Separate functions of BTZ during post-transcriptional gene regulation

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Simona Ciriello

aus Neapel, Italien

Köln, 01.10.14

Berichterstatter :   PD Dr. Niels H. Gehring

Prof. Dr. Jürgen R. Dohmen

Tag der mündlichen Prüfung :   28.11.14

Table of Contents

List of Abbreviations

Aa	Amino acid
Bp	Base pair
EJC	Exon Junction Complex
EMSA	Electrophoretic mobility shift assay
IP	Immunoprecipitation
KD	knockdown
mRNA	Messenger RNA
mRNP	Messenger ribonuceoprotein
NES	Nuclear export signal
NLS	Nuclear localization signal
NMD	Nonsense-mediated mRNA decay
Nt	Nucleotide
o/n	Over night
ORF	Open reading frame
PCR	Polymerase chair reaction
Pol II	RNA Polymerase II
PTC	Premature termination codon
RNP	Ribonucleoprotein
RRM	RNA recognition motif
RT	Room temperature
snRNP	Small nuclear ribonucleic particles
UTR	Untranslated region

1. Abstract

In metazoans, the exon junction complex (EJC) is a central component of spliced messenger ribonucleoprotein particles (mRNPs). EJCs are assembled by the spliceosome and deposited upstream of exon-exon boundaries in the nucleus. The heterotetrameric core of the EJC is composed of the proteins eIF4A3 (DDX48), MAGOH, RBM8 (Y14) and CASC3/MLN51/Barentsz (BTZ). EJCs contribute to different steps of post-transcriptional gene expression including splicing, translation and nonsense-mediated mRNA decay (NMD). BTZ is an important functional component and is involved in the stimulation of translation and nonsense-mediated mRNA decay. Here, I show that both the N-terminal and the SELOR domain of BTZ elicit NMD in a tethering assay. They activate NMD following two different pathways, BTZ-dependent and UPF2 dependent, which get reunited once UPF1 is activated. In contrast, the C-terminal region of BTZ does not seem to be involved in NMD. Instead, this region plays a role in a different process that leads to the polyadenylation of a reporter mRNA at an upstream, non-canonical polyadenylation site. Moreover, I show that binding of the SELOR domain to mRNA in vivo is EJC-dependent. In addition the SELOR domain *in vivo* interacts with several SR proteins for a subset of which an NMD-activating function is observed. These findings uncover novel EJC-dependent and -independent functions of BTZ during post-transcriptional gene expression regulation.

1.1 Deutsche Zusammenfassung

In Metazoen stellt der Exon-Verbindungs Komplex (`exon junction complex´, EJC) eine zentrale Komponente von gespleißten Ribonukleoproteinpartikeln (mRNPs) dar. EJCs werden mithilfe von Spleißosomen im Zellkern zusammengesetzt und nahe Exon-Exon Grenzen auf der mRNA platziert. Der heterotetramere Kern des EJC besteht aus den Proteinen eIF4A3 (DDX48), MAGOH, RBM8 (Y14) und CASC3/MLN51/Barentz (BTZ). EJCs sind in verschiedenste Schritte der posttranskriptionalen Genexpression involviert, unter anderem das Spleißen, die Translation und den nonsense-vermittelten mRNA-Abbau (NMD). BTZ ist eine wichtige funktionale Komponente innerhalb dieser Prozesse und an der Stimulation von sowohl Translation als auch NMD beteiligt. In dieser Arbeit wird gezeigt, dass die N-terminale sowie die SELOR (`Speckle Localizer and RNA-binding´) Domäne von BTZ, NMD im Rahmen eines Tethering-Assays induzieren können und dabei zwei verschiedene Wege der NMD-Aktivierung genutzt werden. Darüber hinaus wird dargestellt, dass die C-terminale Domäne eine nicht-kanonische Polyadenylierungsstelle innerhalb der 3′ UTR einer Reporter-mRNA aktiviert, was darauf hinweist, dass BTZ an alternativer Polyadenylierung beteiligt ist. Zudem wird präsentiert, dass die SELOR Domäne *in virto* präferenziell bestimmte mRNAs bindet und *in vivo* die mRNA- Interaktion von SELOR EJC-abhängig ist. die Zusätzlich konnten verschieden SR-Proteine in SELOR-assoziierten Proteinkomplexen identifiziert werden, wobei für einige dieser SR Proteine eine NMD-aktivierende Funktion nachgewiesen werden konnte. Diese Ergebnisse zeigen eine neue, EJC-abhängige sowie –unabhängige, regulatorische Funktion von BTZ innerhalb der posttranskriptionalen Geneexpression.

2. Introduction

2.1 mRNA life: from transcription to translation

Eukaryotic gene expression is a multistep process that requires the coordination of several events that are spatially and temporally separated (Moore, 2005; Moore and Proudfoot, 2009; Reed, 2003). All these steps are needed in order to correctly decode the information of the DNA into a functional product (Figure 1). The following sections describe these processes with a special emphasis on the steps most important for this work.

Transcription

Gene expression begins with transcription of DNA into messenger pre-RNA (pre-mRNA) by the RNA Polymerase II (Pol II) (Kornberg, 1999; Sims et al., 2004) (Figure 1). The different steps of transcription (initiation, elongation and termination) are marked by rearrangement of the components that binds Pol II (Cheung and Cramer, 2012). Transcription initiation starts upon addition of two initiating nucleoside triphosphates (NTPs) complementary to the DNA sequence and the formation of the first phosphodiester bond. When the pre-mRNA is 23-nt-long the rearrangement of the transcriptional factors around Pol II marks the beginning of the elongation phase (Pal and Luse, 2002; Roberts et al., 2004). Polyadenylation (see below) and specific termination factors mediate the termination of transcription (Kireeva et al., 2000; Lykke-Andersen and Jensen, 2007; Richard and Manley, 2009).

5'capping

The 5' end capping is the first modification that nascent pre-mRNA undergoes, when it is 22-25 nt long and emerging from the RNA exit channel of Pol II (Shatkin, 1976; Shatkin and Manley, 2000). The 5' cap protects mRNA against exonucleases and promotes transcription, polyadenylation, splicing and nuclear export (Gu and Lima, 2005; Lewis and Izaurralde, 1997).
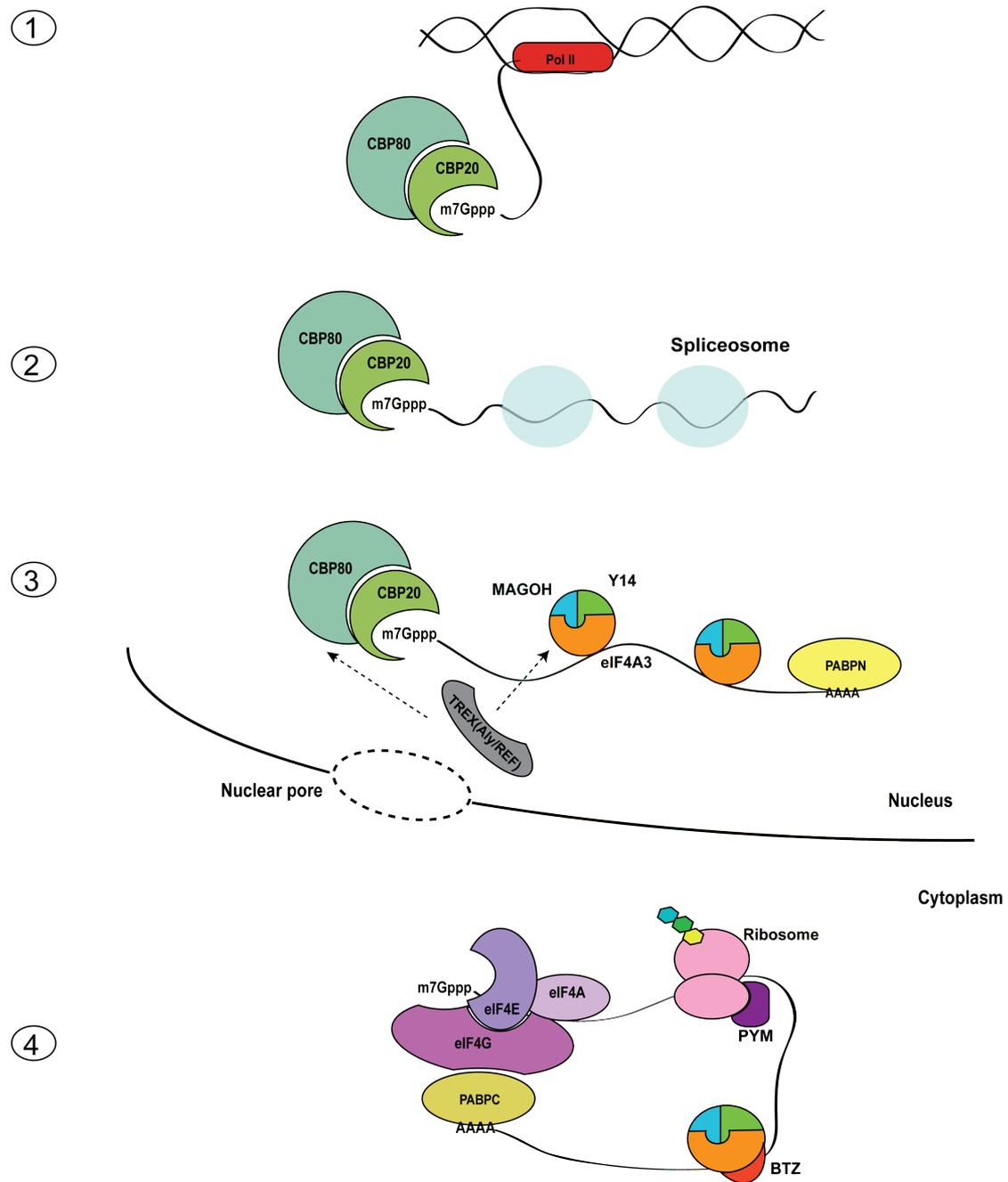
Figure 1. Steps of gene expression. 1) Pol II transcribes the information of DNA in pre-mRNA. 5' cap (m7Gppp) is added when the pre-mRNA is 22-25 nt long and is subsequently bound by the nuclear Cap Binding Complex (nCBC: CBP20 and CBP80). 2) Splicing occurs cotranscriptionally and removes the introns. 3) As a consequence of splicing, the pre-EJC (MAGOH, Y14 and eIF4A3) is deposited on the mRNA. At the end of transcription, the mRNA is polyadenylated, and the poly(A) tail is bound by the nuclear Poly(A) binding protein (PABPN). Finally, the export complex TREX addresses the mRNA to the nuclear pore. 4) In the cytoplasm, BTZ joins the EJC complex, the eIF4F complex (eIF4E, eIF4A and eIF4G) replaces the nCBC, the cytoplasmic PABP (PABPC) replaces PABPN, the ribosome translates the mRNA into protein and PYM displaces the EJC upon translation.

mRNA capping requires three reactions: first, the 5′-γ-phosphate group of the first transcribed nucleotide of pre-mRNA is removed by the RNA triphosphatase; second, a guanine monophosphate (GMP) nucleotide is transferred by a guanylyltransferase to the RNA 5′-diphosphate end; third, a RNA methyl transferase adds a methyl group, yielding the 7-methylguanosine cap (m7GpppN) (Mao et al., 1995; Yue et al., 1997). The 5′ cap is bound to the nuclear cap binding complex (nCBC) in the nucleus, consisting of the cap-binding subunit CBP20 and the auxiliary protein CBP80 that stabilizes the interaction of CBP20 with the cap (Mazza et al., 2001; Mazza et al., 2002). After the mRNA is exported to the cytoplasm, eIF4E (eukaryotic translation initiation factor 4E), a component of the eIF4F complex, binds the cap and promotes the recruitment of the small ribosomal subunit for translation initiation. The other two components of the complex are the DEAD-box helicase eIF4A, responsible for RNA unwinding (Feoktistova et al., 2013) and eIF4G. eIF4G binds eIF4E and PABPC via its C- and N-terminal region, respectively (see polyadenylation). These interactions enhance eIF4F binding to the cap (Kahvejian et al., 2005b), as well as enabling the formation of the so-called closed-loop which facilitates re-initiation of mRNA translation (Hinnebusch and Lorsch, 2012; Jackson et al., 2010) (Figure 1).

Splicing

The second modification that most pre-mRNAs undergo is the excision of the intervening sequences (introns). Specific elements inside the gene sequence mark the position of an intron (Clancy, 2008; Konarska et al., 1985) (Figure 2):

- donor site, at the 5′ end of the intron, it contains a GU sequence, surrounded by a less conserved region;
- branch site, 20-50 nt upstream of the acceptor site, it includes an A;
- acceptor site, at the 3′end of the intron, it contains an AG sequence and is preceded by a polypyrimidine tract.

The spliceosome is a complex consisting of several small nuclear ribonucleoproteins (snRNP), which catalyzes nuclear pre-mRNA splicing (Wahl et al., 2009). The spliceosome assembly is spatially-temporally organized and proceeds through the assembly of the complexes E, A, B, B^act , B* and C (Bessonov

et al., 2010). Complex E is comprised of U1 snRNP binding the donor site, U2AF the polyprimidine tract and SF1 the branch point sequence (BPS). Subsequently, U2 snRNP is recruited to the BPS, forming complex A. U4/U6 and U5 are then recruited as pre-assembled tri-snRNP, leading to the formation of complex B.

Figure 2. Splicing reaction. Schematic representation of an mRNA, with indicated donor site, branch site, polypyrimidine tract and acceptor site. The steps required for the splicing reaction are indicated with the name of the complex formed (E, A, B, B^{act}, B* and C).

This complex is still inactive and requires a conformational/compositional change to be functional: U1 and U4 are released and the complex becomes splicing competent (complex B^{act}). After a structural rearrangement into complex B*, the splicing reaction takes place in complex C. The first reaction consists of cleaving of the pre-mRNA at the 5' end of the intron and lariat formation. During the second reaction the exons are ligated, the intron released and the

spliceosome disassembled with the snRNP being recycled (Bessonov et al., 2010) (Figure 2).

Polyadenylation

Similarly to 5′ capping, 3′ polyadenylation is required for maturation of pre-mRNA and is necessary for nuclear export, stability of the mature transcript and efficient translation (Sachs, 1990). The vast majority of eukaryotic mRNAs contain a long stretch of untemplated adenosines in their 3′ end termed the poly(A) tail. The 3′ end polyadenylation is a two-steps process that involves endonucleolytic cleavage of the transcript and the addition of a poly(A) tail. In human cells, the average length of the poly(A) tail varies between 250 and 300 adenines, while the length in yeast fluctuates between 70 and 80 adenines. In general, the number of adenines is restricted and varies between species (Elkon et al., 2013). mRNAs with a shortened poly(A) tail are normally degraded or stored in a translationally dormant state (D'Ambrogio et al., 2013; Guhaniyogi and Brewer, 2001). Polyadenylation requires several *cis-* and *trans*–acting elements. The Poly(A) signal (PAS) dictates the selection of the cleavage site and is commonly located 15-20 nt upstream of it (Proudfoot and Brownlee, 1976). In $\sim$ 70% of human mRNAs, the PAS consists of the nucleotide sequences AAUAAA or AUUAAA. The remaining $\sim$ 30% of RNAs contain other sequences, such as UAUAAA, AACAAA or ACUAAA (MacDonald and Redondo, 2002), suggesting the possibility of alternative polyadenylation (APA). Alternative polyadenylation can generate transcript isoforms with alternative 3′ ends (Elkon et al., 2013). In addition to the PAS, U- or GU-rich downstream sequence elements (DSEs) and less well-defined upstream sequence elements (USEs) enhance cleavage efficiency. The cleavage and polyadenylation specificity factor (CPSF) recognizes the PAS, the cleavage stimulating factor (CSTF) binds the DSEs (Mandel et al., 2008; Proudfoot, 2011), and cleavage factors Im (CFIm) and IIm (CFIIm) bind the USEs (Brown and Gilmartin, 2003; Yang et al., 2011). The interaction of PABP (in humans primarily PABPC1) with the cap-associated eIF4G in the cytoplasm is thought to facilitate circularization of mRNAs, thereby supporting efficient translation termination, ribosome recycling and translation initiation (Amrani et al., 2004; Kahvejian et al., 2005a; Wells et al., 1998) (Figure 1).

Export

After the completion of all processing steps (capping, splicing and polyadenylation), the mature mRNA needs to be exported to the cytoplasm in order to be translated into protein. mRNA export is already initiated during splicing of the first intron when the transcription export (TREX) complex is deposited at the 5′ end of the mRNA (Cheng et al., 2006; Masuda et al., 2005). TREX is a multiprotein complex that is composed of the THO complex, the RNA helicase UAP56 and the adaptor molecule Aly/REF (Zhou et al., 2000). The mechanism of recruitment of TREX to the mRNA is still unclear. Different evidence suggest an EJC-dependent (Stutz et al., 2000; Zhou et al., 2000) or a cap-dependent (Cheng et al., 2006; Luna et al., 2012) recruitment. Once recruited, Aly/REF interacts with the heterodimeric mRNA export receptor TAP/p15 (also known as NXF1/NXT1) (Viphakone et al., 2012), which binds to mRNA and translocates it across the nuclear pore to the cytoplasm (Hurt et al., 2000; Segref et al., 1997) (Figure 1).

