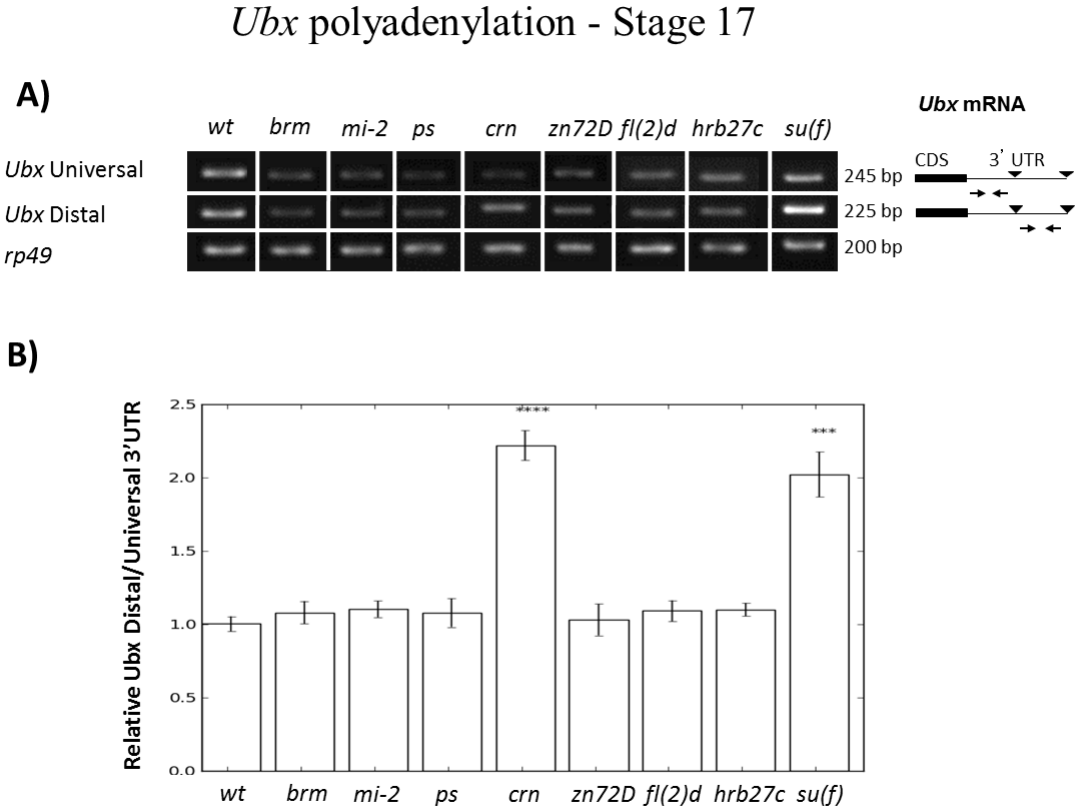


Figure 5.13: Polyadenylation site choice at stage 17: RT-PCR of three independent biological RNA samples. A) RT-PCR of *Ubx* Universal or Distal regions as shown in the right scheme and *rp49* as loading control. B) Quantification of the bands intensities and represented as *Ubx* Distal/*Ubx* Universal ratio showing an important increase in *crn and su(f)* mutants. Bars represent standard error and * $p \leq 0.05$, ** $p \leq 0.01$ of the t-test in respect to the wild type.

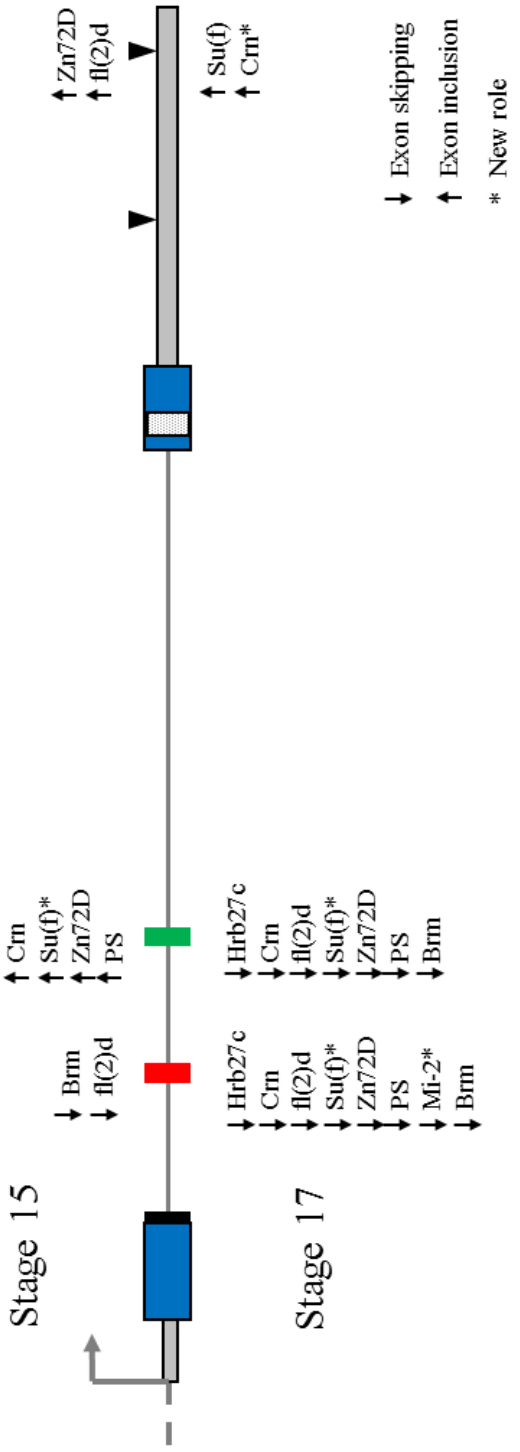


Figure 5.14: Schematic summary of the effects of the candidate factors in *Ubx* splicing and polyadenylation. The effects observed at stage 15 are shown above the scheme of the *Ubx* transcript and the ones observed at stage 17 are below. Arrow pointing upwards represents exon inclusion, arrow pointing downwards represent exon skiping. The asterisk marks the novel regulatory activities for the factor in turn.

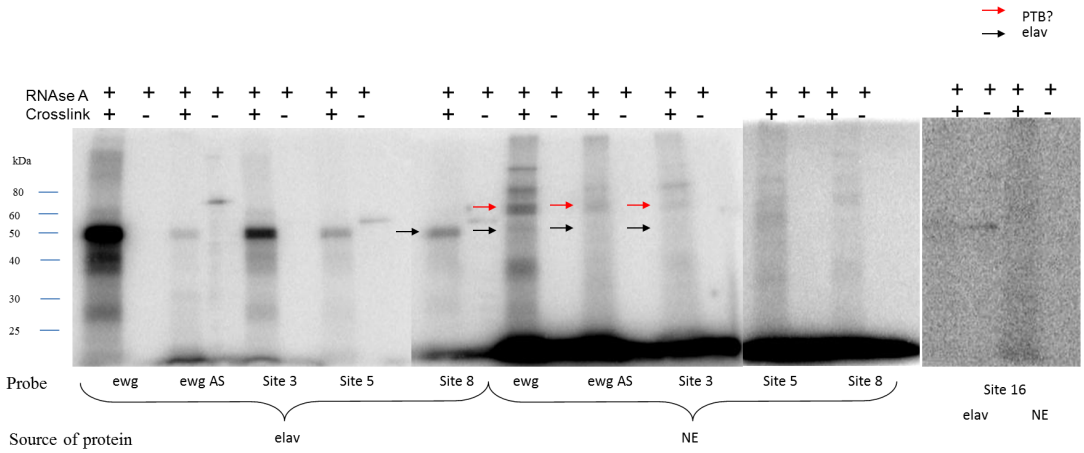


Figure 5.15: UV crosslink using Nuclear Extract reveal proteins binding to *Ubx* probes. UV-crosslink were performed using recombinant Elav as a control and nuclear extracts as marked in source of protein. The RNA probe used is marked at the bottom of the lanes. A no UV control was included and marked as - . Black arrows mark the position of the Elav protein whereas red arrows mark the position of possibly the PTB *Drosophila* homolog.

identification of the bands obtained is required.

## 5.3 Discussion

In this chapter I present preliminary data that suggests that a selection of protein factors affect the RNA processing of Ubx during embryonic development. The factors tested had a modest contribution to the overall splicing and polyadenylation output, and their contributions were not necessarily in the same direction at the two developmental stages tested. Presumably, the factors tested will affect a large spectrum of mRNA targets,leading to indirect effects in the regulation of *Ubx* transcription or RNA processing. I confirmed the role of the factors tested and uncovered new roles for some of them.

When the effect of the candidate proteins is observed globally, some interesting patterns appear (Figure 5.14). At stage 15, fewer factors play a role in exon inclusion and the effects are in contrary directions (exon skipping for mI and exon inclusion for mII) whereas at stage 17, most of the mutant factors tested showed reduced levels of exon inclusion. In other words, the protein factors tested act to stimulate the inclusion of both cassette exons.

*crn*, *fl(2)d* and *hrb27c* played a role in *Ubx* mRNA splicing as expected. How-

ever, *fl(2)d* showed a less effect at the time points sampled whereas it was previously shown to have an important role in larvae and adults. In my study, the two embryonic stages tested showed different effect from the protein factors tested, making evident the different roles they have along embryonic development (Figure 5.14).

*hrb27c* showed a clear effect in splicing at stage 17 but not at stage 15, which is surprising given that it is a factor that regulates alternative splicing. In contrast, *hrb27c* showed a marked effect in the overall levels of *Ubx* mRNA at the two time points tested. The mechanism for this role is unknown. One possible explanation is that it regulates the splicing of a key factor that affects the transcription of *Ubx*.