Translation

After export to the cytoplasm, the information of the mRNA is translated into protein. This process can be divided into different steps, which are initiation, elongation, termination and recycling. Translation initiation is carried out by a network of factors and involves eukaryotic initiation factors (eIFs), the ribosomal subunits and the mRNA (Aitken and Lorsch, 2012). These factors scan the mRNA until an AUG start codon is encountered (Aitken and Lorsch, 2012; Jackson et al., 2010). Specifically, eIF3 in the initiation phase interacts with eIF4F and contributes to the recognition of AUG (Hinnebusch, 2006). Elongation of the amino acid chain proceeds until a termination codon is reached (UAG, UAA, UGA). The eukaryotic release factors (RFs) are responsible for the release of the completed protein product (Jackson et al., 2012). At this step the ribosome is still assembled on the mRNA. The recycling process is still unclear, but what is known is that the ribosome subunits are separated from the mRNA. This enables the ribosomal subunits to be recruited again to the AUG of mRNA to start another round of translation (Nurenberg and Tampe, 2013).

2.2 The Exon Junction Complex

Messenger ribonuceoproteins (mRNPs) are dynamic complexes, changing their composition during the lifetime of an mRNA. As an example, the nuclear cap binding protein (CBP20/80) binds mRNA in the nucleus (Lewis and Izaurralde, 1997; Sonenberg and Hinnebusch, 2009), while the eukaryotic translation initiation complex 4F replaces CBP20/80 in the cytoplasm (Jackson et al., 2010; Sonenberg and Hinnebusch, 2009). Similarly, the nuclear poly(A) binding protein PABPN (Krause et al., 1994; Wahle, 1991) is replaced by PABPC in the cytoplasm (Gorlach et al., 1994). The exon junction complex (EJC) differs from the above-mentioned mRNPs. It remains bound to the mRNA in the cytoplasm after being deposited in the nucleus during splicing (Le Hir et al., 2000a; Le Hir et al., 2000b).

2.2.1 EJC structure

The core of the EJC consists of four proteins known as MAGOH, Y14 (RMB8A), eIF4A3 (DDX48) and Barentsz (BTZ, also known as CASC3 and MLN51) (Gehring et al., 2009a) (Figure 3). In living cells, the EJC functions as a binding platform for many other proteins responsible for different processes during the mRNA life cycle (see paragraphs 2.2.3 and 2.2.4). eIF4A3 is a DEAD-box helicase protein. The proteins of the DExD/H family share a conserved helicase core, consisting of two globular RecA-like domains, the N-terminal RecA domain 1 and the C-terminal RecA domain 2 connected by a flexible linker (Caruthers and McKay, 2002). These domains correspond to amino acids 38-240 and 251-411 in eIF4A3 respectively. Their orientation is not defined in the absence of ATP and RNA. However, upon binding to ATP and RNA, the RecA-like domain 1 and 2 are brought in close proximity forming a cleft where ATP (in red) binds. On the opposite surface the two RecA-like domains bind the RNA (in black). This binding is sequence independent because eIF4A3 binds the ribose-phosphate backbone of the RNA (Bono et al., 2006). The SELOR domain of BTZ (aa 137-286 (Degot et al., 2004))(in orange) extends with two separate stretches over both domains of eIF4A3. The N-terminal stretch (aa 168–196) binds to domain 2 of eIF4A3 and the C-terminal stretch (aa 214–248) binds to domain 1. These

stretches are connected by a flexible linker. In addition, the SELOR domain of BTZ contributes to the interaction of BTZ to the RNA and to MAGOH (in blue) (Bono et al., 2006). The structure of the dimer MAGOH-Y14 in the EJC is almost identical to one of the isolated dimer (Fribourg et al., 2003; Lau et al., 2003). MAGOH folds into flat antiparallel β sheets flanked on one side by two parallel α helices, which bind the RNA binding domain (RBD) of Y14 (in violet). In the EJC the dimer MAGOH/Y14 interacts mainly with the RecA-like domain 2 of eIF4A3 (Bono et al., 2006). Interestingly, MAGOH/Y14 binds eIF4A3 only in the ATP-and RNA-bound state. This explains why *in vitro* MAGOH-Y14 inhibits the ATPase activity of eIF4A3. In contrast, SELOR stimulates ATPase activity of eIF4A3 (Bono et al., 2006).
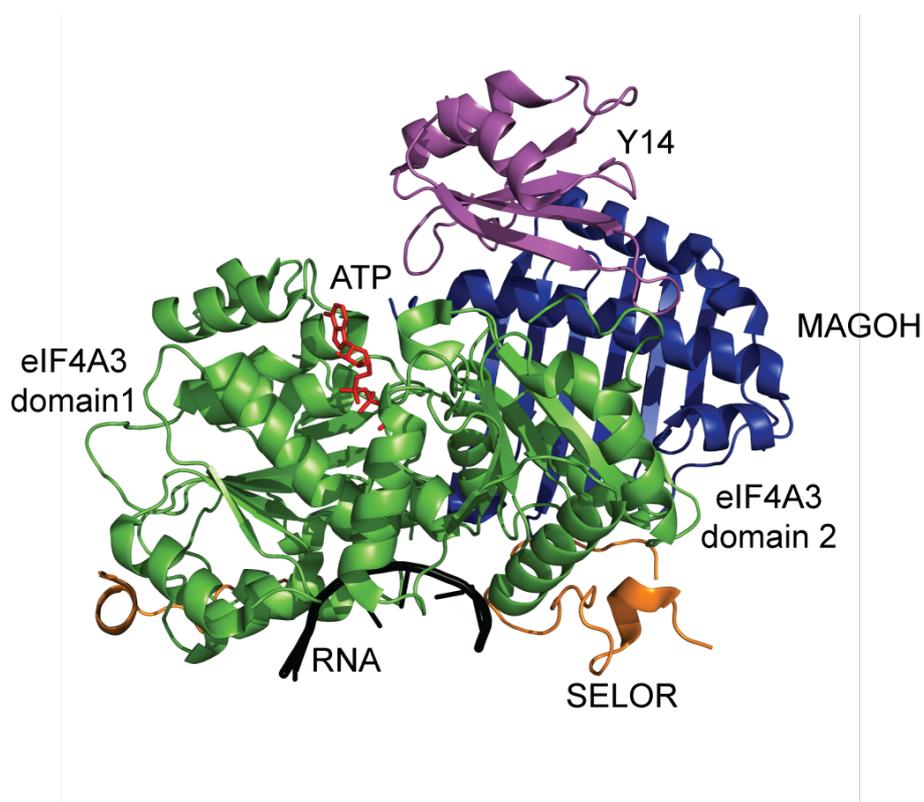
Figure 3. EJC crystal structure. The crystal structure of the EJC complex was modeled with pyMOL (PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) and the PDB file 2J0S (Bono et al., 2006).

2.2.2 EJC assembly and disassembly

As a consequence of splicing, the EJC is deposited on the mRNA 20-24 nt upstream of the exon-exon junction marking the position of the former introns (Le Hir et al., 2000a). The deposition of the EJC on the mRNA is a multi-step process. EJC assembly requires CWC22, a component of the spliceosome and a newly identified binding partner of eIF4A3 (Steckelberg et al., 2012). Before exon-exon ligation takes place, CWC22 brings eIF4A3 in close proximity to the mRNA (in the spliceosomal complex B). The subsequent conformational change, in the presence of ATP, induces the binding of eIF4A3 to the mRNA. As a result, the MAGOH/Y14 heterodimer can interact with eIF4A3 (spliceosomal complex B^{act}), stabilizing the trimeric pre-EJC on the mRNA. The trimeric pre-EJC remains stably associated to the mRNA while the introns are spliced out (spliceosomal complex C). Next, the mRNA associated complex is exported to the cytoplasm where, according to the current model, the protein BTZ joins the complex (Gehring et al., 2009a; Steckelberg et al., 2012) (Figure 4, from 1 to 5). Indeed, there is evidence that BTZ and the spliceosome bind the same eIF4A3 residues, supporting the hypothesis of BTZ joining to the EJC after completion of the splicing reaction. This assumption is corroborated by previous results, where BTZ was not identified in purified spiceosomal complexes (Bessonov et al., 2008; Gehring et al., 2009a). EJC proteins need to be recycled quickly and efficiently because of the limited cellular amount of EJC proteins and the high number of exon junctions in the steady-state transcriptome. EJCs located within the open reading frame (ORF) are removed by the ribosome in the cytoplasm during the first round of translation (Dostie and Dreyfuss, 2002; Lejeune et al., 2002). The protein PYM binds the dimer MAGOH-Y14 with its N-terminal domain (Bono et al., 2004; Forler et al., 2003), but it also interacts with the ribosome (Diem et al., 2007). For this reason, PYM was proposed to bridge the ribosome and the EJC and to stimulate the release of the EJC proteins from the mRNA. Binding of PYM to the ribosome guarantees that PYM removes the EJC only within the ORF from translated mRNA. This type of binding represents a safe mechanism that minimizes the amount of free PYM, in order to avoid EJC removal from not yet translated mRNAs (Gehring et al., 2009b). According to the proposed model, the

binding of PYM to the dimer MAGOH-Y14 induces a conformational change in MAGOH-Y14 that impairs their binding with eIF4A3. In this way, eIF4A3 is not locked on the mRNA anymore and is released together with the other EJC components (Nielsen et al., 2009) (Figure 4, 6-7) .
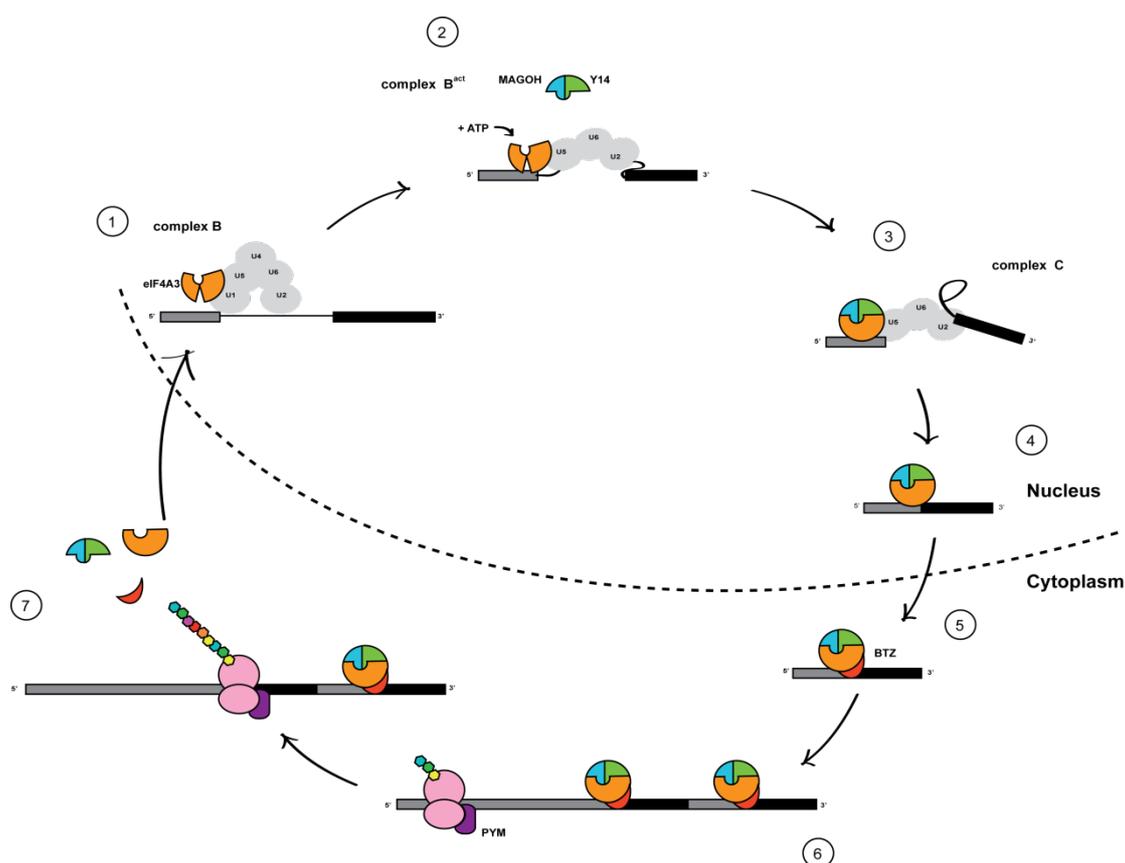
Figure 4. Model of EJC assembly and recycle. Steps 1 to 5 show the assembly of the EJC in the nucleus and the export to the cytoplasm. PYM displaces the EJC in the cytoplasm during translation (steps 6 and 7). This image was inspired by N.H. Gehring drawing.

2.2.3 Peripheral EJC components

As explained in paragraph 2.2.4, the EJC executes several functions. In order to mediate these functions, it needs to interact with other proteins (Degot et al., 2004; Diem et al., 2007; Ferraiuolo et al., 2004; Gatfield et al., 2001; Kataoka et al., 2001; Kataoka et al., 2000; Kim et al., 2001; Le Hir et al., 2001; Le Hir et al., 2000a; Le Hir et al., 2000b; Li et al., 2003; Luo et al., 2001; Lykke-Andersen et al.,

2000; Palacios et al., 2004; Zhou et al., 2000). In Figure 5, EJC-interacting proteins are shown and their functions are described in the next paragraphs.
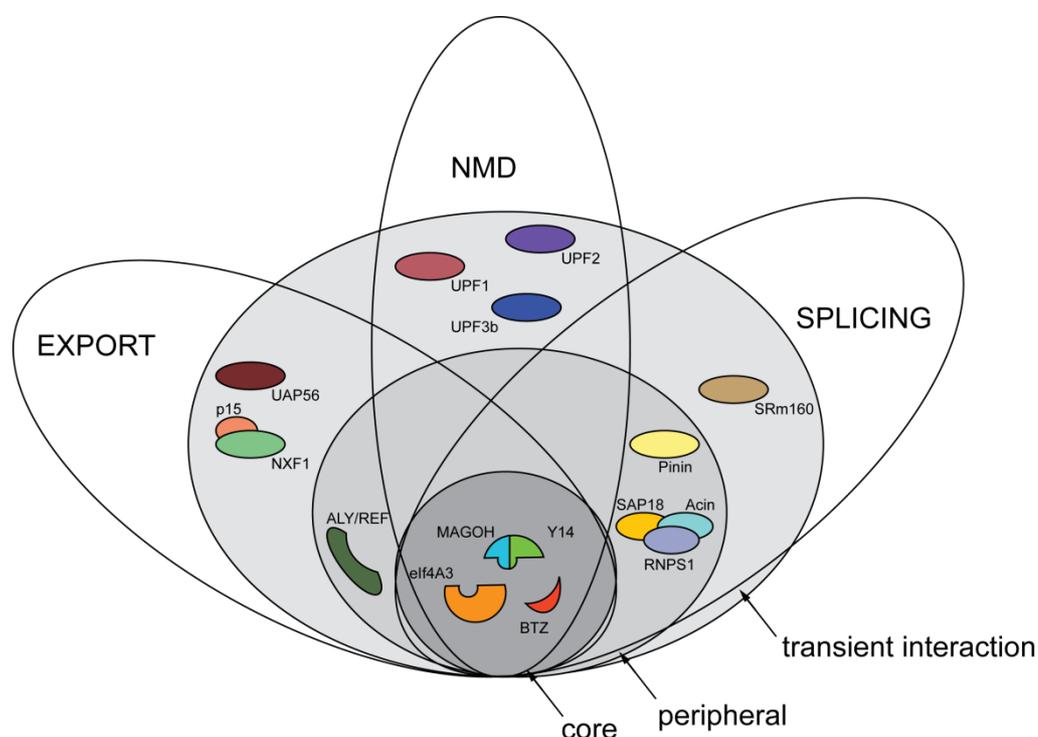
Figure 5. EJC-interacting proteins and functions. The core of the EJC (MAGOH,Y14,BTZ and eIF4A3) (dark gray circle) interacts with several peripheral EJC components (middle gray) and with transient binding partners (light gray) to execute different functions, e.g. NMD, export and splicing (The proteins involved in a common function are enclosed in an oval shape, with the function indicated).

2.2.4 EJC functions

The EJC influences the life of an mRNA through several processes, such as localization, splicing, translation and decay. With the exception of decay, which is described in paragraph 2.3, this paragraph discusses how the EJC influences these processes.

Localization

The EJC is required for proper localization of the *oskar* mRNA in *D.melanogaster* oocytes. *Oskar* mRNA is synthesized in the nurse cell nuclei and then transported along the microtubules into the adjacent oocyte (St Johnston, 2005). Expression

of *oskar* occurs exclusively at the posterior pole is precisely regulated: during transport along the microtubules, Bruno represses the translation of *oskar* (Gunkel et al., 1998; Kim-Ha et al., 1995). Once *oskar* reaches the posterior pole, Staufen activates its translation (Micklem et al., 2000). Translation occuring exclusively at the posterior pole is essential for the formation of the abdomen and the germ line of the embryo (Johnstone and Lasko, 2001). Several *trans*-acting factors control repression, transport and translation of the *oskar* mRNA. Examples of these controlling factors include the *D.melanogaster* EJC core components eIF4A3, Barentz, Mago Nashi and Tsunagi (MAGOH and Y14 in human) (Palacios, 2002; Palacios et al., 2004). In particular, BTZ is required for transport of *oskar* to the posterior pole. BTZ does, however, not affect other processes such as transcription, export, colocalization with Staufen, or translation (van Eeden et al., 2001). Similarly, van Eeden and collegues observed that mutations in *mago nashi* and *btz* resulted in the same phenotype and that the two proteins depend on each other for proper localization at the posterior pole. Subsequently, Y14 and eIF4A3 were identified to be responsible for proper *oskar* localization as well (Hachet and Ephrussi, 2001; Palacios et al., 2004). The *oskar* mRNA localization is the best studied example of the involvement of the EJC in mRNA localization. Another example of the contribution of the EJC proteins in mRNA localization comes from dendrites of mammalian hippocampal neurons. The dendrites are highly polarized cells and require translation of specific mRNAs. MAGOH, Y14 and BTZ accumulated in the dendrites where BTZ associates with the mRNA transport factor Staufen1. Staufen1 assembles in RNPs responsible to transport and localize RNAs into dendrites of mature hippocampal neurons (Giorgi and Moore, 2007; Glanzer et al., 2005; Macchi et al., 2003; Monshausen et al., 2004).