Javier Lopez and this study show that *crn* has a role the splicing of mI only, at larval and adult stages. In this study, I observed the same effect observed in the *crn* mutants by Javier Lopez lab at embryonic stages, which is to some extent, surprising, given that *crn* has been described as a general alternative splicing factor, so the same effect would be expected for both microexons. Interestingly, *crn* has also a role in regulating *Ubx* mRNA levels, probably related to its also observed role in polyadenylation site choice at stage 17. This is a very surprising result because a role for *crn* in polyadenylation has not been reported and this could be a coupled process between transcription, splicing and polyadenylation.

*brahma* showed a role in splicing of microexon I at stage 15, confirming previous reports of *brm* role in splicing in other systems (Batsché et al., 2006). *brahma* also showed a role in determining *Ubx* mRNA levels, probably related to its role in transcription. However, *brahma* did not show any significant role in alternative polyadenylation at the stages sampled. The attempt to map the brm2 mutant used in this study showed a silent point mutation that does not explain the mutant phenotype.

*mi-2* also showed an effect on *Ubx* transcript abundance but in the opposite direction at the two time points sampled. At stage 15, the result is as expected, being a repressor of *Ubx* transcription, *Ubx* is up-regulated in the *mi-2* mutant. However, at stage 17, I observed a decrease in *Ubx* mRNA. I observed an effect on splicing particularly of microexon I at stage 17. *mi-2* did not show any significant role in polyadenylation site choice at the stages sampled. *pasilla* has previously demonstrated to play roles in splicing and polyadenylation but in this study I only

observed an effect on splicing at stage 15. *ps* mutants also showed an effect on *Ubx* transcript levels, role not previously described. This suggests that it either plays a role in transcription regulation or mRNA stability, or the observations are the result of an indirect effect.

The only polyadenylation factor tested, *su(f)* showed a considerable effect in polyadenylation site choice at stage 17 as expected, but also showed an effect in splicing at stage 15, a function not previously reported. My data showed a role in *Ubx* transcript abundance but in the opposite direction as expected. *su(f)* is involved in RNA decay so I would have expected the mutation to up-regulate *Ubx* mRNA levels.

### 5.3.1 Splicing and polyadenylation relationship

Previous studies have found a correlation between splicing and polyadenylation patterns in *Ubx* (O'Connor et al., 1988), (Kornfeld et al., 1989), (Thomsen et al., 2010) where isoforms that have both microexons are proportionally more related to shorter UTRs and isoforms without any microexons are more associated with longer UTRs. These two processes could be controlled independently or they could be dependent on one another. Looking for connections of these two processes in the mutants tested, I observed that in the *crn* mutant, I do see a correlation; the mutant produces more isoform IV and produces also more "long" UTR. In this case, *crn* could link splicing and polyadenylation indirectly, or directly. This is the first evidence of *crn* having a role in polyadenylation. Interestingly, the same pattern is seen with *su(f)* and I observed the production of more isoform II lacking mI and a longer UTR. *su(f)* role in splicing seems to be more specific to the mI, unlike *crn* that seems to have a more general role in the inclusion of both exons. The other factors did not show any consistent relationship between splicing and polyadenylation.

The genome of *Drosophila melanogaster* has many more predicted RNA binding proteins that could also play an important role in shaping the splicing and polyadenylation landscape during development and *Ubx* in particular. The interaction between the factors tested also remains to be elucidated. Deciphering the factors and mechanisms involved in the control of RNA processing during development would shed light on the understanding on the establishment of the morphology

and function of organisms.

# Chapter 6

# Discussion

The aim of this project was to contribute to the understanding of the molecular basis of *Ubx* mRNA processing control during the embryonic development of Drosophila.

In this project I demonstrated that the Elav protein plays a role in splicing, polyadenylation and total mRNA level of *Ubx* during embryonic development. I also described two *cis*-elements within introns of *Ubx* that are able to interact directly with Elav *in vitro*. Other candidate factors were also tested for *Ubx* RNA regulatory activity and most of them play modest roles in splicing and polyadenylation patterning of *Ubx*. Interestingly, we did uncover some previously unknown roles for some of the factors.

This study is focused on the late stages of Drosophila embryonic development because is at this point when all the isoforms of *Ubx* are present, making it easier to observe changes in the mRNA patterns produced. Elav protein also is expressed at higher levels in later stages of development, making easier to measure the effects of its mutation.

## 6.1 The role and mechanism of Elav in *Ubx* RNA processing

In this project I demonstrated that Elav protein participates in the control of *Ubx* splicing, promoting exon exclusion of both microexons. The data suggests that Elav controls both microexons in the same direction. The precise mechanism for this role is unclear. However, in *in vitro* binding assays, I demonstrate that Elav binds to an

element (site 3) located 45 bp upstream of microexon I. This suggests the possibility that Elav masks the downstream splicing site, promoting exon exclusion. This mechanism is similar to the one described for the effect of Elav in polyadenylation, interfering with the usage of the proximal PAS in *erect wing* (Soller and White, 2003). The other regulatory element bound by Elav (site 8) *in vitro* is downstream of microexon II. This site is more than 8 kb away from the microexon II which makes it less likely to play a role in splicing. However, considering the secondary structure that RNA adopts and the factors that are bound to it while its being transcribed, we cannot rule out the possibility that site 8 plays a role on microexon II splicing.

Site 13 did not show significant binding in the *in vitro* binding assays, however, it is possible that a cofactor is needed for this interaction to occur *in vivo*. This site is relatively close (3.5 kb) to a described resplicing site (RP3), which makes it a candidate for resplicing control of the extraordinarily long third Intron of *Ubx* (Burnette et al., 2005).

I also demonstrated that Elav plays a critical role in *Ubx* PAS choice, making the transcript more visible to other regulatory factors such as miRNAs (Thomsen et al., 2010). Experiments by other members of the lab have shown that ectopic expression of Elav in earlier stages of embryonic development leads to a change in polyadenylation choice site, observed by *in situ* hybridisation experiments (Roguljа-Ortmann et al., 2014). The *in vitro* binding experiments did not show any significant binding of Elav to the only site near the proximal PAS tested (site 16). This, however, does not exclude its role in polyadenylation control *in vivo*. In other systems, it has been suggested that the interaction of Elav with a target mRNA is stabilised by ligands such as SETalpha, SETbeta, pp32, and APRIL (protein phosphatase inhibitors) (Brennan et al., 2000). It is possible that *in vivo*, an Elav ligand is required for the recognition and binding of the site 16, playing a role in polyadenylation site choice. Supporting this idea, the sequence between the two A/U sequences in site 16 contains a predicted PTB binding site.

The phenomenon of 3' UTR lengthening as development progresses has been observed in several transcripts within the developing CNS, where a more distal PAS is used at a later point in development, producing transcripts with progressively longer 3' UTRs (Hilgers et al., 2011). The Levine group has demonstrated that Elav

is responsible for such lengthening (Hilgers et al., 2012) in the embryonic CNS. It is to note that they did not pick up *Ubx* in their study. Our work confirms the UTR lenghtening in *Ubx* RNA although a putative binding site was not identified. The work of Levine's group show that the lengthening of *brat, ago1, nej, pum* and *imp* is severely affected by the loss of Elav protein whereas the change in PAS use observed in *Ubx* is less affected. Although in our study the effect might be contaminated by non neuronal tissue, this also could point out to a more complex regulatory system for *Ubx* polyadenylation during development, where the effect of Elav is mediated by a third factor. Our work expands to the role of Elav on splicing along embryonic development, demonstrating that Elav binds to *Ubx RNA* to regulate alternative splicing. We also extended our study to the effect of polyadenylation site choice on RNA levels and protein production. Our observation that total *Ubx* mRNA levels are reduced in an *elav* mutant background was contrary to our expectations, where an *Ubx* mRNA molecule containing a shorter 3' UTR would be less visible to miRNAs, so less down regulated (Thomsen et al., 2010). Modulation of Ubx protein levels leads to changes in the apoptosis pattern of GW neuron in the posterior thoracic and abdominal segments (Rogulja-Ortmann et al., 2014).

A possible explanation for this reduction is a "stalling" of non-appropriately spliced RNA molecule during transcription. Intronic *in situ* hybridisation experiments carried out by Joao Osorio in the Alonso Lab showed an up regulation of RNA associated to chromatin in *elav* mutant background compared to wild type late embryos (Rogulja-Ortmann et al., 2014). This effect has already been described in beta globin transcripts (Carmo-Fonseca et al., 1999).

Another possibility is that the Elav protein is promoting *Ubx* mRNA stability, as demonstrated in mammalian systems (Hinman, M.N; Lou, 2008), (Ziegeler et al., 2010), (Mukherjee et al., 2011), (Zhuang et al., 2013) so in its absence, *Ubx* RNA degradation would increase. However, Elav promotes deadenylation and RNA degradation in other systems, such as Xenopus Eg2 transcript (Legagneux et al., 1992), and translation repression (Standart, 1993). The HuR protein also able to interact with miRNAs to increase or reduce their activity on their targets (Kundu et al., 2012) (and references therein). HuR in some cases prevents or dismantles the miRISC complex from the mRNA (Kundu et al., 2012) and in other systems HuR

is necessary for miRISC recruitment to the target mRNA (Kim et al., 2009). In consequence, the interaction of Elav with the miRNA sites within the *Ubx* 3' UTR may play an important role for *Ubx* mRNA stability.

The Elav protein from Drosophila has important differences from its mammal homologues: it has been only detected within the nucleus of neuronal tissue, whereas the mammal proteins HuB/C and D are able to shuttle between the nucleus and cytoplasm in neuronal tissue (Colombrita et al., 2013). This allows the Hu proteins to participate in cytoplasmic functions, like mRNA translation. For example, the HuD mammal protein is capable of promoting mRNA circularisation in mammals, interacting with eIF4A and the polyA, stimulating translation (Brennan et al., 2000). The ubiquitously expressed HuR protein is able to control translation interacting with internal ribosome entry site (IRES) (Rivas-Aravena et al., 2009). However, a very close member of the Elav family, FNE, is mainly cytoplasmic, making it a possible candidate to carry on the cytoplasmic activities of the mammal counterparts.