Splicing

The recognition of the splice site occurs in two different ways in long and short intron containing genes: in short introns (<200 bp), the 5′ and 3′ splice sites are recognized across the intron (intron definition); in long introns (>250 bp), splice sites of bordering exons have to be recognized before splicing can occur (exon definition) (Fox-Walsh et al., 2005; Sterner et al., 1996). The influence of the EJC

over the splicing process is associated with this second phenomenon of exon definition. In *D.melanogaster* the nuclear EJC influences splicing of the *mapk* gene exclusively. The splicing of the other genes of the RAS1/MAPK pathway is not influenced by the EJC, indicating a selective choice for a specific mRNA (Ashton-Beaucage et al., 2010). EJC removal leads to exon skipping (Ashton-Beaucage et al., 2010) and, of the EJC core components, only the depletion of BTZ does not affect splicing of the *mapk* pre-mRNA. This is in line with the observation that BTZ deposition on the mRNA occurs after completion of the splicing reaction (Gehring et al., 2009a). So far it was not possible to define if the effect of the EJC was direct or indirect. One possibility is that the EJC works as a binding platform for the splicing effectors SRSF proteins (Long and Caceres, 2009). Another option is that the EJC stabilizes the interaction between the spliceosome and the splice site (Ashton-Beaucage et al., 2010). In 2012, Michelle and collegues analyzed the effect of three categories of EJC components on splicing of the apoptotic regulator *Bcl-x*: core, peripheral and NMD associated. The depletion of the core components Y14 and eIF4A3 and the peripheral RNPS1, Acinus and SAP18 was associated with an increase of the proapoptotic splicing variant $Bcl\text{-}x_S$. On the contrary, no effect was observed when the export factors UAP56, ALY/REF and the NMD proteins UPF1, UPF2 and UPF3b were depleted. Similarly, no effect was observed when BTZ was depleted. Interestingly, the function and the deposition of the EJC components in *Bcl-x* modulation depend on *cis*-acting elements, suggesting a different type of EJC assembly and regulation (Michelle et al., 2012).

Translation

It was first observed that intron-containing transcripts had a higher translation rate compared to the intron-less upon injection in *Xenopus* oocytes (Matsumoto et al., 1998). Subsequently, in mammalian cells an increased mRNA translation was observed upon the deposition of the EJC (Lu and Cullen, 2003; Nott et al., 2003). No detectable positive effect on the expression of the mRNA was observed for the human β-globin gene where the intron was positioned so close to the 5′ end of the mRNA that the EJC could not assemble, even though splicing of this mRNA still occurred. In contrast, tethering of the EJC proteins RNPS1 and SRm160 increases expression of intronless β-globin (Wiegand et al., 2003).

Similarly, Nott and colleagues observed that tethering of the EJC components Y14, MAGOH and the more peripheral component RNPS1 led to translational stimulation of the reporter transcript. The EJC components might contribute to the increased mRNA translation in different ways. One explanation is the interaction of PYM with the dimer MAGOH/Y14. The binding of the N-terminus of PYM to MAGOH/Y14 and the C-terminus to ribosome would recruit EJC bound mRNAs to the translation machinery (Diem et al., 2007). This model is supported by previous observation that cytoplasmic polysome association resulted in enhanced spliced mRNAs (Nott et al., 2004). Another possibility is the interaction of BTZ with the translation initiation factor eIF3 (paragraph 2.4.2) (Chazal et al., 2013) or the recruitment of the TREX complex. Export of the mRNA to the cytoplasm would then make it available for the translation machinery (Le Hir et al., 2001; Luo and Reed, 1999; Luo et al., 2001; Zhou et al., 2000).

2.3 Nonsense-mediated mRNA decay

During gene expression, a frequently occurring error is the generation of premature termination codons (PTCs). PTCs can arise at the DNA level as a consequence of nonsense mutations, deletion, insertions or somatic rearrangements. Alternatively, mutations in functional motifs, such as splice site or splice regulatory elements that lead to alternative splicing events, can generate PTCs (Nicholson et al., 2010). One-third of all alternative splicing events lead to the formation of PTCs (Lewis et al., 2003). In principle, the presence of PTCs can lead to synthesis of truncated, non-functional and sometimes deleterious proteins (Frischmeyer and Dietz, 1999; Holbrook et al., 2004). Eukaryotes have acquired an evolutionary conserved surveillance mechanism, named nonsense-mediated mRNA decay (NMD), to identify and degrade these aberrant mRNAs (Amrani et al., 2006; Behm-Ansmant et al., 2007; Conti and Izaurralde, 2005; Culbertson, 1999; Muhlemann et al., 2008). According to the current model, in higher eukaryotes a stop codon is recognized as premature when situated more that 50 nt upstream of an exon-exon junction and, consequently, of a deposited EJC (Figure 6-1) (Nagy and Maquat, 1998; Zhang et al., 1998a; Zhang et al., 1998b). Two main complexes are required for NMD

activation: one is the EJC, downstream of the stop codon, the other the SURF complex. When the ribosome stops either at a normal stop codon or at a PTC, the eukaryotic release factors (eRF) 1 and 3 interact with the A-site of the ribosome (Figure 6-2). For efficient translation termination, the interaction of eRF3 with the cytoplasmic poly(A)-binding protein (PABPC1) is required (Cosson et al., 2002a; Cosson et al., 2002b). However the up-frameshift protein UPF1 competes with PABPC1 for binding to eRF3, affecting the normal termination process (Ivanov et al., 2008). eRF3-bound UPF1 in turn interacts with SMG1, forming the so-called SURF complex (SMG1, UPF1, eRF 1 and 3) (Czaplinski et al., 1998; Kashima et al., 2006) (Figure 6-3).
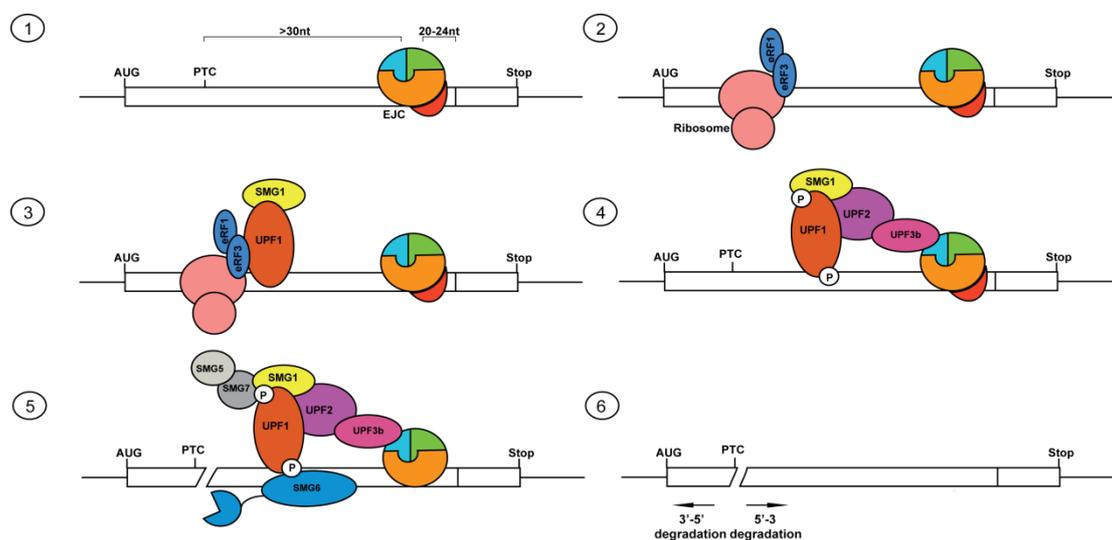
Figure 6. Model of NMD activation. 1- EJC is deposited 20-24 nt upstream of the exon-exon junction. 2- The ribosome stalls at the PTC and recruits the eukaryotic release factors (eFR1-3). 3-The SURF complex (SMG1, UPF1, eRF 1 and 3) is assembled. 4-UPF3b and UPF2 bridge the SURF complex to the EJC. 5- SMG5, SMG 7 and SMG 6 are recruited and activate the degradation process 6- Generation of the two unprotected RNA fragments.

The up-frameshift proteins UPF3b and UPF2 bridge the EJC to the SURF complex through binding to UPF1 (Chamieh et al., 2008; Kashima et al., 2006)(Figure 6-4). This bridging activates the kinase activity of SMG1, which phosphorylates UPF1 (Figure 6-4) (Isken and Maquat, 2008; Kashima et al., 2006; Ohnishi et al., 2003; Yamashita et al., 2001). Once UPF1 is phosphorylated, it serves as a binding platform for the effector proteins SMG5, SMG6 and SMG7 (Figure 6-5).

Two possible pathways are activated to degrade the mRNA: the first is SMG5-SMG7 dependent and activates deadenylases and decapping enzymes, thereby leading to an exonucleolytic degradation of the mRNA (Loh et al., 2013; Unterholzner and Izaurralde, 2004); the second is SMG6 dependent. SMG6 executes an endonucleolytic cleavage of the substrate, thereby generating two unprotected RNA fragments, which are subsequently degraded by the 5′-to-3′ exonuclease XRN1 or by the 3′-to-5′ exosome complex (Figure 6-6) (Eberle et al., 2009; Huntzinger et al., 2008). In addition to the EJC-dependent NMD, another way has been described to activate NMD. In presence of a long 3′ UTR, the ribosome might fail to terminate properly due to the long distance between PABPC1 and the eRFs ("faux" 3′ UTR model) (Amrani et al., 2004; Behm-Ansmant et al., 2007; Eberle et al., 2008; Singh et al., 2008). In this case, as well as described above, UPF1 binds the eRFs, activating the downstream cascade of events that leads to the mRNA degradation (Kashima et al., 2006).

The medical impact of NMD was first observed in β-thalassemia, where a single nucleotide deletion produces a premature stop codon, leading to a reduced half life of the corresponding mRNA (Kinniburgh et al., 1982). β-thalassemia is a valid example of the correlation between NMD and pathologies. In case of a homozygous mutation, the β-globin chain is not produced due to degradation of its mRNA via NMD. In heterozygous mutations, the normal allele can produce enough β-globin subunits to support near normal hemoglobin levels (Hall and Thein, 1994). A third situation is the NMD-insensitive last-exon PTC mutations, where truncated nonfunctional β-globin is produced (Thein et al., 1990). NMD is not only responsible for the degradation of mutated mRNA that would lead to the translation of truncated and deleterious proteins, it also targets 5-10% of naturally occurring transcript (Mendell et al., 2004). One example are members of the SR-family of splicing factor: SRSF2 and SRSF3 promote the alternative splicing of their own mRNA, leading to NMD sensitive variants (Sureau et al., 2001). Another example is the regulation of the mammalian selenium-dependent glutathione peroxidase 1 (Se-GPx1) mRNA. This mRNA contains a UGA codon that is recognized as a codon for the nonstandard amino acid selenocysteine (Sec). But in presence of reduced concentrations of selenium, the UGA codon is

decoded as a stop codon, leading to the subsequent degradation of the transcript (Moriarty et al., 1998).

2.4 Barentsz

Human BTZ was originally identified in a screening of breast cancer metastatic lymph node cDNAs, hence also being known as MLN51 (Metastatic Lymph Node 51). It is located on the long arm of chromosome 17 in the q11-q21.3 region (Tomasetto et al., 1995). The name CASC3 (Cancer Susceptibility Candidate Gene 3) derives from a different screening of genes of which expression was increased in cancer (Arriola et al., 2008), while the name BTZ derives from the *D.melanogaster* ortholog. The protein is highly conserved in mammals (90% homology with cow, mouse rat and pig) and in invertebrates (41% and 48% similarity respectively with *C.elegans* and *D.melanogaster*). The region 168-256 of human BTZ was found to be more conserved than the rest of the protein, with 100% identity among human, rat and mouse. Because this region is responsible for RNA binding and localization to the nuclear speckles, it was named SELOR, for speckle localizer and RNA binding module (Degot et al., 2004).

2.4.1 BTZ structure

The ORF of BTZ is 4119 bp long and encodes a 703 aa protein with a predicted molecular weight of 76 kDa (Degot et al., 2002). Figure 7 displays a schematic representation of BTZ. The full-lenght protein localizes to the cytoplasm, with the nuclei weakly stained, despite two nuclear localization signals (NLSs) (PKGRQRK, 204 – 210; PRRIRKP, 255 –261) being present in the SELOR domain (Degot et al., 2002). However, these NLSs are functional: expressing only the sequence containing the NLS (203-261 and 1-351) of BTZ in cell culture lead to localization of these proteins to the nucleus (Degot et al., 2002). On the contrary, the C-terminal domain, which contains a nuclear export signal (NES) (462-472), localizes to the cytoplasm (Degot et al., 2004). According to the running behavior during SDS PAGE, the apparent molecular weight of BTZ is 110kD, although the calculated weight is just 76 kD. Several factors contribute to the different protein migration. A putative coiled-coil domain is present in the N-terminal part of the

protein (92 – 130) contributing to protein oligomerization (Degot et al., 2002). In addition, BTZ has a high proline content (14.4% of the total amino-acid composition), especially in the C-terminal domain (22.7% from amino acids 352 – 703) and a conserved glutamine-rich region (608-675). The proline-rich region forms relatively rigid structures that confer a retarded migration on conventional SDS-containing gel (Hansen et al., 1998; Schreiber et al., 1998). Moreover, the proline-rich region contributes to the protein oligomerization by binding the Src homology region 3 (SH3) domain (Degot et al., 2002). Two classes of SH3 binding motifs have been described according to the position of a positively charged residue in the environment of the PXXP binding motif. Class I and class II SH3 binding motifs correspond to the following consensus sequences +XXPXXP and PXXPX+ (where + refers to a positively charged amino acid) (Kay et al., 2000). One class I motif (RPVPEPP, 528 – 534) and three class II motifs (PPPPDR, 392 – 398; PTPPTK, 442 – 447; PSPPRR 678 – 683) are present in BTZ. Another factor contributing to the protein oligomerization is one putative tyrosine phosphorylation site (Y404) (Degot et al., 2002). This site can be bound by Src homology region 2 (SH2) domains, which bind to phospho-tyrosine (Mayer et al., 1988; Songyang et al., 1993).
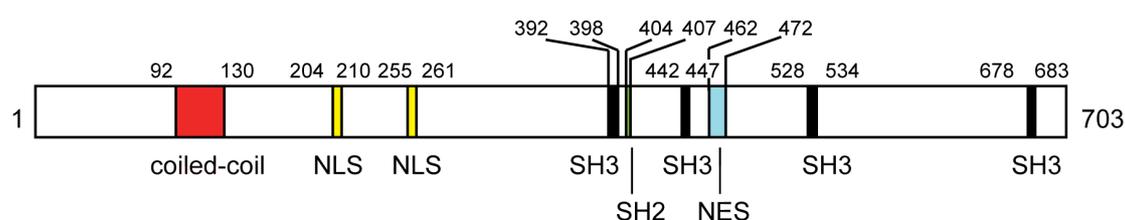
Figure 7. Schematic representation of BTZ. The figure is drawn to scale with the coiled-coil domain in red, the NLSs in yellow, the SH3 domain binding sites in black, SH2 in green and the NES in turquoise. The aa position is indicated above.

2.4.2 BTZ functions

Functionally, BTZ can be divided in three regions: N-terminus, SELOR and C-terminus (Figure 8). The role of the SELOR domain, as a mediator of EJC and RNA binding, was described in paragraph 2.2. Concerning the N-terminus, the only function reported so far was its role in NMD. Gehring and colleagues showed in

2009 that the removal of the N-terminal domain of BTZ significantly impaired BTZ efficiency in tethering assays. Several other BTZ functions have been reported and are described in this paragraph separately.
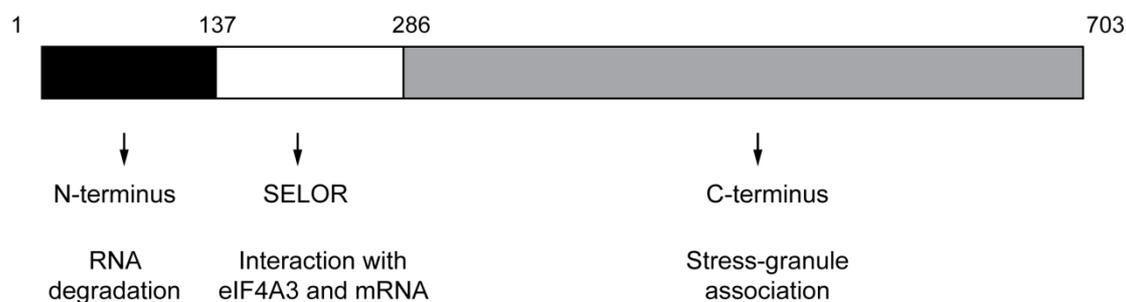
Figure 8. BTZ domain structure. Three functional domains compose the protein BTZ. N-terminus (aa 1-137) involved in mRNA degradation; SELOR (aa 137-286), responsible for eIF4A3 and mRNA binding; C-terminus (aa 286-703), involved in stress-granule assembly.

Stress granule assembly

The C-terminus of BTZ was proposed to be the region responsible for the assembly of BTZ in stress granules (SGs) (Baguet et al., 2007). SGs are cytoplasmic aggregates composed of proteins and RNAs and are formed under unfavorable conditions. They represent an adaptive cellular response to environmental stress and are mainly known as dynamic cytoplasmic foci where stalled 48S preinitiation complexes accumulate (Kedersha et al., 2002). Many SG-resident proteins are RNA-binding proteins involved in different aspects of mRNA function, such as translation (TIA1, TIAR and PABP), stability (HuR, TTP), degradation (G3BP and PMR1) and localization (Staufen, Smaug and FMRP) (Anderson and Kedersha, 2006). These proteins shuttle rapidly in and out of SGs, supporting the idea that SG are not static storage centers, but rather dynamic structures (Anderson and Kedersha, 2006; Kedersha and Anderson, 2002). BTZ also shuttles rapidly in and out of the SGs, with the isolated C-terminus having a higher mobility compared to the full-length protein. Specifically, the 220 C-terminal aa are responsible for BTZ recruitment to the SGs (Baguet et al., 2007). The same region is dispensable for EJC assembly (Degot et al., 2004). TIA1 and Pumilio 2 have been shown to be targeted to SGs via a glutamine-rich prion-related domain (PRD), which is responsible for their self aggregation (Gilks et al.,

2004; Vessey et al., 2006). The C-terminal domain of BTZ contains several phosphorylation sites at serine residues, SH2 and SH3 binding sites and a conserved glutamine-rich region (aa 608-675) (Degot et al., 2004; Degot et al., 2002). Both phosphorylation sites and the conserved glutamine-rich region may govern BTZ association with SGs (Baguet et al., 2007).

NMD

In order to activate NMD, two pathways were suggested: BTZ-dependent and UPF2-dependent. The BTZ-dependent pathway requires a fully assembled EJC, where UPF3b binds UPF1, without bridging by UPF2. In this pathway, the EJC is fundamental and any mutation in the four core components that affect their binding disrupts the NMD function (Gehring et al., 2009a). The existence of this UPF2-independent NMD-activating mRNP can also explain some earlier findings. UPF3 proteins can coimmunoprecipitate UPF1 independently of the binding to UPF2 (Ohnishi et al., 2003). In addition, UPF3a (short isoform) and UPF1 have both been detected in high molecular weight complexes lacking UPF2 (Schell et al., 2003). Finally, the interaction of UPF2 with tethered UPF3b seems to be dispensable for NMD, a finding that was difficult to reconcile with a linear UPF3-UPF2-UPF1 pathway (Gehring et al., 2003). The second proposed pathway is EJC-independent and UPF2-dependent. In this pathway RNPS1 binds the mRNA via its RRM (RNA recognition motif) domain and activates UPF1 through UPF3b and UPF2 (Gehring et al., 2009a).

Translation initiation

BTZ has been identified as binding partner of eIF3 and was proposed to stabilize the initiating complex, promoting translation initiation (Chazal et al., 2013). Chazal and colleagues showed that BTZ overexpression is associated with increased translation of the Firefly luciferase reporter. Similarly, BTZ depletion reduced mRNA translation, whereas complementation of the depleted extracts by de novo-synthesized BTZ restored translation efficiency, showing that BTZ is a bona fide regulator of translation. Although BTZ is able to stimulate translation of mRNAs that have not undergone splicing, its effect is much more pronounced on spliced mRNA. This function of BTZ is EJC-dependent, because BTZ mutants

that cannot be assembled into the EJC (Ballut et al., 2005; Daguenet et al., 2012) do not stimulate translation. Notably, this function is specific of BTZ, since the other EJC core components are not able to stimulate translation (Chazal et al., 2013).

2.5 EJC and NMD in *D.melanogaster* and *C.elegans*

The EJC core proteins are, like the NMD proteins, highly conserved among human, *C.elegans* and *D.melanogaster* (Bono et al., 2006). SMG1, SMG5, SMG6 and SMG7 are required for NMD in *C.elegans*, *D.melanogaster* and in mammalian cells (Conti and Izaurralde, 2005), as well as the UPF proteins (SMG2, SMG3 and SMG4 in *C. elegans*) (Leeds et al., 1991; Leeds et al., 1992). While in mammals and *D.melanogaster* NMD is required for viability (Metzstein and Krasnow, 2006; Weischenfeldt et al., 2012), in *C.elegans* genetic deletions of the NMD components are associated only with morphogenic alteration (Hodgkin et al., 1989). The analysis of the spliceosome C complex in *D.melanogaster* revealed the presence of the EJC core components eIF4A3, MAGO and Y14. BTZ was not identified in the C complex, suggesting a later association with the EJC (Herold et al., 2009). However, it has not been finally solved whether NMD in *D.melanogaster* is EJC-dependent or not. In 2003, Gatfield and colleagues showed that Y14 and RNPS1 are not required for NMD and that the recognition of premature stop codons is EJC-independent. They postulated that PTCs might be recognized through downstream sequence elements, the *D.melanogaster* analogue of the exon junction, to which the UPF proteins could bind (Wagner and Lykke-Andersen, 2002). Interestingly, it was observed that the EJC in *D.melanogaster* was not deposited on all the exon-exon junctions but when present it promoted NMD. Natural intron–containing 3′ untranslated regions (UTRs) were inserted downstream of the stop codon of a firefly luciferase reporter. These reporters were stabilized in UPF proteins depleted cells. Upon immunopurification of RNA associated with the EJC components, the same reporters were specifically enriched (Sauliere et al., 2010).

Instead, in *C.elegans* the EJC components are not required for PTC definition or for NMD (Longman et al., 2007). Rather, the EJC proteins MAGO, Y14 and eIF4A3

are needed for retention in the nucleus of improperly spliced mRNA. Even though they do not affect splicing efficiency, their depletion is associated with a premature export of unspliced or partially spliced mRNA. Only BTZ seems to be irrelevant for this phenomenon (Shiimori et al., 2013).

3. Aims of the Project

The Exon Junction Complex and its components have been extensively studied. Much is known about the role of the EJC components in the processes of splicing, translation and decay. However, only few data are available about the role of BTZ inside the EJC or as an independent protein.

The purpose of this project is to better define the functions of BTZ. First I aim to dissect to role of the single domains. The SELOR domain has been identified as the RNA binding and eIF4A3 binding domain. SELOR has not been studied before *in vivo* as an isolated domain. I plan to understand if some of the previous described functions of BTZ are recapitulated by the short SELOR domain and in addition, if these functions are EJC-dependent or independent. Furthermore, the N-terminus and C-terminus of BTZ are not very well characterized; I aim to get a better inside in the role of these two regions. Moreover, to better define the role of BTZ in cells, I intend to identify new BTZ binding partners and possibly to define the BTZ RNA interactome.

4. Results

BTZ and its molecular function in cells has been mainly studied as a constituent of the EJC (Bono et al., 2006; Bono and Gehring, 2011; Gehring et al., 2009a; Tange et al., 2005). The aim of this work is to further characterize the role of BTZ as a core component of the EJC and additonally to identify EJC-independent functions. The results chapter is divided in 4 parts: (1) functions of individual domains of BTZ, (2) BTZ-induced NMD, (3) BTZ binding to the RNA and (4) BTZ protein interactome.

4.1 Functions of BTZ domains

The BTZ protein can be divided in three regions: N-terminus (N), SELOR domain (S) and C-terminus (C) (see paragraph 2.4.2). These regions have not been studied in isolation, with the exception of the SELOR domain, which plays a role in eIF4A3 and RNA binding as previously described (see paragraph 2.2). In the following paragraphs I aim to identify their specific functions.

4.1.1 Reduction of mRNA levels upon BTZ mutants tethering

To elucidate the functions of BTZ regions, several length mutants were generated. Besides the full-length construct, which contains the three aforementioned regions (FL), constructs were cloned containing one (137-286 (S), 286-703 (C) and 1-137 (N)) or two complete regions (137-703 (SC), 1-286 (NS)) or regions of intermediate size (NSC-short, N-long, S-long and C-short) (Figure 9A). To study their function, eight of the BTZ constructs were selected and expressed in cultured human cells as λNV5-tagged fusion proteins and thereby tethered to the co-transfected β-globin mRNA reporter construct containing 4boxB binding site in its 3′ UTR (Gehring et al., 2008). In this way the protein of interest is recruited downstream of the natural stop codon, where it can influence the composition of the mRNP. It is well established that tethering of NMD or EJC components lead to decreased reporter mRNA levels due to the initiation of mRNA decay (Gehring et al., 2008; Lykke-Andersen et al., 2000). The

unfused λNV5-tag was used as negative control and the corresponding mRNA reporter level set as 100%. As previously reported (Gehring et al., 2009a), tethering of BTZ FL results in a robust reduction of the reporter mRNA levels (Figure 9B, upper panel, compare lane 2 with 1). Furthermore, all deletion variants of BTZ caused a strong, albeit variable decrease of the reporter mRNA level (Figure 9B, lanes 2-9). Interestingly four of these variants (Figure 9B, lanes 2-6) contain the SELOR domain of BTZ, which mediates binding to the EJC via a direct interaction with eIF4A3 (Ballut et al., 2005; Degot et al., 2004) and two contain only small parts of the N-terminus (Figure 9B, lanes 8 and 9). The C-terminal region instead will be analyzed in isolation in the next paragraph.
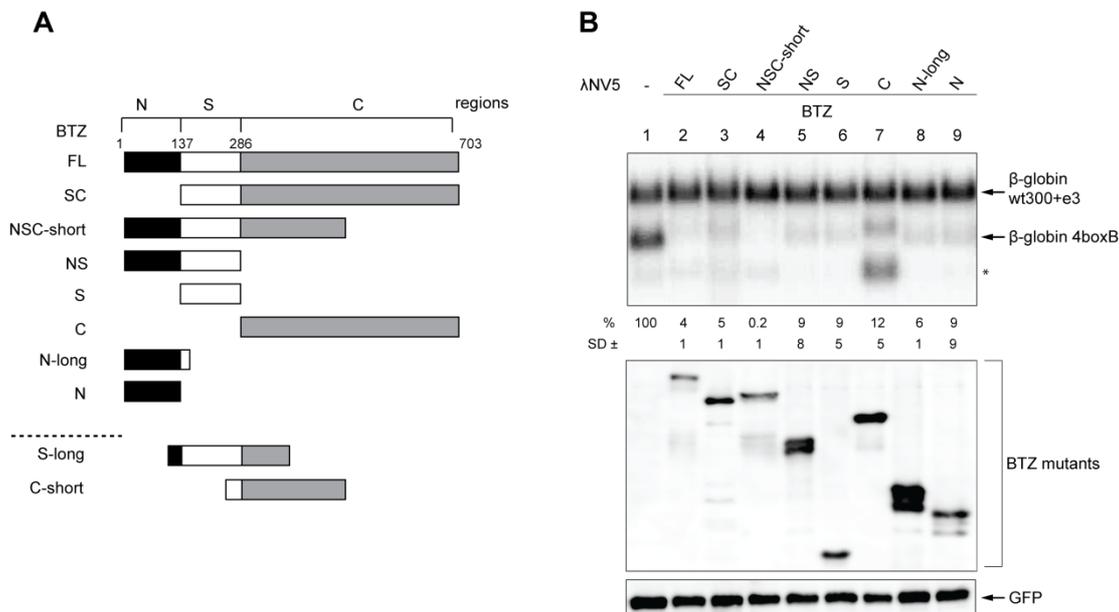
Figure 9. BTZ domains elicit NMD. (A) Schematic representation of full-length BTZ and mutants (black: N-terminus, white: SELOR, grey: C-terminus domains). NSC-short corresponds to the aa 1-480; N-long to the aa 1-160; S-long to the aa 110-372; C-short to the aa 255-480. (B) Northern blot analysis (upper panel) of RNA samples extracted from HeLa cells transfected with the indicated BTZ mutants, β-globin 4boxB reporter mRNA and β-globin wt 300+e3 as control mRNA (n=3). Western blot analysis (bottom panel) of the same samples with an anti-V5 antibody, anti-GFP is used as transfection control. The asterisk indicates a fast running band detected only upon tethering of BTZ C. For the quantification in lane 7 the upper running band was considered.

4.1.2 BTZ C-terminus activates polyadenylation

When the C-terminus of BTZ was tethered, two reporter bands were observed (Figure 9B, lane 7). In addition, the upper reporter band appeared to run slightly slower than in all other lanes (Figure 9B, compare lane 7 with the others). To define if tethering of BTZ C induced the decrease of the reporter mRNA, the upper band was used for the quantification (Figure 9B, lane 7). However when the amounts of both bands were combined (i.e. upper and lower bands), 66% of the reporter mRNA was still present at the steady state level. It seems that tethering of BTZ C does not degrade the reporter mRNA, but a different process occurs leading to the formation of the faster migrating band. To confirm that the two bands derived from the β-globin 4boxB reporter and that they were not unspecific products of the β-globin wt 300+e3 control mRNA, the λNV5 BTZ C was tethered to the reporter construct in absence of the control mRNA (Figure 10A, lane 2). Even in this condition, two bands were detected, one with faster, the other with slower motility compared to the reporter band in the negative control (Figure 10A, compare lane 2 with lane 1). The same mRNA samples that were analysed by Northern blotting, were used for PCR amplification and cloning into the pGEM-T Easy Vector (Promega). Clones were sequenced using the following strategy. For the retrotranscrition a oligo(dT) adaptor primer was used (oPR1628). This primer adds an adaptor sequence 5′ to the oligo(dT) sequence. This adaptor sequence is complementary to the antisense primer oRP1630 used for PCR amplification (Harigaya and Parker, 2012). As sense primer an internal globin primer was used. The result of the sequencing is schematically shown in Figure 10B. The β-globin 4boxB reporter has a cryptic polyadenylation signal (AUUAAA, white box) upstream of the 4boxB binding site (black box), while the canonical polyadenylation signal (AAUAAA, second white box) is downstream of 4boxB binding site. The faster migrating band lacked completely the 4boxB binding site and was polyadenylated downstream of the cryptic polyadenylation signal AUUAAA (β-globin short). In contrast, the slower migrating band corresponded to the full-length reporter β-globin 4boxB. A possible reason for its slower migration, might be a longer poly(A) tail, but the length of the poly(A) tail could not be measured by this experimental approach. I hypothesized that tethering of BTZ C to the reporter stimulates 3′ end processing

at the cryptic upstream site and that BTZ C acts similar to 3′ end processing factors that bind to the downstream sequence element of the polyadenylation signal (see paragraph 2.1) (Chan et al., 2011; Elkon et al., 2013). In order to test this hypothesis, I tethered the 3′ processing factors CFIm68, CSTF1 and CSTF3 (aa 20-600, comprising the HAT domain of CSTF3, required for the interaction with CSTF1 (Bai et al., 2007)) to the reporter mRNA. CFIm68 forms together with CFIm25 the CFI complex that binds upstream, whereas CSTF1 (CstF50) and CSTF3 (CstF77) are part of the CSTF complex that binds downstream of canonical 3′ end processing signals.
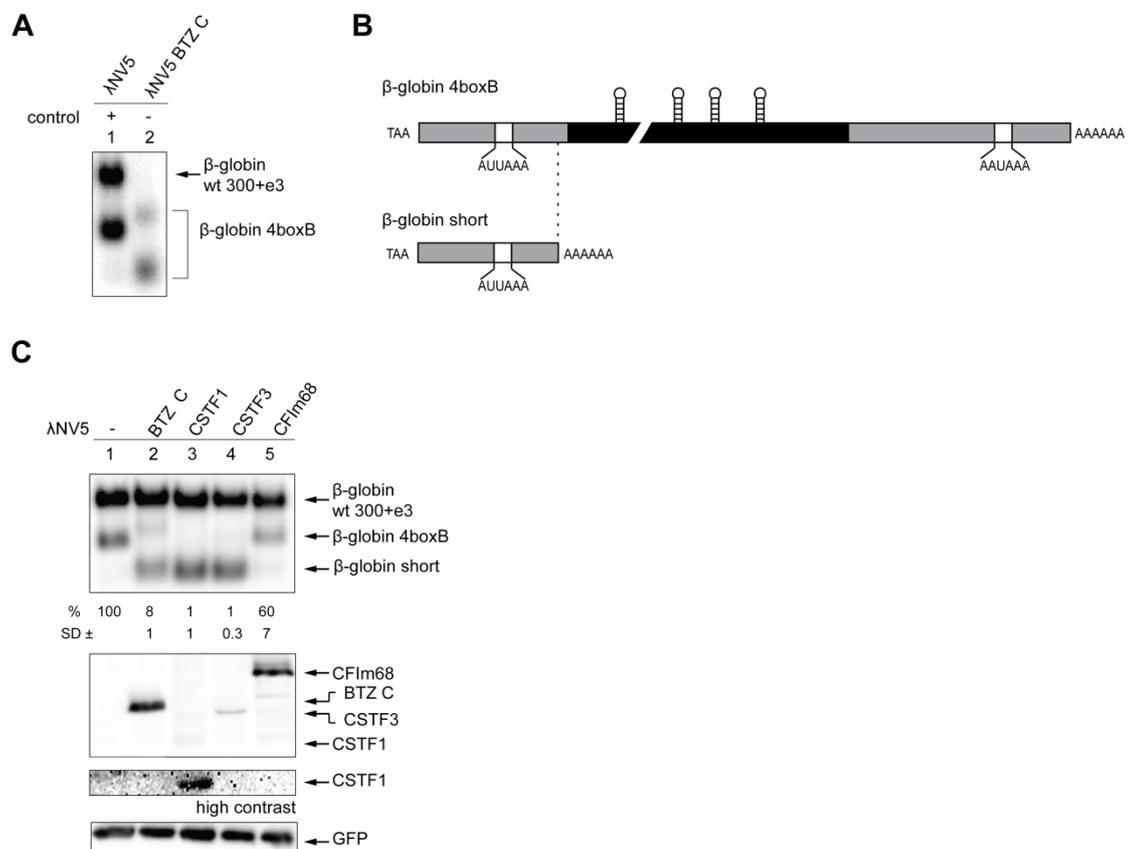
Figure 10. BTZ C-terminus induces polyadenylation at a cryptic polyadenylation signal. (A) Northern blot analysis of RNA samples transfected with the indicated constructs. In lane 2 the control mRNA β-globin wt 300+e3 was omitted. (B) Schematic representation of β-globin reporter products detected upon BTZ C tethering. In black the boxB binding site is shown while the non-canonical polyadenylation site (AUUAAA) and the canonical (AAUAAA) are depicted in white. The 3′UTR of the β-globin 4boxB reporter is representedin grey. (C) Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B.