## 6.2 Bio-informatic tools in the search of Elav binding sites

The efforts to elucidate a precise consensus sequence for Elav/HuR have been partially succesful and have found that the protein is flexible to some extent in its specificity requirements. A sequence rich in A/U is the common characteristic in all the sites tested, from Drosophila (Haussmann et al., 2011) to Xenopus (Legagneux et al., 1992) and mammalian systems (Wang and Hall, 2001a), (López de Silanes et al., 2004). The minimal internal motif makes difficult to identify effectively functional sites. Even within one of the most well characterised sites, within the *erect wing* transcript, a comparison between the *Drosophila melanogaster* and *virilis* sequence, reveals that the sequence elements is only partially conserved, but the interaction with Elav remains the same (Haussmann et al., 2011).

Within the cell, the biochemical composition is different from the conditions in the *in vitro* experiments described here, so it is likely that the secondary structure of the RNAs within the cell, under physiological conditions would be different (Lai et al., 2013), (Rouskin et al., 2013). Moreover, additional factors, such as RNA bind-

ing proteins that are bound to the RNA molecule as it is transcribed might change the secondary structure of the final transcript, probably modifying the structure where the putative Elav binding sites lie, modifying their affinity to Elav and other RNA binding proteins. There is evidence supporting this, suggesting that the RNA starts folding as it is transcribed, following a determined folding pathway that will dictate the final folding and in consequence, the function of the RNA (Lai et al., 2013). One example of the role of protein in RNA folding is the folding of the yeast mitochondrial group II intron ai5, aided by the protein cofactor Mss116 (Fedorova, 2014).

## 6.3 Other factors affecting *Ubx* RNA processing

In the experiments carried out to test the role of other factors in *Ubx* RNA processing, the results suggest that each factor has a modest contribution to the overall *Ubx* splicing and polyadenylation patterns. Moreover, their contributions were not necessarily in the same direction at the two developmental stages tested. Presumably, the factors tested will affect a large spectrum of targets, making possible that those targets are available at different time points, that could in turn, affect *Ubx* transcription or RNA processing.

Interestingly, I under covered new roles for some of the factors. For example, the Crn factor that had been described as a splicing factor, in a crn mutant background, the levels of *Ubx* are reduced. Crn also showed an effect on polyadenylation site choice, role not described before. At stage 17, the levels of longer 3' UTR in the crn mutant background are increased very significatively, if compared to the wild type. The effect is comparable with the one observed for the *su(f)* mutant, factor that has a known role in polyadenylation. The Su(f) factor in turn, shows a significant effect in splicing in my experiments. In a *su(f)* mutant background, mII inclusion is promoted.

The chromatin remodelling factor, Mi-2 revealed a role in splicing, not described before. The Mi-2 mutant shows an increase on mII inclusion, but no effect on mI.

The molecular basis for the effects observed are not known and its study will help to further understand the cooperation of those factors to shape the RNA processing

patterns during development.

## 6.4 Future work

The work presented here sets the start of a bigger project that aims to elucidate the factors that shape the control of *Ubx* mRNA processing. The strategy used here was useful but requires improvement. The bioinformatic search of Elav putative binding sites can be improved adding a more detailed secondary structure analysis, which has helped in other systems (Meisner et al., 2004), (Li et al., 2010). This strategy could be also applied to search putative binding sites for the other candidate factors demonstrated in this study to participate in *Ubx* mRNA processing, for their posterior testing for direct interaction. A more sensitive technique such as Real Time PCR can be used for the analysis of splicing and polyadenylation isoforms to improve the quantifications and reduce the huge variation obtained by agarose gel-ethidium bromide quantifications done in this study.

The RT-PCR experiments were carried out using RNA extracted from whole embryos, in consequence, the tissue specific information was lost. It would be very useful to study the mRNA processing of *Ubx* in isolated tissues, comparing neuronal versus non neuronal context. We could then over express Elav in non neuronal tissues and see if it is able to recapitulate the *Ubx* neuronal processing profile.

In our lab, a collection of *Ubx* "minigenes" has been constructed by Dr. Stefan Thomsen modifying the Ubx minigene produced by (Burnette et al., 2005) (Figure 6.1). The sequence of the introns of *Ubx* was shortened and cloned into constructs that were injected into flies. The intronic sequences of the minigenes contains site 3, 5 and 16 sequences. If we express the minigenes in an *elav* null mutant, we can test if the splicing and polyadenylation of the minigene is affected. We can also use the minigene constructs to mutate the identified sites and test them in flies or in S2 cells and test if its processing is affected.

Another way of testing the sites, could be within a S2 cell system transfecting a splicing construct and modulating Elav levels, and using RT-PCR to monitor the splicing isoforms of *Ubx* produced. The constructs can also contain mutated versions of the sites to test for changes in splicing activity.

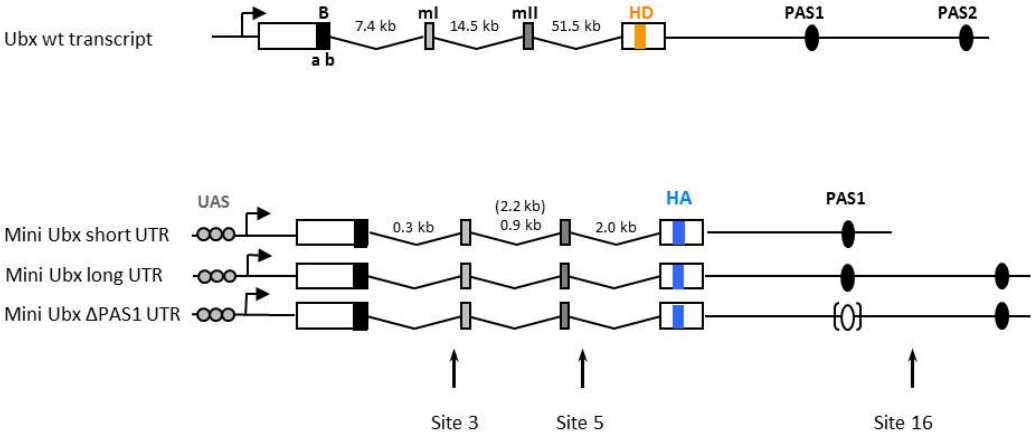


Figure 6.1: The *Ubx* minigene system. Top scheme represents the *Ubx* wt transcript, numbers show the sizes of the introns. HD, homeodomain. Below, modified minigenes with shortened introns. The Homeodomain is substituted with two copies of the HA (Haemaglutinin peptide) to allow the detection of the mRNA by PCR or protein using an anti HA antibody. The positions of putative Elav binding sites are marked with arrows. Mini *Ubx* short stops just after the proximal polyadenylation site, the long minigene stops after the distal polyadenylation site whereas the third construct has the proximal polyadenylation site deleted.

To test whether the sites identified to have binding activity *in vivo*, have any role in splicing, *in vitro* splicing assays could be carried out using Elav protein and a splicing construct containing the sites identified and test them against their mutations (Ares, 2013). The role of the sites identified in polyadenylation can also be addressed using *in vitro* systems, like the ones described in (Roca and Karginov, 2012).

ARE binding proteins have been described as factors that promote mRNA decay, since deadenylation is frequently the first step in RNA decay (Standart, 1993), we can test the polyadenylation status of each o the isoforms produced by *Ubx* in the presence of absence of Elav or when the putative sites are mutated (Huang and Richter, 2007).

Our data show that Elav protein produces RNA/protein complexes of different sizes, suggesting the association of different number of Elav peptides in each complex. We also observed that the number of intermediate complexes was different in the RNA probes tested. In order to investigate the number of Elav peptides in each complex, size exclusion chromatography experiments using the Elav/RNA reaction

mixtures for EMSAS can be performed (Sahin and Roberts, 2012).

Elav could potentially use sites that we did not detect with the methods used in this study, that could be identified by other approaches. Using whole embryo lysates, UV-crosslink followed by immunoprecipitation using and anti-Elav antibody and RNA sequencing could lead to the discovery of other sites that can be tested for *in vitro* binding activity or for functional splicing or polyadenylation roles (Sugimoto et al., 2012), (Ankö and Neugebauer, 2012). Modifications of the CLIP (Cross Link Immuno Precipitation) approach have yielded good results in other systems (Singh et al., 2013) and could be implemented here to identify Elav binding sites *in vivo*.

A possible strategy to search for candidate factors affecting *Ubx* mRNA processing is UV-crosslink using Nuclear Extracts from whole embryos or tissue enriched in neuronal cells (adult or larval heads) and labelled RNA from fragments of *Ubx* (Walker et al., 1998). After RNAse digestion, a PAGE is run and the proteins protecting the fragments can be identified by size or antibody staining. Alternatively, RNA-IP using antibodies against Elav or the other candidate factors offer a good strategy to identify RNA elements bound to this factors(Rodgers et al., 2002), (Huber and Zhao, 2010),(Sibley et al., 2012).

To be able to identify other factors that are involved in *Ubx* mRNA processing, another approach can be used, where sequences of the *Ubx* RNA are used to collect and identify the proteins bound to it (Sibley et al., 2012). The proteins identified then could be tested further for *in vitro* binding activity or *in vivo* assays to test their precise role in RNA processing.