29

Tethering of CFIm68 did not alter the behavior of the reporter mRNA, indicating that polyadenylation occurs at the canonical polyadenylation signal. On the northern blot this band migrated at the same level of the negative control sample, suggesting that the poly(A) tail length was also unaltered (Figure 10C, compare lane 5 with lane 1). On the contrary, tethering of CSTF1 and CSTF3 resulted in a faster migrating band (Figure 10C, compare lane 3 and 4 with lane 1), suggesting that the upstream cryptic polyadenylation signal was used. A similar phenotype was observed upon tethering of BTZ C and therefore implies that the C-terminus of BTZ is directly or indirectly involved in alternative polyadenylation.

In order to investigate if this function depends on the interaction with the EJC, I tested the binding of the C-terminal domain of BTZ to the other EJC components. Due to the low expression of FLAG-tagged BTZ C, a shorter C-terminal mutant was used, C-short (Figure 9A). FLAG-tagged S-long (see paragraph 4.1.3) and C-short expressing Flp-In T-REx 293 stable cell lines were established to facilitate the analysis of interaction with cellular EJC components. Endogenous eIF4A3 and MAGOH served as positive controls for the EJC binding in co-immunoprecipitation experiments. While S-long co-immunoprecipitated the EJC components, C-short failed to do so (Supplementary figure S1). This is in line with previous observations showing that the C-terminal domain of BTZ is not required for the EJC binding (Degot et al., 2004). Hence, the polyadenylation altering function of the C-terminal domain of BTZ is EJC-independent. This 3' end processing function of the isolated C-terminus of BTZ has not been previously described and further analyses will be required to elucidate the molecular mechanism of this phenomenon.

4.1.3 Point mutations in the SELOR domain affect NMD function

So far, I observed that two independent regions of BTZ can induce the reduction of reporter mRNA levels, namely the N-terminus and the SELOR domain. Since no structural information is available for the N-terminus, I focused on the SELOR domain and aimed to identify the residues responsible for its activity. The crystal structure of SELOR bound to eIF4A3 (Andersen et al., 2006; Bono et al., 2006) shows two main sites of interaction between the two proteins, the SELOR N-

terminal and C-terminal binding sites. Taking advantage of the crystal structure, point mutations were designed to disrupt this binding. In Figure 11A eIF4A3 is shown in green, SELOR in red and the designed mutations of SELOR in blue. The residues for the mutational study were selected due to the position at the SELOR-eIF4A3 interface and because they were previously tested in the context of the full-length BTZ (Gehring et al., 2009a). Gehring and colleagues showed that these mutations weaken the binding of BTZ with the EJC, but they had no effect on the function of the protein in the tethering assay, since the reporter mRNA levels were still reduced. I decided to study the effect of those mutations in the isolated SELOR domain. The mutants were cloned as λNV5-tagged fusion proteins and tested for their ability to decrease the reporter mRNA levels in the tethering assay. Surprisingly, the effect of the mutations was more prominent in the SELOR domain alone than in the full-length protein. While the mutations in the C-terminal binding site (e.g. HD220AK) strongly affected or completely abolished the ability of the SELOR domain to decrease the reporter mRNA level in the tethering assay, the mutations in the N-terminal binding site (e.g. DED173KRK) resulted in a weaker impairment of SELOR function (Figure 11B). This is in line with previous observations according to which the C-terminal binding site is required for the binding of SELOR and eIF4A3 *in vivo* and *in vitro*, while mutations in the N-terminal binding site, despite disrupting the interaction *in vivo*, do not affect the *in vitro* binding (Bono et al., 2006). To analyze the ability of these mutants to interact with the core EJC proteins, I decided to establish Flp-In T-REx 293 cell lines expressing FLAG tagged SELOR and mutants. Since FLAG-tagged SELOR was weakly expressed either as stable cell line or in transient transfection, a longer version, S-long (Figure 9A), was used. FLAG-tagged S-long in contrast to FLAG-tagged SELOR could indeed be expressed in Flp-In T-REx 293 cell lines (Figure 11C, lanes 2-8, input). To investigate the ability to assemble the EJC, immunoprecipitation of endogenous eIF4A3 and MAGOH was tested. S-long wt immunoprecipitated endogenous eIF4A3 and MAGOH (Figure 11C, lane 2, IP), whereas the binding to eIF4A3 and MAGOH was observed to be gradually weaker in the mutants DED173KRK and PR183AE (Figure 11C, lanes 3 and 4, IP). The binding was completely lost for all the other point mutants (Figure 11C, lanes 5-8, IP). Therefore the result of the tethering assay in Figure 11B is in line

with the interaction data, indicating that SELOR requires the binding to the EJC in order to exert a function on the reporter mRNA. Moreover to confirm that S-long is functionally identical to the shorter SELOR domain in tethering assays, I selected two repesentative mutations, one for each binding site, F188D and W218D. As expected, it was likewise observed that the mutation in the C-terminal binding site more extensively affects the function of the tethered protein than the mutation in the N-terminal binding site (Supplementary figure S2).
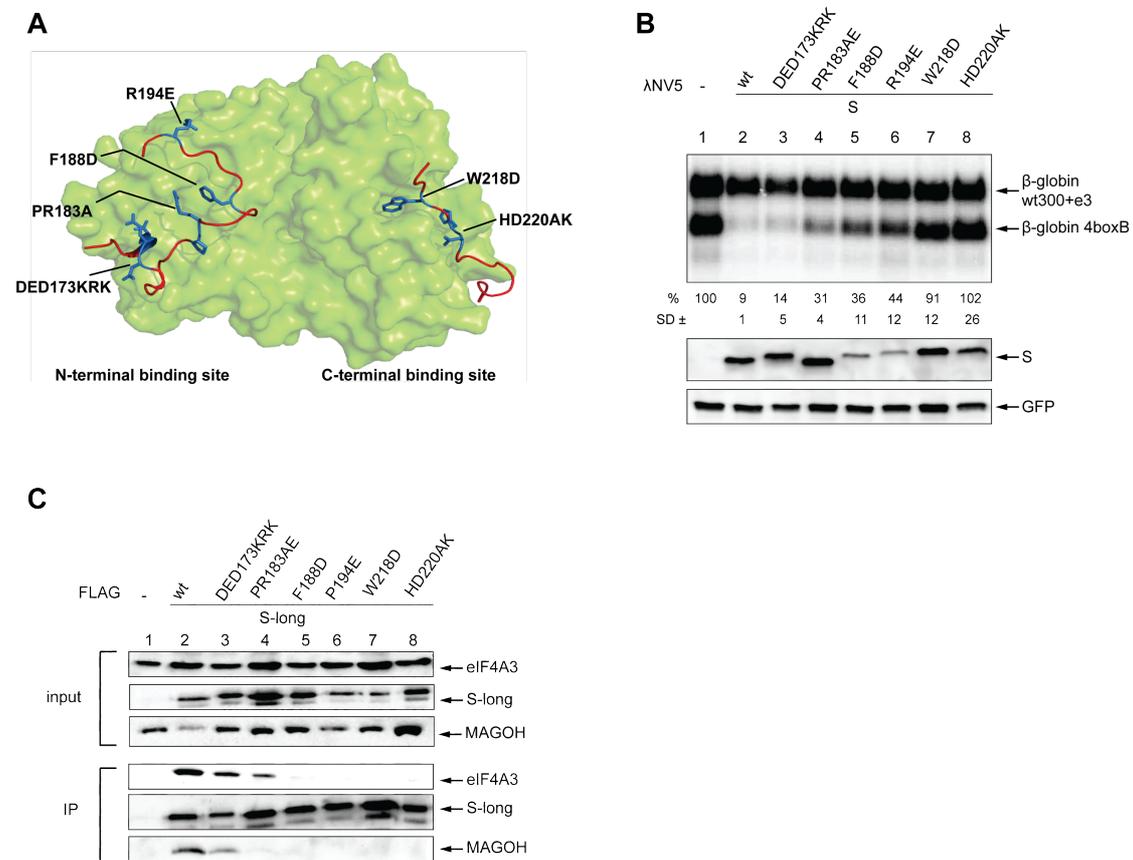
Figure 11. SELOR point mutations affect NMD function. (A) Crystal structure of eIF4A3 (green) and the SELOR domain (red), with SELOR mutations indicated (blue). The crystal structure of the EJC complex was modeled with PyMOL and the PDB file 2J0S, as in figure 3. (B) Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B. (C) Input and immunoprecipitation (IP) of S-long and mutants from Flp-In T-REx 293 cell lines. Anti-FLAG antibody was used to detect S-long expression and IP, endogenous eIF4A3 and MAGOH were detected with specific antibodies.

In addition, I observed that S-long F188D and S-long W218D have a different nuclear localization compared to the wt protein. S-long localizes to the nuclear speckles, similar to the other EJC components (Holzmann et al., 2000; Schmidt et al., 2009). The mutants, on the other hand, despite being in the nucleus, have lost their speckle localization, but are located diffusely in the entire nucleus (Supplementary figure S3). This phenotype is most probably the consequence of the inability of the mutants to bind the EJC (Figure 11C). Taken together, these results show that BTZ induced mRNA reduction upon tethering is directly proportional to BTZ capability to form a completely assembled EJC with the C-terminal binding site of SELOR playing a major role for this binding. A similar connection between mRNA degradation and EJC assembly was reported for mutations in MAGOH and Y14 that affect their ability to be assembled in the EJC. Mutations, which abolished binding to the EJC, also resulted in inefficient mRNA degradation in the tethering assay (Gehring et al., 2005; Gehring et al., 2009a). Therefore, the reduction of mRNA levels upon SELOR tethering can be attributed to recruitment of EJCs to positions downstream of the stop codon, which ultimately leads to NMD upon translation termination.

4.1.4 EJC binding and NMD inducing domains influence RNA decay

Hitherto it was shown that the N-terminal region and the SELOR domain in isolation can induce NMD and that the SELOR domain requires the interaction with the EJC to induce NMD. I decided to study the effect of these two regions in combination. Two mutations (F188D and W218D) were selected for further studies, one for each binding site of the SELOR-eIF4A3 interaction. Flp-In T-REx 293 cell lines expressing FLAG-tagged BTZ FL wt and the two mutants were established. Similarly to what was observed for the SELOR domain, the point mutations impaired the binding to the EJC (Figure 12A). Interestingly, when BTZ FL F188D and W218D were tethered to the reporter β-globin 4boxB mRNA, the mRNA was degraded to the same extent as with wt BTZ (Figure 12B). It is conceivable that in order to induce the reporter degradation, the full-length BTZ does not require the EJC binding and that the N-terminus functionally

compensates the inability of the protein to be assembled in the EJC. To validate this hypothesis, the N-terminus was added to the SELOR containing construct, which alone strictly requires proper EJC assembly for NMD activation, and BTZ NS wt and mutants were tethered to the mRNA (Figure 12C).
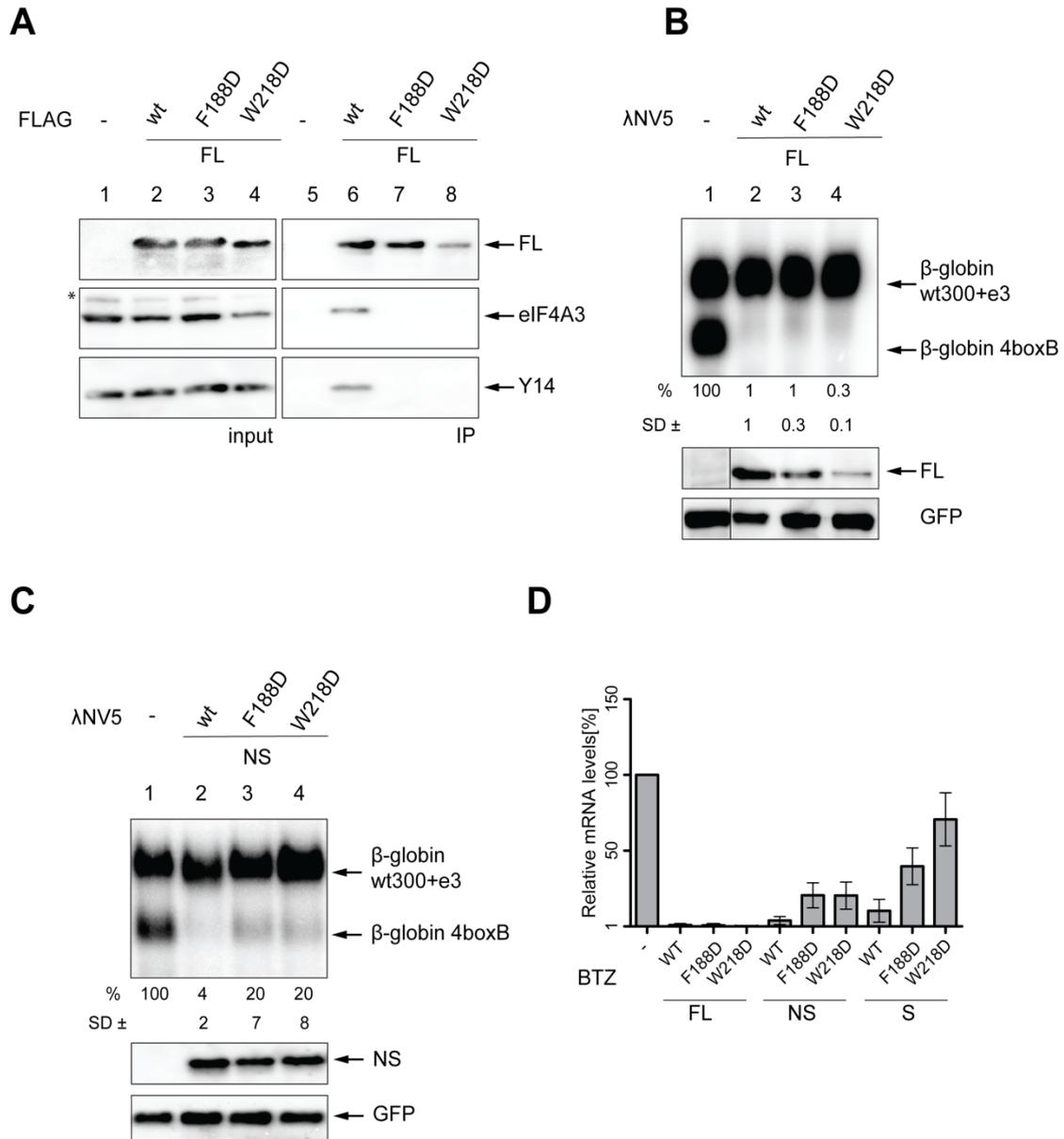
Figure 12. NMD dependence upon number of NMD inducing domains and EJC binding. (A) WB analyses of full-length BTZ as in Figure 11C. The asterisk indicates an unspecific band. (B, C) Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B. (D) Quantification of mRNA level at the steady state level of figure 12B and C and of Figure 11C, normalized to the negative control λNV5 (100 %).

The steady state mRNA levels detected upon tethering of the NS mutants were elevated compared to tethering of the full-length protein, but still reduced compared to the same mutations in the isolated SELOR domain. Furthermore, I noticed that comparing the wt proteins, FL, NS and S, respectively, 1 %, 4 % and 9 % were the mRNA levels detected (Figure 12B, lane 2, Figure 12 C lane 2 and Figure 11B lane 2), suggesting that the NMD induction was increased when the tethered protein was longer. It is possible to conclude that the activation of the NMD by BTZ depends on two determinants: the presence of the NMD inducing domains, SELOR and N-terminal region, and the ability to associate with the other EJC components. The SELOR domain mediates this association, for this reason the mutants containing the N-terminus (FL and NS) are less sensitive than SELOR to the F188D and W218D mutations (Figure 12D).

4.1.5 Conservation of the SELOR domain in different species

The EJC core components are highly conserved even between distant species like Human, *Drosophila* and *C.elegans*. Two regions of BTZ are conserved between the three species: aa 173-192 and aa 218-244 (in the human sequence), comprising most of the SELOR domain. Indeed these are the same two stretches responsible for eIF4A3 binding (Bono et al., 2006) (Figure 13A). In human *Drosophila* and *C.elegans* NMD exists, but the molecular mechanism of NMD activation is different. Although in mammals NMD can be activated by the EJC, in *C.elegans* EJC is dispensable while contradicting results were reported for *Drosophila* (see paragraphs 2.3 and 2.5). To study if the three orthologues are functionally similar, I tethered human (aa 110-372), *Drosophila* (aa 64-270) and *C.elegans* (aa 36-203) SELOR to the β-globin 4 boxB reporter to observe the effect of these proteins in the human cells. Human SELOR served as positive control for EJC binding (Figure 13C, lane 2) and for inducing EJC-dependent NMD as showed in Figure 11. The *Drosophila* protein induced the degradation of the reporter RNA to levels comparable to tethered human SELOR (Figure 13B, compare lanes 2 and 3). Interestingly, the *Drosophila* SELOR protein was able to immunoprecipitate human eIF4A3 (Figure 13C, lane 3). On the contrary *C.elegans* SELOR was not able to induce mRNA reporter degradation (Figure 13B, lane 4) nor to immunoprecipitate human EJC components (Figure 13C, lane 4).