## 6.5 Final conclusions

The RNA processing of the *Ubx* gene is very dynamic during embryonic development of *Drosophila*. The establishment of the appropriate patterns of RNA splicing and polyadenylation isoforms is a crucial event for the correct functioning of the *Ubx* gene. Several factors play a role in the establishment of such patterns and the results of this project contribute to the identification of some of them. My data also contributes to the understanding of the molecular mechanisms for the Elav protein role in splicing by demonstrating a direct interaction between Elav and two

cis elements within *Ubx* introns.

# Appendix A

# Chapter 1: Primers used

| Name | Sequence | Position/use |
|---|---|---|
| UbxE1F | tggaatgccaattgcaccatc | 5' Exon Ubx |
| Ubx.E3.ORF.R | cctgcttctcctgttcgttc | 3' Exon Ubx |
| Ubx.short.F | gaaatgacgcggagacagat | proximal UTR |
| Ubx.E3.R-5 | aatctgcgctccttccacta | proximal UTR |
| Ubx.E3.F-4 | gaacgaaggcagatgcaaat | distal UTR |
| Ubx.E3.R-4 | ggtaagtggtcggatgcagt | distal UTR |
| Rp49L | ccagtcggatcgatatgctaa | RNA input |
| Rp49R | tctgcatgagcaggacctc | RNA input |
| ubx.E5.fw | gccaccggaatggcaatggg | Ubx levels |
| ubx.E5.rv | agggactcatgcccagcgag | Ubx levels |

# Appendix B

# Chapter 2: Primers used

| Name | Sequence |
|---|---|
| ewg.F | atGGTACCattaatttttatttttaatt |
| ewg.R | tatGTCGACagtataaaattataagataa |
| site3.F | tatGGTACCcatatggtccaaacaccatcc |
| site3.R | tatGTCGACaagcaaagcggagagaagaa |
| site5.F | atGGTACCcccttagcctatagttttgcctat |
| site5.R | tatGTCGACgcaacgcatatcgaataagtga |
| site8.F | acatttatccccacgtcgtt |
| site 8.R | tcagcgaaatgtgacagaaa |
| site13.F | gcgaccactgcagaatgac |
| site13.R | ccaacaaaagctgacagcag |
| site16.F | atGGTACCGAAGATGTATACGTACTACCGCCTAA |
| site16.R | tatGTCGACGACTCGATCCCCACATACATA |
| site3mut.F | tgagagcaaacaattttgatttcacagtgtgtaaacgttttcttctctccgctttgc |
| site3mut.R | gcaaagcggagagaagaaaacgtttacacactgtgaaatcaaaattgtttgctctca |
| site8Amut.F | gcattatcagttaaaccaaatgtgtgtgtgtatcaaatttctgtcac |
| site8Amut.R | gtgacagaaatttgatacacacacacatttggtttaactgataatgc |
| site8Bmut.F | acatttatccccacgtcgttgctgtgtgtgttttaccctctccgg |
| site8Bmut.R | ccggagagggtaaaacacacacagcaacgacgtggggataaatgt |

# Appendix C

# Chapter 2: Constructs used

| Insert | Vector | Promoter |
|---|---|---|
| ewg | pBluescript | T7 |
| ewg as | pBluescript | SP6 |
| Site 3 | pBluescript | T7 |
| Site 5 | pGEMT | T7 |
| Site 8 | pGEMT | T7 |
| Site 13 | pGEMT | SP6 |
| Site 16 | pBluescript | T3 |
| | | |
| ewg | pGEMT easy | T7 |
| ewg as | pGEMT easy | T7 |
| Site 3 | pGEMT easy | T7 |
| Site 8 | pGEMT easy | T7 |
| Site 13 | pGEMTeasy | T7 |
| Site 16 | pGEMT easy | T7 |

# Appendix D

# Chapter 3: Primers used

| Name | Sequence |
|---|---|
| crn.F | tctccatctcggcgtatttc |
| crn.R | aggagcagcagcaggagat |
| suf.FW | ctagtgtgaccgcagcgtaa |
| suf.R | agcacaacgggcaaaataat |
| hrb27c.F | cggtcacacagcattaaacg |
| hrb27c.R | cggtcgcttttcgttcttac |
| fl(2)d.F | gtcggttgggctcactttta |
| fl(2)d.R | catgttggattgcgaacaa |
| PS.F | cacacacacattcccgagtc |
| PS.R | ctccgctctgctctctgg |
| mi-2.F | gatcgattctctcccgactg |
| mi-2-R | gtgaatgctggcaatgttgt |
| brm.F | ccgttttgtgtgcctctttt |
| brm.R | ttagcgtgcagttgatttcg |
| zn72d.F | ctagtgtgaccgcagcgtaa |
| zn72d.R | agcacaacgggcaaaataat |
| pz5p.rv | tttgggagttttcaccaagg |

# Appendix E

# Publication

# Bibliography


Adams, M. D. and Sekelsky, J. J. (2002). From sequence to phenotype: reverse genetics in Drosophila melanogaster. *Nature Reviews. Genetics*, 3(3):189–98.

Akam, M. E. and Martinez-arias, A. (1985). The distribution of Ultrabithorax transcripts in Drosophila embryos. *The EMBO Journal*, 4(7):1689–1700.

Allen, S. E. and Lipscombe, D. (2010). The neuronal splicing factor Nova controls alternative splicing in N-type and P-type Ca V 2 calcium channels. *Channels*, 4(6):52–51.

Alonso, C. R. (2002). Hox Proteins : Sculpting Body Parts by Activating Localised Cell Death Hox proteins shape animal structures by eliciting. *Current Biology*, 12(02):776–778.

Alonso, C. R. (2012). A complex 'mRNA degradation code' controls gene expression during animal development. *Trends in Genetics : TIG*, 28(2):78–88.

Alonso, C. R. and Akam, M. (2003). A Hox gene mutation that triggers nonsense-mediated RNA decay and affects alternative splicing during Drosophila development. *Nucleic Acids Research*, 31(14):3873–3880.

Alonso, C. R. and Wilkins, A. S. (2005). The molecular elements that underlie developmental evolution. *Nature Reviews, Genetics*, 6(September):709–715.

Ambros, V. (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell*, 113(6):673–6.

Andreassi, C. and Riccio, A. (2009). To localize or not to localize: mRNA fate is in 3'UTR ends. *Trends in Cell Biology*, 19(9):465–74.

Ankö, M.-L. and Neugebauer, K. M. (2012). RNA-protein interactions in vivo: global gets specific. *Trends in Biochemical Sciences*, 37(7):255–62.

Antic, D. and Keene, J. D. (1997). Embryonic lethal abnormal visual RNA-binding proteins involved in growth, differentiation, and posttranscriptional gene expression. *American Journal of Human Genetics*, 61(2):273–8.

Ares, M. (2013). Analysis of splicing complexes on native gels. *Cold Spring Harb Protoc*, 10:986–9.

Armstrong, J. a., Papoulas, O., Daubresse, G., Sperling, A. S., Lis, J. T., Scott, M. P., and Tamkun, J. W. (2002). The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *The EMBO Journal*, 21(19):5245–54.

Artero, R., Akam, M., and Perez-Alonso, M. (1992). Oligonucleotide probes detect splicing variants in situ in Drosophila embryos. *Nucleic Acids Research*, 20(21):5687–5690.

Barker, A., Epis, M. R., Porter, C. J., Hopkins, B. R., Wilce, M. C. J., Wilce, J. a., Giles, K. M., and Leedman, P. J. (2012). Sequence requirements for RNA binding by HuR and AUF1. *Journal of Biochemistry*, 151(4):423–37.

Barreau, C., Paillard, L., and Osborne, H. B. (2005). AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Research*, 33(22):7138–50.

Batsché, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nature Structural & Molecular Biology*, 13(1):22–9.

Bava, F.-A., Eliscovich, C., Ferreira, P. G., Miñana, B., Ben-Dov, C., Guigó, R., Valcárcel, J., and Méndez, R. (2013). CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature*, 495(7439):121–5.

Bender, W. (2008). MicroRNAs in the Drosophila bithorax complex. *Genes & Development*, 22(1):14–9.

Bender, W., Akam, M., Karch, F., Beachy, P. a., Peifer, M., Spierer, P., Lewis, E. B., and Hogness, D. S. (1983). Molecular Genetics of the Bithorax Complex in Drosophila melanogaster. *Science (New York, N.Y.)*, 221(4605):23–9.

Benoit, B., Juge, F., Iral, F., Audibert, A., and Simonelig, M. (2002). Chimeric human CstF-77/Drosophila Suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10593–8.

Benoit, R. M., Meisner, N.-C., Kallen, J., Graff, P., Hemmig, R., Cèbe, R., Ostermeier, C., Widmer, H., and Auer, M. (2010). The x-ray crystal structure of the first RNA recognition motif and site-directed mutagenesis suggest a possible HuR redox sensing mechanism. *Journal of Molecular Biology*, 397(5):1231–44.

Bevilacqua, A., Ceriani, M. C., Capaccioli, S., and Nicolin, A. (2003). Post-transcriptional regulation of gene expression by degradation of messenger RNAs. *Journal of Cellular Physiology*, 195(3):356–72.

Bhattacharjee, S., Renganaath, K., Mehrotra, R., and Mehrotra, S. (2013). Combinatorial Control of Gene Expression. *BioMed Research International*, 2013:407263.

Bisturia, A., Vernos, I., Macias, A., Casanova, J., and Morata, G. (1990). Different forms of Ultrabithorax proteins generated by alternative splicing are functionally equivalent. *EMBO Journal*, 9(11):3551–3555.

Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, 72:291–336.

Blanchette, M., Green, R. E., Brenner, S. E., and Rio, D. C. (2005). Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila. *Genes & Development*, 19(11):1306–14.

Blanchette, M., Green, R. E., MacArthur, S., Brooks, A. N., Brenner, S. E., Eisen, M. B., and Rio, D. C. (2009). Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. *Molecular Cell*, 33(4):438–49.

Bomze, H. M. and Lopez, A. J. (1994). Evolutionary Conservation of the structure and expression of alternatively spliced Ultrabithorax isoforms from Drosophila. *Genetics*, 136:965–977.

Borgeson, C. D. and Samson, M.-L. (2005). Shared RNA-binding sites for interacting members of the Drosophila ELAV family of neuronal proteins. *Nucleic acids research*, 33(19):6372–83.

Brehm, a., Längst, G., Kehle, J., Clapier, C. R., Imhof, A., Eberharter, A., Müller, J., and Becker, P. B. (2000). dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *The EMBO Journal*, 19(16):4332–41.

Brennan, C. M., Gallouzi, I. E., and Steitz, J. a. (2000). Protein ligands to HuR modulate its interaction with target mRNAs in vivo. *The Journal of Cell Biology*, 151(1):1–14.

Brizuela, B. J., Elfring, L., Ballad, J., Tamkun, J. W., and Kennison, J. A. (1994). Genetic Analysis of the brahma Gene of Drosophila melanogaster and Polytene Chromosome Subdivisions 72AB B. *Genetics*, 137:803–813.

Brody, J. R. and Kern, S. E. (2004). Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *BioTechniques*, 36(2):214–6.

Brooks, A. N., Yang, L., Duff, M. O., Hansen, K. D., Park, J. W., Dudoit, S., Brenner, S. E., and Graveley, B. R. (2011). Conservation of an RNA regulatory map between Drosophila and mammals. *Genome Research*, 21(2):193–202.

Burnette, J. M., Hatton, R., and Lopez, J. (1999). Trans-acting factors required for inclusion of regulated exons in the Ultrabithorax mRNAs of Drosophila melanogaster. *Genetics*, 151(4):1517–29.

Burnette, J. M., Miyamoto-Sato, E., Schaub, M. a., Conklin, J., and Lopez, a. J. (2005). Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. *Genetics*, 170(2):661–74.

Bushati, N., Stark, A., Brennecke, J., and Cohen, S. M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. *Current Biology : CB*, 18(7):501–6.

Campos, A.R., Grossman, D., White, K. (1985). Mutant alleles at the locus elav in Drosophila melanogaster lead to nervous system defects. A developmental-genetic analysis. *J. Neurogenet*, 2:197—-218.

Carey, M. (2005). Chromatin marks and machines, the missing nucleosome is a theme: gene regulation up and downstream. *Molecular Cell*, 17(3):323–30.

Carmo-Fonseca, M., Geraghty, F., Pereira, H. S., Grosveld, F., and Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. *The EMBO Journal*, 18(10):2855–2866.

Chung, S., Zhou, Z., Huddleston, K. a., Harrison, D. a., Reed, R., Coleman, T. a., and Rymond, B. C. (2002). Crooked neck is a component of the human spliceosome and implicated in the splicing process. *Biochimica et Biophysica Acta*, 1576(3):287–97.

Colegrove-Otero, L. J., Minshall, N., and Standart, N. (2005). RNA-binding proteins in early development. *Critical reviews in biochemistry and molecular biology*, 40(1):21–73.

Colgan, D. F. and Manley, J. L. (1997). Mechanism and regulation of mRNApolyadenylation. *Genes & Development*, 11(21):2755–2766.

Colombrita, C., Silani, V., and Ratti, A. (2013). ELAV proteins along evolution: Back to the nucleus? *Molecular and Cellular Neurosciences*, 56(447455).

Dai, W., Zhang, G., and Makeyev, E. V. (2012). RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. *Nucleic Acids Research*, 40(2):787–800.

de la Mata, M., Alonso, C., Kadener, S., Fededa, J., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell*, 12:525–532.

de Navas, L. F., Reed, H., Akam, M., Barrio, R., Alonso, C. R., and Sánchez-Herrero, E. (2011). Integration of RNA processing and expression level control modulates the function of the Drosophila Hox gene Ultrabithorax during adult development. *Development (Cambridge, England)*, 138(1):107–16.

Di Giammartino, D. C., Nishida, K., and Manley, J. L. (2011). Mechanisms and consequences of alternative polyadenylation. *Molecular Cell*, 43(6):853–66.

Dieterich, C. and Stadler, P. F. (2013). Computational biology of RNA interactions. *Wiley iIterdisciplinary Reviews. RNA*, 4(1):107–20.

Dixit, R., VijayRaghavan, K., and Bate, M. (2013). Hox Genes and the Regulation of Movement in Drosophila. *Developmental Neurobiology*, 68(3):309–316.

Dudickz, M. E., Wright, T. R. F., and Brothers, L. L. E. E. (1974). The developmental genetics of the temperature sensitive lethal allele of the Suppressor of Forked l(1)su(f) ts679, in Drosophila melanogaster. *Genetics*, 76:487–510.

Eberl, D. F. and Hilliker, A. J. (1988). Characterization of X-Linked Recessive Lethal Mutations Affecting Embryonic Morphogenesis in Drosophila melanogaster Daniel. *Genetics*, 118:109–120.

Edenfeld, G., Volohonsky, G., Krukkert, K., Naffin, E., Lammel, U., Grimm, A., Engelen, D., Reuveny, A., Volk, T., and Klämbt, C. (2006). The splicing factor crooked neck associates with the RNA-binding protein HOW to control glial cell maturation in Drosophila. *Neuron*, 52(6):969–80.

Elkon, R., Drost, J., van Haaften, G., Jenal, M., Schrier, M., Vrielink, J. a. O., and Agami, R. (2012). E2F mediates enhanced alternative polyadenylation in proliferation. *Genome Biology*, 13(7):R59.

Ellis, J. D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J. a., Wang, X., Pan, Q., O'Hanlon, D., Kim, P. M., Wrana, J. L., and Blencowe, B. J. (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. *Molecular Cell*, 46(6):884–92.

Enroth, S., Bornelöv, S., Wadelius, C., and Komorowski, J. (2012). Combinations of histone modifications mark exon inclusion levels. *PloS One*, 7(1):e29911.

Estacio-Gómez, A., Moris-Sanz, M., Schäfer, A.-K., Perea, D., Herrero, P., and Díaz-Benjumea, F. J. (2013). Bithorax-complex genes sculpt the pattern of leucokinergic neurons in the Drosophila central nervous system. *Development (Cambridge, England)*, 140(10):2139–48.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2009). Deadenylation is a widespread effect of miRNA regulation. *RNA (New York, N.Y.)*, 15(1):21–32.

Euskirchen, G., Auerbach, R. K., and Snyder, M. (2012). SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *The Journal of Biological Chemistry*, 287(37):30897–905.

Fasulo, B., Deuring, R., Murawska, M., Gause, M., Dorighi, K. M., Schaaf, C. a., Dorsett, D., Brehm, A., and Tamkun, J. W. (2012). The Drosophila MI-2 chromatin-remodeling factor regulates higher-order chromatin structure and cohesin dynamics in vivo. *PLoS Genetics*, 8(8):e1002878.

Fedorova, O. (2014). A Chemogenetic Approach to Study the Structural Basis of Protein-Facilitated RNA Folding. *Methods in Molecular Biology*, 1086:177–191.

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature reviews. Genetics*, 9(2):102–14.

Fleming, R. J., Zusman, S., and White, K. (1983). Developmental Genetic Analysis of Lethal Alleles at the ewg Locus and Their Effects on Muscle Development in Drosophila melanogaster. *Developmental Genetics*, 363.

Flynt, A. S. and Lai, E. C. (2008). Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nature Reviews. Genetics*, 9(11):831–42.

Fukao, A., Sasano, Y., Imataka, H., Inoue, K., Sakamoto, H., Sonenberg, N., Thoma, C., and Fujiwara, T. (2009). The ELAV protein HuD stimulates cap-dependent translation in a Poly(A)- and eIF4A-dependent manner. *Molecular cell*, 36(6):1007–17.

Gabut, M., Samavarchi-Tehrani, P., Wang, X., Slobodeniuc, V., O'Hanlon, D., Sung, H.-K., Alvarez, M., Talukder, S., Pan, Q., Mazzoni, E. O., Nedelec, S., Wichterle, H., Woltjen, K., Hughes, T. R., Zandstra, P. W., Nagy, A., Wrana, J. L., and Blencowe, B. J. (2011). An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell*, 147(1):132–46.

Gamberi, C., Johnstone, O., and Lasko, P. (2006). Drosophila RNA binding proteins. *International Review of Cytology*, 248(06):43–139.

Garcia-Bellido, A., Ripoll, P., and Morata, G. (1976). Developmental compartmentalization in the dorsal mesothoracic disc of Drosophila. *Developmental Biology*, 48(1):132–147.

Goldstrohm, a. C., Greenleaf, a. L., and Garcia-Blanco, M. a. (2001). Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene*, 277(1-2):31–47.

Goodrich, J. S., Clouse, K. N., and Schüpbach, T. (2004). Hrb27C, Sqd and Otu cooperatively regulate gurken RNA localization and mediate nurse cell chromosome dispersion in Drosophila oogenesis. *Development (Cambridge, England)*, 131(9):1949–58.

Granadino, B., Campuzano, S., and Sánchez, L. (1990). The Drosophila melanogaster fl(2)d gene is needed for the female-specific splicing of Sex-lethal RNA. *The EMBO Journal*, 9(8):2597–602.

Hammond, L. E., Rudner, D. Z., Kanaar, R., and Rio, D. C. (1997). Mutations in the hrp48 gene, which encodes a Drosophila heterogeneous nuclear ribonucleoprotein particle protein, cause lethality and developmental defects and affect P-element third-intron splicing in vivo. *Molecular and Cellular Biology*, 17(12):7260–7.

Hatton, R., Subramaniam, V., and Lopez, J. (1998). Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Molecular Cell*, 2(6):787–96.