Although the SELOR domain is conserved in the three species, human and *Drosophila* SELOR are functionally more related, given that *Drosophila* SELOR can bind eIF4A3 and through it activate NMD.
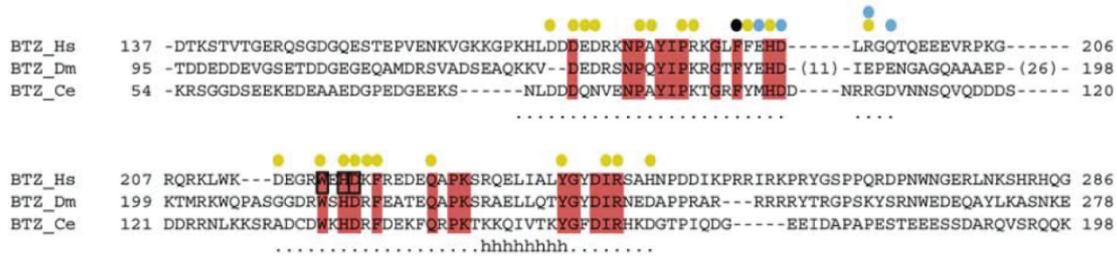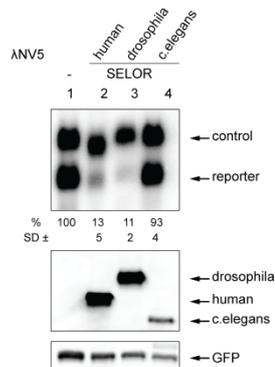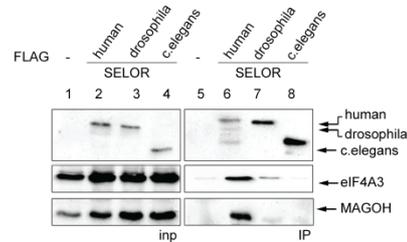
Figure 13. BTZ orthologues are not functionally identical (A) Alignment of Human (Hs), *D.melanogaster* (Ds) and *C.elegans* (Ce) BTZ. Conserved aa are marked in red. Above the alignment, yellow circles indicate residues required for eIF4A3 binding, black circle for RNA binding and blue circles for MAGO binding. Black squares indicate tested positions where mutations affect protein-protein interactions (Ballut et al., 2005). The figure is adapted from Bono et al., 2006. (B) Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B. (C) WB analyses of human, *D.melanogaster* and *C.elegans* SELOR, FLAG tagged transfected in HeLa cells. Endogenous eIF4A3 and MAGOH were detected with specific antibodies.

4.2 BTZ-induced NMD

The full-length BTZ induces NMD when tethered downstream of a stop codon. The isolated SELOR domain (S) and the small N-terminus (N) retain this function (Figure 9B). My aim is to identify the specific pathways activated upon tethering of BTZ and its deletion mutants.

4.2.1 BTZ domains induce NMD via different pathways

Heretofore I showed that NMD could be triggered by the SELOR domain and by the N-terminal region. When tethered to the mRNA, they activate a cascade of reactions that terminate with the degradation of the reporter mRNA. Starting from the tethered protein, two pathways have been described for the activation of NMD, BTZ-dependent and UPF2-dependent pathway. The first one is mediated by a fully assembled EJC, the second one is EJC-independent (Gehring et al., 2005)(see paragraph 2.4.2). I performed UPF2 knockdown to estabilish if SELOR and N-terminus NMD require UPF2 protein. In UPF2 knockdown cells, the UPF2 protein levels were reduced to less than 10%, as shown by immunoblotting (Figure 14A).
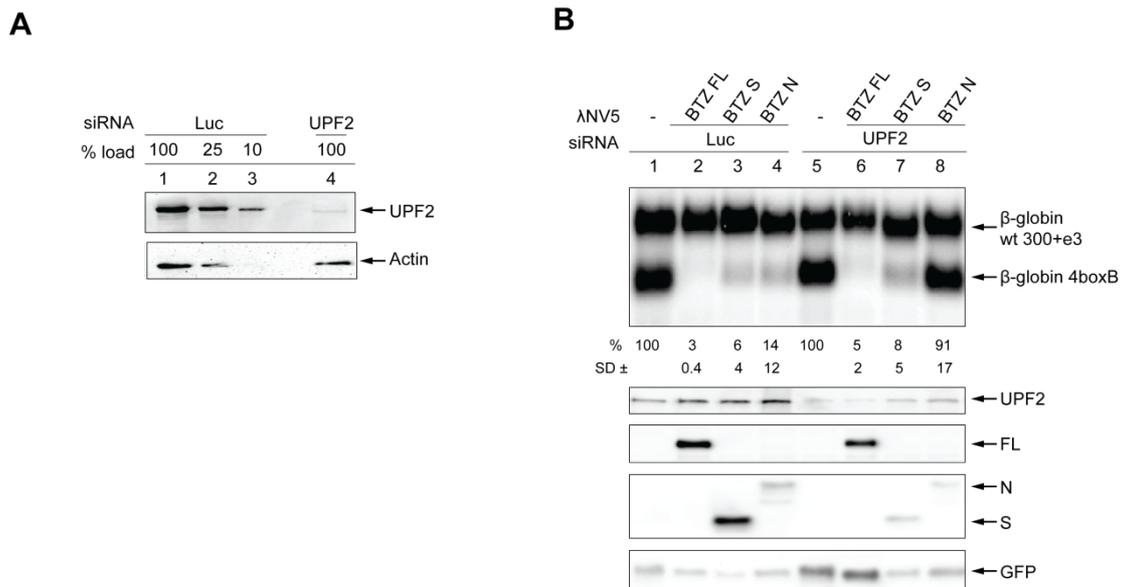
Figure 14. UPF2 (in)dependent NMD. (A) siRNA knockdown of UPF2 in human cell culture. HeLa cells were transfected with siRNAs targeting UPF2 or Luciferase (negative control). The knockdown efficiency was assessed by immunoblotting with an UPF2-specific antibody. Actin served as a loading control. (B) Northern blot (upper panel) of RNA samples extracted from HeLa cells transfected with the indicated siRNAs and constructs. Co-transfected β-globin wt300+e3 served as control mRNA and β-globin 4 boxB as reporter mRNA (n=3). Western blot analysis (bottom panel) was performed as in Figure 9B.

Upon control knockdown of Luciferase (Luc), the BTZ FL, as well as BTZ S and BTZ N stimulated degradation of the reporter mRNA comparable to earlier results without knockdown (compare figure 9B, lanes 2,6 and 9 with figure 14B, lanes 2-4). Upon depletion of UPF2, BTZ FL retained its ability to induce NMD in agreement with Gehring at al., 2005 (Figure 14B, lane 6). Interestingly, the knockdown of UPF2 did not influence the degradation of reporter mRNA by tethered BTZ S, while BTZ N completely lost its ability to induce NMD (Figure 14B, lanes 7 and 8). This finding suggests that the two NMD inducing domains, SELOR and N-terminus, can in isolation induce NMD via two independent pathways, one via the assembly of a functional EJC (SELOR), the other via UPF2 (BTZ N).

4.2.2 XRN1-dependent degradation

To define how BTZ S and BTZ N lead to the degradation of the reporter mRNA, I followed the cascade of reactions that they activate. Both pathways, EJC-dependent and UPF2-dependent rejoin at the step of UPF1 recruitment (Gehring et al., 2005). The events that follow UPF1 recruitment culminate in the degradation of the mRNA. SMG6 is responsible of the endocleavage at the termination codon, which is stimulated by the presence of EJCs or by tethered EJC components (see paragraph 2.3). Two fragments are generated, of which the 3′ fragment can be stabilized by depletion of the 5′-3′ endonuclease XRN1 (V Boehm, N Haberman, F Ottens, J Ule and NH Gehring, in press). Two different reporters were used to detect the specificity of the 3′ fragment formation: TPI-4MS2-SMG5 and TPI-SMG5. The triosephosphate isomerase (TPI) gene was modified by the addition of 1037 nt of the SMG5 3′UTR (such long sequence increase the detection of the 3′ fragment, data not shown). 4 copies of a 100 nt sequence of β-globin third exon and 3′ UTR were included to enhance the detection by the northern blotting. The 4MS2 binding sites in TPI-4MS2-SMG5 enable to distinguish between the specific effect of the tethered protein and the unspecific effect due to the protein overexpression (V Boehm, N Haberman, F Ottens, J Ule and NH Gehring, in press) (Figure 15A). In XRN1 knockdown cells, the XRN1 protein levels were reduced to less than 10%, as shown by immunoblotting (Figure 15B). In Luc KD cells, a reduction of the reporter (TPI-

4MS2-SMG5) mRNA was observed for all three tethered variants (Figure 15C, lanes 2-4). In XRN1 KD cells when BTZ full-length, BTZ S and BTZ N were tethered, the 3' fragment generated from the TPI-4MS2-SMG5 reporter construct specifically appeared in comparison to the control (Figure 15C, lanes 5-8).
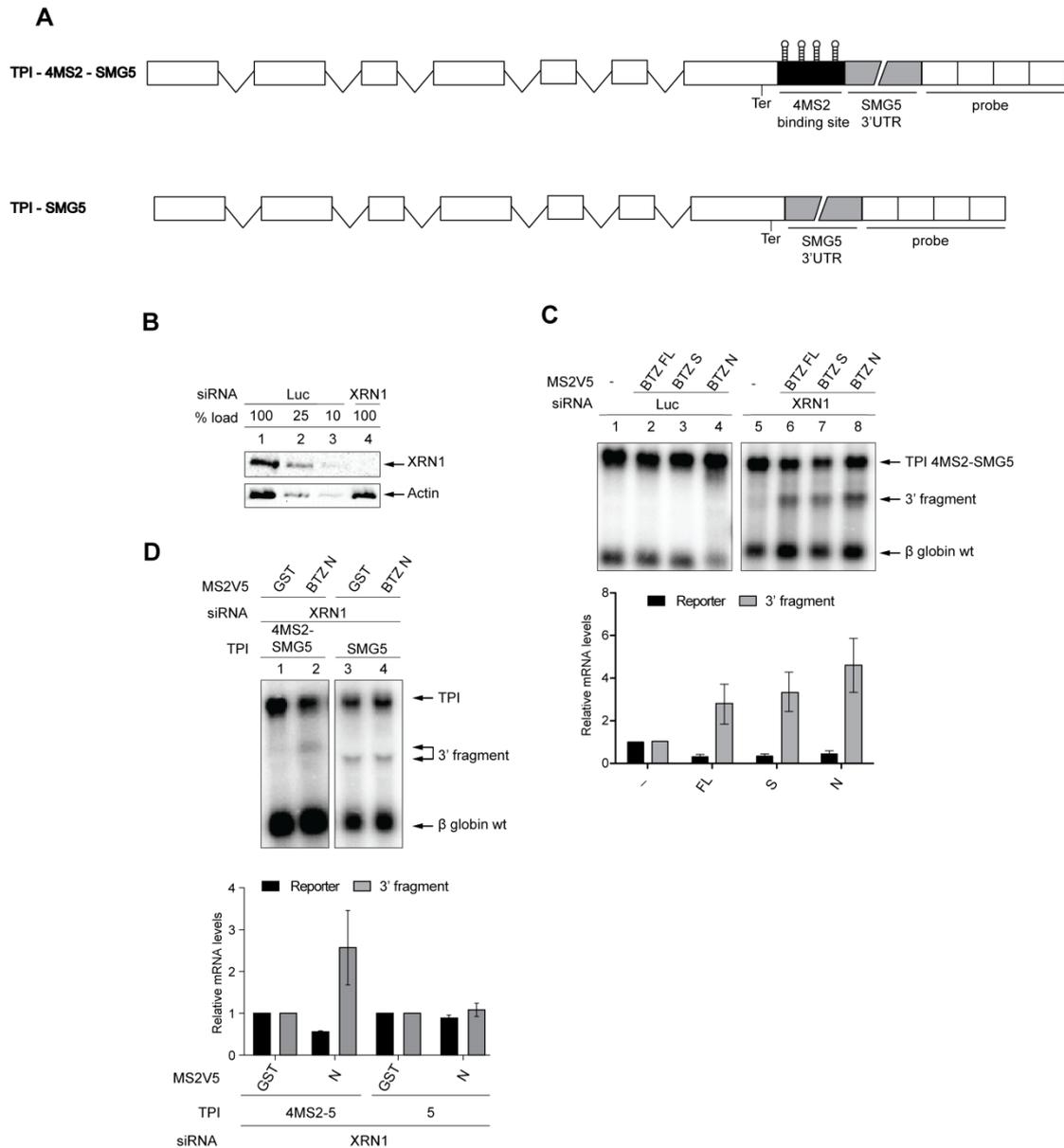
Figure 15. XRN1 dependent 3' fragment formation. (A) Schematic representation of the triosephosphate isomerase (TPI) reporter constructs. Exons are depict as white boxes and the connecting introns as black lines. Northern probe binding sites as white boxes without intron lines. MS2-binding is shown as a black box and SMG5 3' UTR is depicted as a gray box (B) Same as in figure 14A for siRNA targeting XRN1. (C-D) Northern blot (upper panel) of RNA samples extracted from HeLa cells transfected with the indicated siRNAs and constructs. Co-transfected β-globin wt served as control mRNA and TPI 4MS2 as reporter mRNA. The position of the 3' fragment is indicated (n=3).

It was previously shown that tethered EJC components could lead to endocleavage at the termination codon and the generation of 3′ fragments (V Boehm, N Haberman, F Ottens, J Ule and NH Gehring, in press). The 3′ fragment generation upon BTZ N tethering was surprising considering that this mutant does require other EJC proteins to induce NMD (Figure 14B). To confirm that the generation of the 3′ fragment is not an unspecific consequence of the over expression of BTZ N, two different reporters, TPI-4MS2-SMG5 and TPI-4MS2, were compared, of which TPI-4MS2 does not contain the tethering site for the protein (Figure 15A). The effect of BTZ N is indeed specific, because no 3′ fragment generation was detected with TPI-4MS2 (Figure 15D). Taken together, my results show that BTZ, as well as the SELOR domain, induces NMD in an UPF2-independent manner. In contrast, the N-terminus requires UPF2 to elicit NMD (Figure 14B). Altough the specific NMD activating pathways are different, the involvement of UPF1 and the recruitment of the SURF complex is common to both (Kashima et al., 2006), leading to the degradation of the substrate mRNA (Figure 15C).

4.3 BTZ binding to the RNA

In addition to the EJC-interacting function, the SELOR domain of BTZ has been reported to directly bind the mRNA. BTZ has been originally identified as component of mRNPs *in vivo*. Additionally it was shown that *in vitro* translated BTZ bound selectively poly G and poly U oligonucleotides immobilized on agarose beads (Degot et al., 2004). Moreover Ballut and colleagues showed in 2005 that *in vitro* purified SELOR domain could be crosslinked to a synthetic ^{32}P end-labeled 44-mer RNA. However *in vivo*, BTZ RNA interactome has not yet been defined. Interestingly would be the identification of the RNA binding partnerns of the wt and EJC impaired proteins, in order to distinguish between the EJC-dependent and –independent interactome.

4.3.1 *in vitro* RNA binding

In order to validate BTZ binding to the RNA *in vitro,* I performed electrophoretic mobility shift assays (EMSA). This technique allows estimating the ratio of unbound and bound protein to a specific nucleic acid probe. For initial experiments and in order to establish a positive control I used a recombinant protein comprising RRM 1-4 of PABPC (cytoplasmic poly A binding protein) (Deo et al., 1999) purified from E.coli. γ^{32}P-ATP phosphorylated RNA oligo (A25) was used as probe. Increasing concentrations of proteins were incubated with constant amount of RNA. In the absence of the purified protein or at low concentrations of the PABPC protein only the free probe was detected at the bottom of the gel (Figure 16A, lane 1 and lanes 2-4). A partial shift of the RNA band was observed at the highest protein concentration (Figure 16A, lanes 5). To study the affinity of BTZ to the RNA, S-long was purified from bacteria and two different RNA oligos were used: A25 and U27. A25 was used as positive control for PABPC binding, whereas U27 was previously shown to be bound by BTZ (Degot et al., 2004).
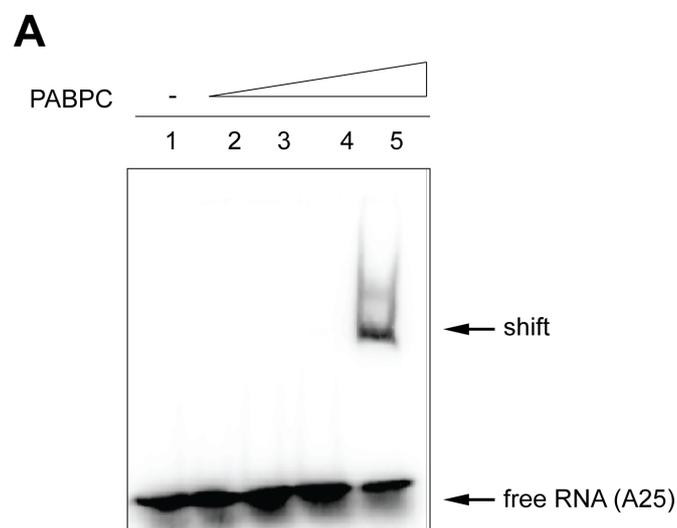
Figure 16. PABPC binding to A25 RNA. (A) 5nM of γ^{32}P-ATP phosphorylated A25 RNA were incubated with increasing amount of purified PABPC (0, 10, 50, 100, 1000 nM). Electrophoretic shift was observed at the highest protein concentration.

Taking into consideration that in the initial experiment with PABPC, only with 1000 nM of the protein a partial shift of the RNA band was observed, the subsequent experiments were performed with up to 10000 nM concentration of the purified proteins, PABPC was used as positive control for the binding to the A25 RNA (Figure 17A, lanes 5-8). At a concentration of 1000 nM of PABPC, a weak shift of the RNA probe was observed, whereas most of the free probe ran at the bottom of the gel (Figure 17A, lane 7). However, at a concentration of 10000 nM of PABPC, the entire free probe was supershifted to the top of the gel (Figure 17A, lane 8). When S-long was incubated with the A25 probe, no shift was observed at the concentration of 1000 nM protein but the free probe band was reduced compared to the negative control (Figure 17A, compare lane 3 with lane 1). However, at the concentration of 10000 nM, a weak shifted band at the top of the gel, but no free probe at the bottom of the gel were detected (Figure 17A, lane 4). The experiment was repeated with an U27 RNA oligo. Compared to the A25 probe, S-long bound the U27 probe with higher affinity, confirming earlier observations (Degot et al., 2004). A shift was observed at 1000 nM and a supershift at 10000 nM (Figure 17B, lanes 3 and 4).
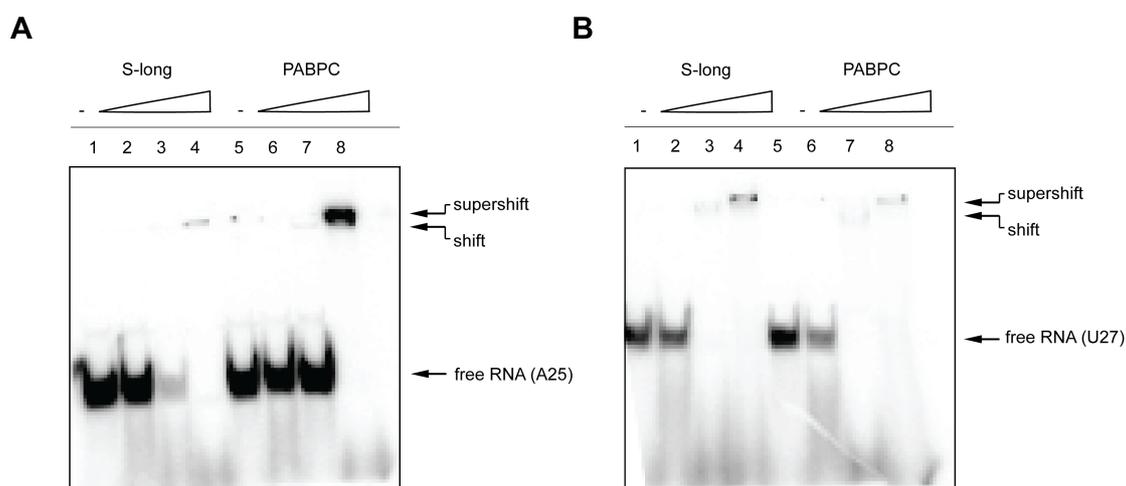
Figure 17. S-long and PABP binding to the RNA. (A) 5nM of A25 RNA were incubated with increasing amount of purified S-long and PABP (0, 100,1000, 10000 nM). (B) Same as in A, for γ^{32}P-ATP phosphorylated U27 RNA.

In contrast, PABPC interacted with the U27 RNA probe, with much weaker affinity compared to the A25 probe (compare figure 17B, lanes 7 and 8, with figure 17A, lanes 7 and 8). Of note, in both experiments (Figure 17A, lanes 3 and 4 and figure 17B lanes 3, 4, 7 and 8), the amount of total probe (free + bound) was reduced at high concentration of the proteins compared to the corresponding negative control, but this phenomenon would require a more in depth analyses. The supershift observed with S-long and the U27 RNA might derive from the assembly of a high molecular weight complex formed of multiple proteins and RNAs. To validate this hypothesis the experiment was performed incubating S-long with either A25 or U27 RNA probe. Each condition was performed in duplicate, so that after the incubation one half of the sample was loaded on a EMSA native gel (Figure 18A, upper panel), while the other half was UV crosslinked at 254 nm and loaded on a 12% SDS gel (Figure 18, bottom panel). When S-long was incubated with U27 and then crosslinked to the probe, the formation of two protein-RNA complexes with different molecular weight was observed.
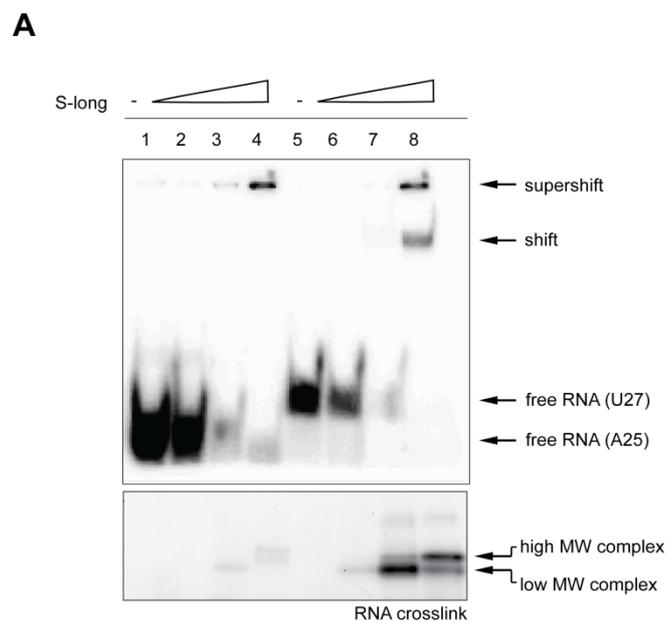
Figure 18. S-long forms a super complex with the RNA probe (A) 5nM of RNA probe were incubated with increasing amount of purified S-long (0, 100,1000, 10000 nM) and loaded on a native gel (upper panel) or UV crosslinked at 254nm and loaded on a 12% SDS gel (bottom panel).

The ratio between these complexes changed according to the protein concentration: at 10000 nM of S-long the higher molecular weight (MW) complex was more prominent (Figure 18A, compare lane7 with 8). This phenomenon was clearly visible when S-long was incubated with U27, but a weaker signal was detected with A25 as well (Figure 18A, lanes 3 and 4). Summarizing the results obtained so far, S-long showed differential and sequence specific binding affinities to the RNAs utilized, with preferential binding of poly(U) over poly(A) (Figure 17A and B). This indicates that BTZ is not an unspecific RNA binding protein, but selectively binds certain mRNA. Moreover I could show that S-long forms high molecular weight complexes in which more proteins and/or RNAs are involved (Figure 18A), indicating a possible multimerization of BTZ on the mRNA.

4.3.2 *in vivo* RNA binding

My results so far raise the question how the NMD activation by BTZ is regulated in living cells. I have observed that *in vitro* S-long shows a preferential binding for certain mRNAs and previous reports suggested that BTZ can directly interact with RNA *in vitro* in the absence of any other EJC component (Degot et al., 2004). To gain insight into RNA binding by BTZ *in vivo*, I used Flp-In T-REx 293 cell lines, which express the S-long protein as wildtype or as EJC-binding deficient mutants F188D and W218D. I performed crosslinking and immunoprecipitation to RNAs (CLIP) in living cells as readout for RNA binding. The CLIP allows to identify RNA sites in direct contact with a RNA binding protein (Ule et al., 2005). Following the protocol, a FLAG-immunoprecipitation was performed on the UV crosslinked cell lysate, thereby the RNA bound to the FLAG tagged protein will be co-immunoprecipitated, γ^{32}P-ATP phosphorylated and subsequently detected by SDS-PAGE. As negative control a cell line expressing non-fused FLAG was used. In the control cells no signal was observed after crosslinking and IP, confirming the specificity of the approach (Figure 19A, lane 1 and 2). A strong signal was detected for S-long wt (Figure 19A, upper panel, lane 4), whereas the signals of W218D and F188D mutants were markedly reduced (Figure 19A, upper panel, compare lanes 6 and 8 with 4). Hence, the reduced ability of both mutants to

bind to RNA in living cells suggests that S-long interacts with RNAs mainly as part of the EJC. I aimed to confirm these results with full-length BTZ. After a short UV exposure BTZ was very inefficiently crosslinked to the RNA (Figure 19B, lanes 2-4) when compared to the amount of crosslinked S-long (Figure 19A, upper panel, lane 4). In addition, prolonged UV irradiation, aimed to increase the crosslinking efficiency, resulted instead in a reduced crosslinking efficiency (Figure 19B, upper panel, lanes 5 and 6) probably due to destabilization of the protein (Figure 19B, bottom panel, lanes 5 and 6).
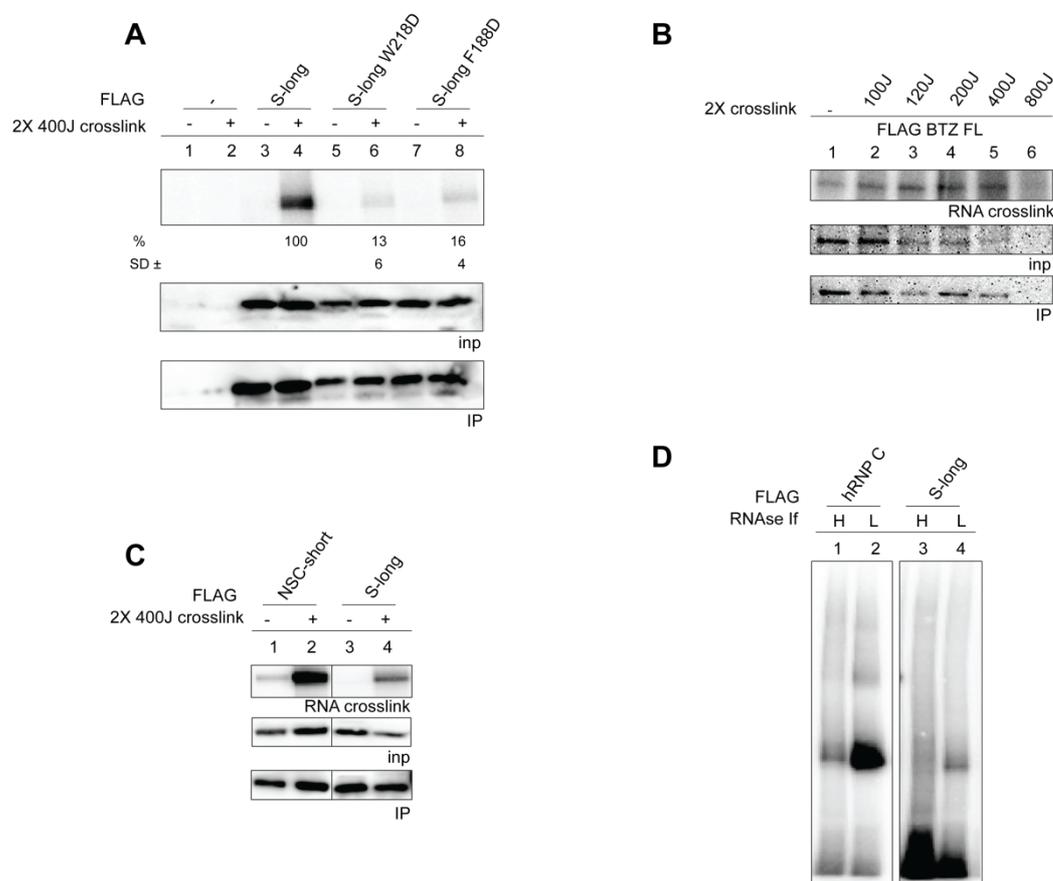
Figure 19. BTZ *in vivo* binding to the RNA (A) Flp-In T-REx 293 cell lines expressing FLAG, S-long, W218D and F188D were subjected to UV crosslinking (+) or not (-). mRNA was labeled with γ^{32}P-ATP (upper panel). FLAG-tagged S-long was detected with an anti-FLAG antibody by immunoblotting prior to labeling (bottom panel). (B) Same as in A for FLAG-tagged full-length BTZ, increasing UV crosslinking intensities are shown. (C) Same as in A for NSC-short and S-long (D) Flp-In T-REx 293 cell lines expressing FLAG, S-long and hn RNP C were subjected to UV crosslinking. Two different RNA dilutions were used, 1:50 (L) and 1:500 (H).

Nevertheless, the strong RNA crosslink of the C-terminally truncated BTZ (NSC-short) and the shorter S-long mutant (Figure 19C, lane 2 and 4) suggested that the C-terminal part of BTZ prevented its efficient crosslinking to RNA in living cells. Considering that S-long could be more efficiently crosslinked than full-length BTZ, I decided to use it for mapping the transcriptome wide binding sites. However several attempts to obtain a library of mRNA bound by S-long failed. The RNA-binding protein hnRNP C (Huang et al., 1994) has been previously used in iCLIP experiments due to its strong affinity for RNA (Huppertz et al., 2014). Comparing the binding of hnRNP C and S-long to the RNA, I observed that S-long was less efficiently crosslinked to the mRNA (Figure 19 D). Such limited RNA binding is most probably the reason why it was not possible to identify the RNA bound by BTZ.

4.4 BTZ protein interactome

To identify specific interaction partners of BTZ and in particular of the SELOR domain, I performed a mass spectrometry analysis (MS) of FLAG-tagged full-length BTZ and S-long in Flp-In T-REx 293 cell lines. I compared FLAG-BTZ and FLAG-S-long with their corresponding W218D mutants (Figures 11C and 12A) to specifically identify EJC-dependent interaction partners of both proteins. In the MS IP, each single protein was identified with a certain number of peptides and a specific concentration. The proteins were selected according to the relative fold increase (the ratio between the concentration of a certain protein in the wt and the mutant IP) and only the proteins that were identified with more than one peptide were considered. Among the most highly enriched proteins in S-long I found all EJC core proteins (MAGOH, Y14, eIF4A3), confirming the specificity of the approach (Fig. 20A). Furthermore, several of the previously identified peripheral EJC proteins, many spliceosomal proteins, such as components of the nineteen complex (NTC) or the U2 snRNP, and several members of the SR protein family were enriched in the interactome of FLAG-S-long (Fig. 20A). In order to define the functions of the enriched proteins, I performed a functional annotation search using DAVID (Huang da et al., 2009a, b), considering only the proteins that were enriched more than two fold. Figure 20B shows the result

obtained for the S-long interactome, while similar results were obtained for the BTZ FL interactome. Most of the interacting partners are involved in processes, for which BTZ was already described to be part of, such as transcription, splicing, spliceosome assembly, RNA processing or NMD (Ciriello S, Boehm V. and Gehring N.H., in preparation).

A

Name	Unique peptides	Increase
PININ	27	∞[1]
ACINUS	11	∞[1]
RNPS1	3	∞[1]
RBM8	8	32
MAGOH	7	13.63
eIF4A3	21	7.35
SAP18	4	6.89

Name	Unique peptides	Increase
UPF3B	5	∞[1]
UPF2	4	∞[1]

Name	Unique peptides	Increase
UPF3B	5	∞[1]
TDRD3	3	∞[1]

Name	Unique peptides	Increase
SF3B3	2	2.5
SF3B5	2	2.36
SF3B1	21	1.75
SF3B14	3	0.93

Name	Unique peptides	Increase
PRPF19	6	8.18
CDC5L	6	∞[1]

Name	Unique peptides	Increase
SRSF9	2	∞[1]
SRSF6	4	∞[1]
SRSF3	6	20.8
SRSF1	3	6.15
SRSF7	2	0.6

B

Figure 20. Selected S-long binding partners. (A) Selection of proteins interacting with S-long as identified by mass spectrometry. Increase denotes the enrichment in wt versus mutant (218) complexes. Different categories of proteins are shown in different colors (EJC component: yellow; NMD factors: green; EBM-containing proteins: blue; U2snRNP components: pink; NTC complex: orange; SR proteins: red) [1]These proteins were identified only in S-long wt, but not in the mutant 218. (B) Functions of the proteins enriched more then two times in S-long IP.

4.4.1 SRSF proteins contribute to NMD

In order to assess the functional importance of these interactomes, I further characterized the serine/arginine-rich (SR) proteins that were found to be enriched in S-long interactome. SRSF proteins contain in their N-terminal domain one or two RRMs, with which they bind to mRNA (Dreyfuss et al., 2002; Singh and Valcarcel, 2005) and a C-terminal domain rich in serines and arginines (Long and Caceres, 2009). They are loaded on the mRNA cotranscriptionally and exported with the mRNA to the cytoplasm; in particular SRSF1, SRSF3 and SRSF7 interact with mRNA export factors and promote transport of mRNAs to the cytoplasm (Huang and Steitz, 2005). Moreover they are known regulators of the splicing process (Zhou and Fu, 2013) and the overexpression of SRSF1 contributes to NMD, enhancing the degradation of PTC expressing reporter mRNA (Zhang and Krainer, 2004). Singh and colleagues (Singh et al., 2012) proposed that endogenous EJCs multimerize with each another and with SRSF proteins to form megadalton sized complexes in which SRSF proteins are stoichiometric (SRSF9, SRSF10, SRSF12) or super-stoichiometric (SRSF1, SRSF3 and SRSF7) to EJC core factors (multimeric EJCs). In this complex the EJC and the SRSFs collaborate to form higher-order mRNP structures, which may facilitate overall mRNA packaging and compaction (Singh et al., 2012). In the immunprecipitation presented here, the core EJC components represented the most prominent bands on silver-stained gels of my immunoprecipitated samples and I could not detect any other proteins with the same intensity (Figure 21A). Therefore, the precise abundance and stochiometry of SRSF proteins in purified EJCs needs to be evaluated in future studies. Previously, SRSF1 has been studied in the context of NMD, whereas other members of the SRSF protein family have not been systematically analyzed. Therefore, I used the tethering assay to investigate the effects of SRSF1, 2, 3, 6 and 7. Variable levels of reporter mRNAs were detected upon tethering of different SRSF proteins. The strongest downregulation of the reporter mRNA was observed when SRSF1 or SRSF7 were tethered (Figure 20B, upper panel, lanes 2 and 6), despite SRSF7 being weakly expressed (Figure 20B, lane 6, bottom panel). A moderate downregulation of the reporter mRNA was elicited by SRSF2 and SRSF6 (Figure 20B, upper panel, lanes 3 and 5), the latter showing a very weak expression.