Haussmann, I. U., Li, M. I. N., and Soller, M. (2011). ELAV mediated 3 -end pro-

cessing of ewg transcripts is evolutionary conserved despite sequence degeneration of the ELAV binding site. *Genetics*, 189:97 – 107.

Haussmann, I. U., White, K., and Soller, M. (2008). Erect wing regulates synaptic growth in Drosophila by integration of multiple signaling pathways. *Genome Biology*, 9(4):R73.

He, L. and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nature reviews. Genetics*, 5(7):522–31.

Hilgers, V., Lemke, S. B., and Levine, M. (2012). ELAV mediates 3'UTR extension in the Drosophila nervous system. *Genes and Development*, 26:2259–2264.

Hilgers, V., Perry, M. W., Hendrix, D., Stark, A., Levine, M., and Haley, B. (2011). Neural-specific elongation of 3' UTRs during Drosophila development. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38):15864–9.

Hinman, M.N; Lou, H. (2008). Diverse molecular functions of Hu proteins. *Cell Mol Life Sci*, 65(20):3168–3181.

Hirokawa, N., Niwa, S., and Tanaka, Y. (2010). Molecular motors in neurons: transport mechanisms and roles in brain function, development, and disease. *Neuron*, 68(4):610–38.

Hnilicová, J., Hozeifi, S., Dušková, E., Icha, J., Tománková, T., and Staněk, D. (2011). Histone deacetylase activity modulates alternative splicing. *PloS One*, 6(2):e16727.

Holt, C. E. and Schuman, E. M. (2013). The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron*, 80(3):648–57.

Huang, Y. and Richter, J. D. (2007). Analysis of mRNA Translation in Cultured Hippocampal Neurons. *Methods in Enzymology*, 431:143–162.

Huber, P. and Zhao, W. (2010). Detection of protein - RNA complexes in Xenopus oocytes. *Methods*, 51:82–86.

Hueber, S. D., Weiller, G. F., Djordjevic, M. a., and Frickey, T. (2010). Improving Hox protein classification across the major model organisms. *PloS One*, 5(5):e10820.

Hughes, C. L. and Kaufman, T. C. (2002). Hox genes and the evolution of the arthropod body plan. *Evolution & Development*, 4(6):459–99.

Huntzinger, E. and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews. Genetics*, 12(2):99–110.

Izquierdo, J. M. (2010). Cell-specific regulation of Fas exon 6 splicing mediated by Hu antigen R. *Biochemical and Biophysical Research Communications*, 402(2):324–8.

Ji, Z., Lee, J. Y., Pan, Z., Jiang, B., and Tian, B. (2009). Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):7028–33.

Jiang, C. and Pugh, B. F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews. Genetics*, 10(3):161–72.

Jimenez, F. and Campos-Ortega, J. A. (1987). Genes in subdivision 1B of the Drosophila melanogaster X-chromosome and their influence on neural development. *Journal of Neurogenetics*, 4:179–200.

Johnstone, O. and Lasko, P. (2001). Translational regulation and RNA localization in Drosophila oocytes and embryos. *Annual Review of Genetics*, 35:365–406.

Jones, C. I. and Newbury, S. F. (2010). Functions of microRNAs in Drosophila development. *Biochemical Society Transactions*, 38(4):1137–43.

Jung, H., Gkogkas, C. G., Sonenberg, N., and Holt, C. E. (2014). Remote control of gene function by local translation. *Cell*, 157(1):26–40.

Jurica, M. S. and Moore, M. J. (2003). Pre-mRNA Splicing : Awash in a Sea of Proteins. *Molecurar Cell*, 12:5–14.

Kalsotra, A. and Cooper, T. a. (2011). Functional consequences of developmentally regulated alternative splicing. *Nature reviews. Genetics*, 12(10):715–29.

Kalsotra, A., Wang, K., Li, P.-F., and Cooper, T. a. (2010). MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes & Development*, 24(7):653–8.

Kandul, N. P. and Noor, M. a. F. (2009). Large introns in relation to alternative splicing and gene evolution: a case study of Drosophila bruno-3. *BMC Genetics*, 10:67.

Kazan, H., Ray, D., Chan, E. T., Hughes, T. R., and Morris, Q. (2010). RNAcontext: a new method for learning the sequence and structure binding preferences of RNA-binding proteins. *PLoS computational biology*, 6(7):e1000832.

Kedde, M. and Agami, R. (2008). Interplay between microRNAs and RNA-binding proteins determines developmental processes. *Cell Cycle*, 7(7):899–903.

Kehle, J. (1998). dMi-2, a Hunchback-Interacting Protein That Functions in Polycomb Repression. *Science*, 282(5395):1897–1900.

Kennison, J. a. and Tamkun, J. W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 85(21):8136–40.

Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews. Genetics*, 11(5):345–55.

Khattak, S., Lee, B. R., Cho, S. H., Ahnn, J., and Spoerel, N. a. (2002). Genetic characterization of Drosophila Mi-2 ATPase. *Gene*, 293(1-2):107–14.

Kim, H. H., Kuwano, Y., Srikantan, S., Lee, E. K., Martindale, J. L., and Gorospe, M. (2009). HuR recruits let-7/RISC to repress c-Myc expression. *Genes & Development*, 23(15):1743–8.

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., and Nogues, G. (2004). Multiple links between transcription and splicing. *RNA (New York, N.Y.)*, 10(10):1489–98.

Kornfeld, K., Saint, R. B., Beachy, P. a., Harte, P. J., Peattie, D. a., and Hogness, D. S. (1989). Structure and expression of a family of Ultrabithorax mRNAs generated by alternative splicing and polyadenylation in Drosophila. *Genes & Development*, 3(2):243–258.

Koushika, S. P., Lisbin, M. J., and White, K. (1996). ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform. *Current Biology : CB*, 6(12):1634–41.

Koushika, S. P., Soller, M., Desimone, S. M., Daub, D. M., and White, K. (1999). Differential and Inefficient Splicing of a Broadly Expressed Drosophila erect wing Transcript Results in Tissue-Specific Enrichment of the Vital EWG Protein Isoform. *Molecular and Cellular Biology*, 19(6):3998–4007.

Koushika, S. P., Soller, M., and White, K. (2000). The neuron-enriched splicing pattern of Drosophila erect wing is dependent on the presence of ELAV protein. *Molecular and Cellular Biology*, 20(5):1836–45.

Krasnow, M. a., Saffman, E. E., Kornfeld, K., and Hogness, D. S. (1989). Transcriptional activation and repression by Ultrabithorax proteins in cultured Drosophila cells. *Cell*, 57(6):1031–43.

Kundu, P., Fabian, M. R., Sonenberg, N., Bhattacharyya, S. N., and Filipowicz, W. (2012). HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Research*, 40(11):5088–5100.

Lai, D., Proctor, J. R., and Meyer, I. M. (2013). On the importance of cotranscriptional RNA structure formation. *RNA*, 19:1461–1473.

Lai, W. S., Carrick, D. M., and Blackshear, P. J. (2005). Influence of nonameric AU-rich tristetraprolin-binding sites on mRNA deadenylation and turnover. *The Journal of Biological Chemistry*, 280(40):34365–77.

Legagneux, V., Bouvet, P., Omilli, F., Chevalier, S., and Osborne, H. B. (1992). Identification of RNA-binding proteins specific to Xenopus Eg maternal mRNAs:

association with the portion of Eg2 mRNA that promotes deadenylation in embryos. *Development (Cambridge, England)*, 116(4):1193–202.

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature*, 276:565–570.

Lewis, E. B. (1998). The bithorax complex: the first fifty years. *The International Journal of Developmental Biology*, 42(3):403–15.

Li, J. and Lu, X. (2013). The emerging roles of 3' untranslated regions in cancer. *Cancer letters*, 337(1):22–5.

Li, X., Quon, G., Lipshitz, H. D., and Morris, Q. (2010). Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure. *RNA*, 16(6):1096–1107.

Licatalosi, D. D. and Darnell, R. B. (2010). RNA processing and its regulation: global insights into biological networks. *Nature Reviews. Genetics*, 11(1):75–87.

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., Clark, T. a., Schweitzer, A. C., Blume, J. E., Wang, X., Darnell, J. C., and Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, 456(7221):464–9.

Lisbin, M. J., Qiu, J., and White, K. (2001). The neuron-specific RNA-binding protein ELAV regulates neuroglian alternative splicing in neurons and binds directly to its pre-mRNA. *Genes & Development*, 15(19):2546–61.

Lopez, J. and Hogness, D. S. (1991). Immunochemical dissection of the Ultrabithorax homeoprotein family in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 88(22):9924–8.

López de Silanes, I., Zhan, M., Lal, A., Yang, X., and Gorospe, M. (2004). Identification of a target RNA motif for RNA-binding protein HuR. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2987–92.

Loureiro, J. and Peifer, M. (1998). Roles of Armadillo, a Drosophila catenin, during central nervous system development. *Current Biology*, 8(11):622–32.

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, 144(1):16–26.

Lutz, C. S. (2008). Alternative Polyadenylation: A Twist on mRNA 3' End Formation. *ACS Chemical Biology*, 3(10):609–617.

Maeda, R. K. and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. *Development (Cambridge, England)*, 133(8):1413–22.

Maniatis, T. and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature*, 416:499–506.

Mann, R. S. and Hogness, D. S. (1990). Functional Dissection of Ultrabithorax Proteins in D. melanogaster. *Cell*, 60:597–610.

Marillat, V., Sabatier, C., Failli, V., Matsunaga, E., Sotelo, C., Tessier-lavigne, M., and Chedotal, A. (2004). The Slit Receptor Rig-1 / Robo3 Controls Midline Crossing by Hindbrain Precerebellar Neurons and Axons. *Neuron*, 43:69–79.

Martin, K. C. and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell*, 136(4):719–30.

Matlin, A. J., Clark, F., and Smith, C. W. J. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology*, 6(5):386–98.

Matunis, M. J., Matunis, E. L., and Dreyfuss, G. (1992). Isolation of hnRNP complexes from Drosophila melanogaster. *The Journal of Cell Biology*, 116(2):245–55.

Mauger, D. M., Siegfried, N. a., and Weeks, K. M. (2013). The genetic code as expressed through relationships between mRNA structure and protein function. *FEBS letters*, 587(8):1180–8.

Mayr, C. and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138(4):673–84.

McGinnis, W. and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell*, 68(2):283–302.

Meisner, N.-C., Hackermüller, J., Uhl, V., Aszódi, A., Jaritz, M., and Auer, M. (2004). mRNA openers and closers: modulating AU-rich element-controlled mRNA stability by a molecular switch in mRNA secondary structure. *Chembiochem : a European Journal of Chemical Biology*, 5(10):1432–47.

Meng, Z., King, P. H., Nabors, L. B., Jackson, N. L., Chen, C.-Y., Emanuel, P. D., and Blume, S. W. (2005). The ELAV RNA-stability factor HuR binds the 5'-untranslated region of the human IGF-IR transcript and differentially represses cap-dependent and IRES-mediated translation. *Nucleic Acids Research*, 33(9):2962–79.

Merritt, C., Rasoloson, D., Ko, D., and Seydoux, G. (2008). 3' UTRs are the primary regulators of gene expression in the C. elegans germline. *Current Biology : CB*, 18(19):1476–82.

Millevoi, S. (2009). Molecular mechanisms of eukaryotic pre-mRNA 3 0 end processing regulation. *Nucleic Acids Research*, 4242(4431):1–18.

Millevoi, S. and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, 38(9):2757–74.

Mitchelson, a., Simonelig, M., Williams, C., and O'Hare, K. (1993). Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppression in Drosophila melanogaster by suppressor of forked occurs at the level of RNA stability. *Genes & Development*, 7(2):241–249.

Mueller, A. a., Cheung, T. H., and Rando, T. a. (2013). All's well that ends well: alternative polyadenylation and its implications for stem cell biology. *Current Opinion in Cell Biology*, pages 1–11.

Mujtaba, S., Zeng, L., and Zhou, M.-M. (2007). Structure and acetyl-lysine recognition of the bromodomain. *Oncogene*, 26(37):5521–7.

Mukherjee, N., Corcoran, D. L., Nusbaum, J. D., Reid, D. W., Georgiev, S., Hafner, M., Ascano, M., Tuschl, T., Ohler, U., and Keene, J. D. (2011). Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Molecular Cell*, 43(3):327–39.

Nilsen, T. W. and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280):457–63.

Norris, A. D. and Calarco, J. a. (2012). Emerging Roles of Alternative Pre-mRNA Splicing Regulation in Neuronal Development and Function. *Frontiers in Neuroscience*, 6(August):122.

O'Connor, M. B., Binari, R., Perkins, L. a., and Bender, W. (1988). Alternative RNA products from the Ultrabithorax domain of the bithorax complex. *The EMBO Journal*, 7(2):435–45.

Orphanides, G. and Reinberg, D. (2002). A Unified Theory of Gene Expression. *Cell*, 108:439–451.

Ortega, A., Niksic, M., Bachi, A., Wilm, M., Sánchez, L., Hastie, N., and Valcárcel, J. (2003). Biochemical function of female-lethal (2)D/Wilms' tumor suppressor-1-associated proteins in alternative pre-mRNA splicing. *The Journal of Biological Chemistry*, 278(5):3040–7.

Parton, R. M., Davidson, A., Davis, I., and Weil, T. T. (2014). Subcellular mRNA localisation at a glance. *Journal of cell science*, 127(Pt 10):2127–33.

Pascale, A. and Govoni, S. (2012). The complex world of post-transcriptional mechanisms: is their deregulation a common link for diseases? Focus on ELAV-like RNA-binding proteins. *Cellular and Molecular Life Sciences : CMLS*, 69(4):501–17.

Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax - Extradenticle homeodomain complex. *Nature*, 397:714–719.

Patraquim, P., Warnefors, M., and Alonso, C. R. (2011). Evolution of Hox post-transcriptional regulation by alternative polyadenylation and microRNA modulation within 12 Drosophila genomes. *Molecular Biology and Evolution*, 28(9):2453–60.

Penalva, L. O. F., Ruiz, M. F., Ortega, A., Vicente, L., Sa, L., and Segarra, C. (2000). The Drosophila fl(2)d Gene, Required for Female-Specific Splicing of

Sxl and tratra Pre-mRNAs , Encodes a Novel Nuclear Protein With a HQ-Rich Domain n Valca. *Genetics*, 155:129–139.

Perrimon, N., Engstrom, L. E. E., and Mahowald, A. P. (1984). Developmental genetics of the 2E-F region of the Drosophila X chromosome: a region rich in "Developmentally important"genes. *Genetics*, 108:559–572.

Persson, M., Andrén, Y., Mark, J., Horlings, H. M., Persson, F., and Stenman, G. (2009). Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proceedings of the National Academy of Sciences of the United States of America*, 106(44):18740–4.

Phatnani, H. P. and Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes & Development*, 20(21):2922–36.

Pinto, P. a. B., Henriques, T., Freitas, M. O., Martins, T., Domingues, R. G., Wyrzykowska, P. S., Coelho, P. a., Carmo, A. M., Sunkel, C. E., Proudfoot, N. J., and Moreira, A. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. *The EMBO Journal*, 30(12):2431–44.

Proudfoot, N. J. (2011). Ending the message: poly(A) signals then and now. *Genes & Development*, 25(17):1770–82.

Raisin-Tani, S. and Léopold, P. (2002). Drosophila crooked-neck protein co-fractionates in a multiprotein complex with splicing factors. *Biochemical and Biophysical Research Communications*, 296(2):288–92.

Ratti, A., Fallini, C., Cova, L., Fantozzi, R., Calzarossa, C., Zennaro, E., Pascale, A., Quattrone, A., and Silani, V. (2006). A role for the ELAV RNA-binding proteins in neural stem cells: stabilization of Msi1 mRNA. *Journal of Cell Science*, 119:1442–52.

Reed, H. C., Hoare, T., Thomsen, S., Weaver, T. a., White, R. a. H., Akam, M., and Alonso, C. R. (2010). Alternative splicing modulates Ubx protein function in Drosophila melanogaster. *Genetics*, 184(3):745–58.

Rivas-Aravena, A., Ramdohr, P., Vallejos, M., Valiente-Echeverría, F., Dormoy-Raclet, V., Rodríguez, F., Pino, K., Holzmann, C., Huidobro-Toro, J. P., Gallouzi,

I.-E., and López-Lastra, M. (2009). The Elav-like protein HuR exerts translational control of viral internal ribosome entry sites. *Virology*, 392(2):178–85.

Roca, X. and Karginov, F. (2012). RNA biology in a test tube–an overview of in vitro systems/assays. *Wiley Interdiscip Rev RNA*, 4:509–27.

Rodgers, N. D., Jiao, X., and Kiledjian, M. (2002). Identifying mRNAs bound by RNA-binding proteins using affinity purification and differential display. *Methods*, 26:115–122.

Rogulja-Ortmann, A., Picao-Osorio, J., Villava, C., Patraquim, P., Lafuente, E., Aspden, J., Thomsen, S., Technau, G. M., and Alonso, C. R. (2014). The RNA-binding protein ELAV regulates Hox RNA processing, expression and function within the Drosophila nervous system. *Development (Cambridge, England)*, 141(10):2046–56.

Rogulja-Ortmann, A., Renner, S., and Technau, G. M. (2008). Antagonistic roles for Ultrabithorax and Antennapedia in regulating segment-specific apoptosis of differentiated motoneurons in the Drosophila embryonic central nervous system. *Development (Cambridge, England)*, 135(20):3435–45.

Rogulja-ortmann, A. and Technau, G. M. (2008). Multiple roles for Hox genes in segment-specific shaping of CNS lineages. *Fly*, 2(6):316–319.

Ronshaugen, M., Biemar, F., Piel, J., Levine, M., and Lai, E. C. (2005). The Drosophila microRNA iab-4 causes a dominant homeotic transformation of halteres to wings. *Genes & Development*, 19(24):2947–52.

Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2013). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo . *Nature*, advance on.

Sahin, E. and Roberts, C. J. (2012). Size-Exclusion Chromatography with Multi-angle Light Scattering for Elucidating Protein Aggregation Mechanisms. *Methods in Molecular Biology*, 899:403–423.

Sakai, K., Gofuku, M., Kitagawa, Y., Ogasawara, T., Hirose, G., Yamazaki, M., Koh, C.-s., Yanagisawa, N., and Steinman, L. (1994). A hippocampal protein associated with paraneoplastic neurologic syndrome and small cell lung carcinoma.pdf. *Biochemical and Biophysical Research Communications*, 199(3):1200–1208.

Salecker, I. (2006). HOW to Wrap Axons with croocked neck. *Neuron*, 52:933–935.

Salomonis, N., Schlieve, C. R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A. C., Vranizan, K., Spindler, M. J., Pico, A. R., Cline, M. S., Clark, T. a., Williams, A., Blume, J. E., Samal, E., Mercola, M., Merrill, B. J., and Conklin, B. R. (2010). Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(23):10514–9.

Samson, M.-L. (2008). Rapid functional diversification in the structurally conserved ELAV family of neuronal RNA binding proteins. *BMC Genomics*, 9:392.

Samson, M.-L. and Chalvet, F. (2003). found in neurons, a third member of the Drosophila elav gene family, encodes a neuronal protein and interacts with elav. *Mechanisms of Development*, 120(3):373–383.

Seelig, H. P., Moosbrugger, I., Ehrfeld, H., Fink, T., Renz, M., and Genth, E. (1995). The major dermatomyositis-specific mi-2 autoantigen is a presumed helicase involved in transcriptional activation. *Arthritis & Rheumatism*, 38(10):1389–1399.

Seshaiah, P., Miller, B., Myat, M. M., and Andrew, D. J. (2001). pasilla, the Drosophila homologue of the human Nova-1 and Nova-2 proteins, is required for normal secretion in the salivary gland. *Developmental Biology*, 239(2):309–22.

Shen, H., Kan, J. L. C., Ghigna, C., Biamonti, G., and Green, M. R. (2004). A single polypyrimidine tract binding protein ( PTB ) binding site mediates splicing inhibition at mouse IgM exons M1 and M2. *RNA*, 10:787–794.

Shepard, P. J. and Hertel, K. J. (2009). Protein family review The SR protein family. *Genome Biology*, 10(242):1–9.

Shukla, S. and Oberdoerffer, S. (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochimica et Biophysica Acta*, 1819(7):673–83.

Sibley, C. R., Attig, J., and Ule, J. (2012). The greatest catch: big game fishing for mRNA-bound proteins. *Genome Biology*, 13(7):163.

Siebel, C. W., Kanaar, R., and Rio, D. C. (1994). Regulation of tissue-specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes & Development*, 8(14):1713–1725.

Simionato, E., Barrios, N., Duloquin, L., Boissonneau, E., Lecorre, P., and Agnès, F. (2007). The Drosophila RNA-binding protein ELAV is required for commissural axon midline crossing via control of commissureless mRNA expression in neurons. *Developmental Biology*, 301(1):166–77.

Simone, L. E. and Keene, J. D. (2013). Mechanisms coordinating ELAV/Hu mRNA regulons. *Current opinion in genetics & development*, 23(1):35–43.

Singh, G., Ricci, E. P., and Moore, M. J. (2013). RIPiT-Seq: A high-throughput approach for footprinting RNA:protein complexes. *Methods*, ahead of p.

Smibert, P., Miura, P., Westholm, J. O., Shenker, S., May, G., Duff, M. O., Zhang, D., Eads, B. D., Carlson, J., Brown, J. B., Eisman, R. C., Andrews, J., Kaufman, T., Cherbas, P., Celniker, S. E., Graveley, B. R., and Lai, E. C. (2012). Global patterns of tissue-specific alternative polyadenylation in Drosophila. *Cell reports*, 1(3):277–89.

Smith, C. W. and Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in Biochemical Sciences*, 25(8):381–8.

Soller, M., Li, M., and Haussmann, I. U. (2008). Regulation of the ELAV target ewg: insights from an evolutionary perspective. *Biochemical Society Transactions*, 36(Pt 3):502–4.

Soller, M., Li, M., and Haussmann, I. U. (2010). Determinants of ELAV gene-specific regulation. *Biochemical Society transactions*, 38(4):1122–4.

Soller, M. and White, K. (2003). ELAV inhibits 3'-end processing to promote neural splicing of ewg pre-mRNA. *Genes & Development*, 17(20):2526–38.

Soller, M. and White, K. (2005). ELAV Multimerizes on Conserved AU 4 6 Motifs Important for ewg Splicing Regulation. *Molecular and Cellular Biology*, 25(17):7580–7591.

Spasic, M., Friedel, C. C., Schott, J., Kreth, J., Leppek, K., Hofmann, S., Ozgur, S., and Stoecklin, G. (2012). Genome-wide assessment of AU-rich elements by the AREScore algorithm. *PLoS Genetics*, 8(1):e1002433.

Spector, D. L. and Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harbor Perspectives in Biology*, 3(2):1–12.

St Johnston, D. (2005). Moving messages: the intracellular localization of mRNAs. *Nature Reviews. Molecular Cell Biology*, 6(5):363–75.

Standart, N. (1993). The RNA-protein partners in mRNP. *Molecular Biology Reports*, 18(2):135–42.

Steinmetz, E. J. (1997). Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell*, 89(4):491–4.

Subramaniam, V., Bomze, H. M., and Lopez, A. J. (1994). Functional Differences Between Ultrabithorax Protein Isoforms. *Genetics*, 136:979–991.

Sugimoto, Y., König, J., Hussain, S., Zupan, B., Curk, T., Frye, M., and Ule, J. (2012). Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biology*, 13(8):R67.

Szostak, E. and Gebauer, F. (2013). Translational control by 3'-UTR-binding proteins. *Briefings in Functional Genomics*, 12(1):58–65.

Takagaki, Y., Seipelt, R. L., Peterson, M. L., and Manley, J. L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell*, 87(5):941–52.

Takagaki, Yoshio ; Manley, J. (1994). A polyadenylation factor subunit is the human homologue of the Drosophila suppresor of forked protein. *Nature*, 372:471–474.

Taliaferro, J. M., Alvarez, N., Green, R. E., Blanchette, M., and Rio, D. C. (2011). Evolution of a tissue-specific splicing network. *Genes & Development*, 25(6):608–20.

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, a. M., Kaufman, T. C., and Kennison, J. a. (1992). brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell*, 68(3):561–72.

Thomsen, S., Azzam, G., Kaschula, R., Williams, L. S., and Alonso, C. R. (2010). Developmental RNA processing of 3'UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to microRNAs. *Development*, 137(17):2951–60.

Toba, G. and White, K. (2008). The third RNA recognition motif of Drosophila ELAV protein has a role in multimerization. *Nucleic Acids Research*, 36(4):1390–9.

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E., and Rubin, G. M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biology*, 8(7):R145.

Tyler, D. M., Okamura, K., Chung, W.-J., Hagen, J. W., Berezikov, E., Hannon, G. J., and Lai, E. C. (2008). Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. *Genes & Development*, 22(1):26–36.

Waldholm, J., Wang, Z., Brodin, D., Tyagi, A., Yu, S., Theopold, U., Farrants, A. K. O., and Visa, N. (2011). SWI/SNF regulates the alternative processing of a specific subset of pre-mRNAs in Drosophila melanogaster. *BMC Molecular Biology*, 12(1):46.

Walker, J., de Melo Neto, O., and Standart, N. (1998). Gel retardation and UV-crosslinking assays to detect specific RNA-protein interactions in the 5' or 3' UTRs of translationally regulated mRNAs. *Methods in Molecular Biology (Clifton, N.J.)*, 77:365–78.

Wang, Q., Hobbs, K., Lynn, B., and Rymond, B. C. (2003). The Clf1p splicing factor promotes spliceosome assembly through N-terminal tetratricopeptide repeat contacts. *The Journal of Biological Chemistry*, 278(10):7875–83.

Wang, X. and Hall, T. M. T. (2001a). Structural basis for recognition of AU-rich element RNA by the HuD. *Nature Structural Biology*, 8(2):141–145.

Wang, X. and Hall, T. M. T. (2001b). Structural basis for recognition of AU-rich element RNA by the HuD. *Nature Structural Biology*, 8(2):141–145.

Wang, Z. and Burge, C. B. (2008). Splicing regulation : From a parts list of regulatory elements to an integrated splicing code. *RNA*, 14:802–813.

White, R. A. H. and Wilcox, M. (1985). Distribution of Ultrabithorax proteins in Drosophila. *The EMBO Journal*, 4(8):2035–2043.

Worringer, K. and Panning, B. (2007). Zinc finger protein Zn72D promotes productive splicing of the maleless transcript. *Molecular and Cellular Biology*, 27(24):8760–9.

Worringer, K. a., Chu, F., and Panning, B. (2009). The zinc finger protein Zn72D and DEAD box helicase Belle interact and control maleless mRNA and protein levels. *BMC Molecular Biology*, 10:33.

Yao, K.-m. and Samson, M.-l. (1992). A Prototype for Neuronal-Specific RNA Binding Protein Gene Family That I s Conserved in Flies and Humans. *Journal of Neurobiology*, 24(6):723–739.

Yao, K. M. and White, K. (1991). Organizational analysis of elav gene and functional analysis of ELAV protein of Drosophila melanogaster and Drosophila virilis. *Molecular and Cellular Biology*, 11(6):2994–3000.

Zhang, K., Smouse, D., and Perrimon, N. (1991). The crooked neck gene of Drosophila contains a motif found in a family of yeast cell cycle genes. *Genes & Development*, 5(6):1080–1091.

Zhang, Y., Li, S., Yuan, L., Tian, Y., Weidenfeld, J., Yang, J., Liu, F., Chokas, A. L., and Morrisey, E. E. (2010). Foxp1 coordinates cardiomyocyte prolifera-

tion through both cell-autonomous and nonautonomous mechanisms. *Genes & Development*, 24(16):1746–57.

Zhu, H., Zhou, H.-L., Hasman, R. a., and Lou, H. (2007). Hu proteins regulate polyadenylation by blocking sites containing U-rich sequences. *The Journal of Biological Chemistry*, 282(4):2203–10.

Zhuang, R., Rao, J. N., Zou, T., Liu, L., Xiao, L., Cao, S., Hansraj, N. Z., Gorospe, M., and Wang, J.-Y. (2013). miR-195 competes with HuR to modulate stim1 mRNA stability and regulate cell migration. *Nucleic Acids Research*, 7000:1–15.

Ziegeler, G., Ming, J., Koseki, J. C., Sevinc, S., Chen, T., Ergun, S., Qin, X., and Aktas, B. H. (2010). Embryonic lethal abnormal vision-like HuR-dependent mRNA stability regulates post-transcriptional expression of cyclin-dependent kinase inhibitor p27Kip1. *The Journal of Biological Chemistry*, 285(20):15408–19.