Figure 21. SRSF proteins contribute to NMD. (A) Silver staining of FLAG-immunoprecipitated FLAG and FLAG-S-long Flp-In T-REx 293 cell lines (B) Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B. (C) Northern blot of RNA samples and western blot analysis of proteins extracted from HeLa cells transfected with the indicated siRNAs and constructs as in Figure 14B. (D) Northern blot of RNA samples and western blot analysis of proteins extracted from HeLa cells transfected with the indicated siRNAs and constructs as in Figure 15C.

Tethering of SRSF3 had no detectable effect on reporter mRNA levels (Figure 20B, upper panel, lane 4). Because SRSF1 and SRSF7 showed the strongest effects, I continued my analysis with these two proteins. In order to define how their tethering is coupled with the reduction of the reporter mRNA levels, I performed a tethering assay in UPF2 depleted cells. The activity of both proteins was partially impaired by the depletion of UPF2 (Figure 21C), indicating that they require UPF2 in order to activate NMD. As further prove of their ability to induce NMD, they were tethered in XRN1 depleted cells. Similarly to BTZ (Figure 15C) and other EJC proteins (V Boehm, N Haberman, F Ottens, J Ule and NH Gehring, in press), more 3′ fragment is generated when the SRSF proteins were tethered (Figure 21D). In summary, I provide evidence that selected SRSF proteins can elicit NMD via an UPF2-dependent and endocleavage using pathway, but the molecular basis of this mechanism is unclear at the moment.

4.4.2 The NTC complex induces mRNA degradation in a NMD-independent way

I also noticed that several components of the NTC (PRPF19, CDC5L, BCAS2, PLRG1) were enriched in the S-long interactome. The NTC complex associates with the spliciosome in the complex B and remains bound until the termination of the splicing reaction. Moreover it is involved in genome manteinance, transcription elongation and recruitment of ubiquitylated proteins to the proteasome (Chanarat and Strasser, 2013; Hogg et al., 2010). Since I failed to express detectable amounts of PRPF19, I used the NTC components BCAS2 and PLRG1, which were highly enriched in the interactome, but not included in the final list as they were only detected by a single unique peptide. I also analyzed SNW1 (SKIP), a component of the NTC-related SKIP complex that is presumably involved in the activation step of the spliceosome (Bessonov et al., 2010). All three tested proteins reduced the levels of the reporter mRNA in the tethering assay (Figure 22A). However, in contrast to the SRSF proteins, this effect seems to be NMD-independent. The reduction of mRNA levels was UPF2-independent (Figure 22 B, C and D) and there was no 3′ fragment generation upon their tethering in XRN1 depleted cells (Figure 22E).
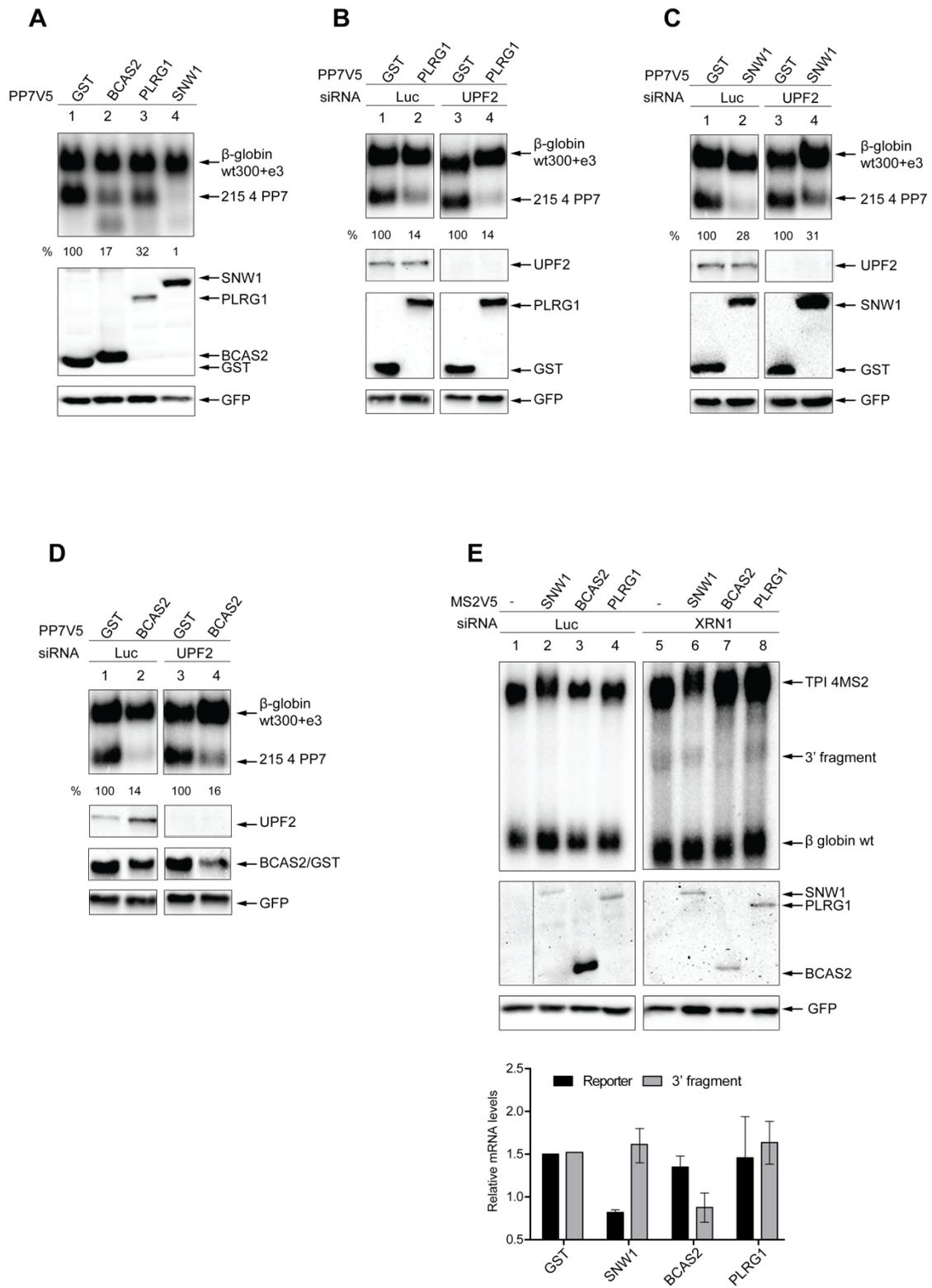
Figure 22. NTC complex proteins and SNW1 do not activate NMD (A) Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B. (B,C,D) Northern blots of RNA samples and western blot analysis of proteins extracted from HeLa cells transfected with the indicated siRNAs and constructs as in Figure 14B. (E) Northern blots of RNA samples and western blot analysis of proteins extracted from HeLa cells transfected with the indicated siRNAs and constructs as in Figure 15C.

5. Discussion

The EJC is deposited on the mRNA as a consequence of splicing, where it marks the exon-exon junctions (Le Hir et al., 2000a). The pre-EJC (eIF4A3, MAGOH and Y14) is assembled in the nucleus and then is exported to the cytoplasm. The compartment, in which BTZ binds the complex, is subject of ongoing discussion. According to the current model, BTZ joins the complex in the cytoplasm (Gehring et al., 2009b; Steckelberg et al., 2012), where it localizes at the steady state level. BTZ and the spliceosome were reported to bind the same eIF4A3 residues. Moreover, BTZ was not identified in purified spiceosomal complexes (Bessonov et al., 2008; Gehring et al., 2009a). This suggests that BTZ joins the complex after the completion of the splicing reaction, but does not exclude the possibility that such joining happens in the nucleus. BTZ is localized to the cytoplasm, even though it contains two functional NLSs, which mediate its localization to the nucleus when nuclear export is inhibited by Leptomycin B treatment. Even the isolated C-terminal domain with its NES, localizes to the nucleus after Leptomycin B treatment (Degot et al., 2004). Furthermore, Tange and colleagues proposed in 2004 that BTZ is present in the nucleus in low abundance and this might be a control mechanism of the cell to avoid premature EJC assembly.

5.1 Separate functions of the three BTZ regions

BTZ can be divided in three functional regions: N-terminus, SELOR domain and C-terminus. In this thesis, I showed that they are functionally independent from each other and I reported previously undescribed functions for these regions.

5.1.1 BTZ induced NMD

NMD inducing regions

Upon tethering of the N-terminal region and the SELOR domain of BTZ to the reporter mRNA, I observed a reduction of the mRNA levels at the steady state level. Interestingly, after transfection of two different constructs (BTZ N and BTZ S) the levels of reporter mRNA detected were in the same order of magnitude

(Figure 9B, lane 5 and 9). It has been reported that the EJC components tethered downstream of the stop codon activate NMD (Gehring et al., 2009a). The SELOR domain is the region of BTZ responsible for mediating binding to the EJC via eIF4A3 (Ballut et al., 2005; Degot et al., 2004). Therefore, SELOR-containing BTZ mutants tethered to the mRNA can function as a binding platform for the EJC proteins and activate NMD. In contrast, the N-terminal region was only studied by Gehring and colleagues in 2009, but not as isolated region. They observed that removing the N-terminal region decreased the ability of BTZ to induce NMD. Here, I studied this region in isolation and showed that it can indeed activate NMD to the same extend as the SELOR domain. This was a surprising result considering that the N-terminus is not involved in EJC binding, suggesting an EJC-independent way of NMD activation. Moreover, Gehring and colleagues observed in 2009 that mutated full-length BTZ proteins, which were unable to interact with the EJC, were still completely functional in the tethering assay. They concluded that another region of the protein beside the SELOR domain had to be responsible for NMD activation. Here, I showed that this region is the N-terminus. When the same mutations were inserted in the isolated SELOR domain the ability to induce NMD gave different results for different SELOR mutants. The constructs with mutations in the SELOR C-terminal binding site (i.e. W218D and HD220AK) were strongly impaired and inefficient in NMD activation. The mutants in the SELOR N-terminal binding site were still functional instead (i.e. DED173KRK), or partially affected by the mutations (i.e. PR183AE, F188D and R194D). The correlation between NMD activation and EJC binding was confirmed in the tethering assay by the unique functional mutant (DED173KRK) which was the only one able to immunoprecipitate all the other EJC components (Figure 11 B and C). In good agreement with these observations, mutations in the SELOR C-terminal binding site have a stronger impact on the function of the protein than the mutations in the SELOR N-terminal binding site. Indeed, the C-terminal binding site is required for binding of SELOR and eIF4A3 *in vivo* and *in vitro*. Mutations in the N-terminal binding site, despite disrupting the interaction *in vivo*, do not affect the *in vitro* binding (Bono et al., 2006). So far, it is possible to conclude that in order to induce NMD the SELOR domain necessitates the binding with the EJC proteins. In contrast, in the context of BTZ FL, mutants that

were unable to assemble an EJC could efficiently induce NMD as well (Figure 12 A and B). Similarly, also BTZ NS (containing both SELOR domain and the N-terminal region) EJC-impaired mutants (F188D and W218D) activated NMD (Figure 12C). This NMD functionality can be explained by the central role of the N-terminal domain during activation of NMD. Indeed, the effect of the point mutations was more severe in the isolated SELOR domain than in BTZ NS and BTZ FL. The N-terminal region could activate the EJC-independent NMD and, hence, was dominant over the non-fuctional SELOR domain. Furthermore, comparison among wt BTZ FL, NS and S revealed that the efficiency of NMD in the S protein was lower than the FL and NS protein. These data suggest that the mRNA degradation efficiency correlates with the number of NMD-inducing regions (Figure 12D).

SELOR and N-terminal region activate different NMD pathways

To identify the pathways activated by tethering BTZ N and BTZ S, I performed RNA silencing of the proteins UPF2 and XRN1. UPF2 KD allows to distinguish between UPF2-dependent and -independent NMD (Gehring et al., 2005). UPF2-independent NMD involves the direct interaction of the EJC with UPF1, whereas UPF2 is dispensable to mediate this binding and to activate NMD. In fact, UPF1 mutants lacking the interaction with UPF2 (Ivanov et al., 2008) can still immunoprecipitate BTZ (Gehring et al., 2009a). BTZ S similarly to BTZ FL was not influenced by UPF2 depletion. Indeed, BTZ S and BTZ FL could assemble an EJC (Figure 11C and 12A) and consequently activate NMD independentely of UPF2. On the contrary, BTZ N could not induce degradation of the reporter mRNA in UPF2 depleted cells (Figure 14B), suggesting an UPF2-depedent NMD activation. The UPF2-dependent pathway does not necessitate a fully assembled EJC and indeed BTZ N is not the region responsible for EJC assembly. Both degradation pathways rejoin at the step of UPF1 involvement (Gehring et al., 2005), leading to the recruitment of SMG6. To confirm that BTZ N, BTZ S and BTZ FL shared the same NMD pathway once UPF1 is activated, I evaluated mRNA endocleavage executed by SMG6. Two fragments are generated by this cleavage (Eberle et al., 2009; Huntzinger et al., 2008), of which the 3′ fragment can be stabilized by XRN1 KD (V Boehm, N Haberman, F Ottens, J Ule and NH Gehring, in

press). Upon tethering BTZ FL, BTZ S and BTZ N in XRN1 depleted cells, the 3′ fragment generated from the TPI-4MS2-SMG5 reporter construct is stabilized in comparison to the control (Figure 15C). This result is additional proof that BTZ tethering elicits bona fide NMD and suggests that the two different pathways activated by BTZ FL, BTZ S and BTZ N rejoin at the step of reporter mRNA degradation via XRN1.

5.1.2 BTZ induced polyadenylation

The C-terminal region of BTZ tethered to the reporter mRNA showed a different effect compared to the N-terminal region and the SELOR domain. Two bands, specific products of the β-globin 4boxB reporter mRNA, were detected in the northern blot, when BTZ C was tethered to the mRNA. These two bands showed faster and slower motility on the northern gel, respectively, when compared to the negative control (Figure 9B and 10A). BTZ C, however, did not induced degradation of the reporter mRNA. When the intensities of both bands were added, 66% of the reporter mRNA was detected at the steady state level (Figure 9B). Cloning and sequencing of both bands showed that the faster migrating band lacked the 4boxB binding site and was polyadenylated downstream of a cryptic polyadenylation signal AUUAAA (β-globin short) (Figure 10B). Analogously, when protein components of the CSTF complex were tethered to the reporter, a fast running band was detected (Figure 10C). The CSTF complex binds downstream of the polyadenylation signal and induce cleavage and polyadenylation upstream of its position (Chan et al., 2011; Elkon et al., 2013). I suggest that BTZ C and the CSTF proteins activate usage of the cryptic polyadenylation signal situated upstream of the 4boxB binding sequence, leading to cleavage and polyadenylation at this site (Figure 10). Degot and colleagues showed in 2004 that the NES of BTZ is functional and responsible for the localization of the protein to the cytoplasm. The NES corresponds to the aa 462-472 included in BTZ C, which indeed localizes to the cytoplasm (data not shown). However, polyadenylation activated upon BTZ C tethering is unlikely to occur in the cytoplasm. Cytoplasmic polyadenylation is determined by the cytoplasmic polyadenylation element (CPE; consensus sequence UUUUUAU) that resides

within 100 bases 5′ of the AAUAAA poly(A) signal. The protein CBPE (CPE binding protein) binds this sequence in the nucleus and is then exported with the mRNA to the cytoplasm where it contributes to elongate the short poly (A) tail already added to the mRNA in the nucleus (Ivshina et al., 2014; Lin et al., 2012). The β-globin 4boxB reporter does not contain such a sequence, so activation of CBPE via BTZ seems rather unlikely. On the other hand, BTZ C might be able to shuttle into the nucleus, as was proposed for the full-length BTZ (Macchi et al., 2003), where it could be involved in nuclear polyadenylation. More in depth analyses will be necessary to define the role of BTZ C in polyadenylation.

5.2 RNA binding

In vitro binding

In vitro the SELOR domain mediates binding of BTZ to RNA (Ballut et al., 2005; Degot et al., 2004). In agreement with previous results, I observed that S-long showed a preference for the U27 oligo compared to A25 (Figure 17) (Degot et al., 2004). However, I noted that S-long had much weaker binding to the RNA compared to PABPC (Figure 17). The differences in RNA affinity might reflect the structure and function of the two proteins. PABPC has four RRMs, of which RRM1 and RRM2 are mostly responsible for binding to the poly(A) sequence (Deo et al., 1999). On the other hand, the SELOR domain, despite being responsible for RNA recognition, does not contain any RNA recognition motif. The binding to the RNA was stronger, albeit still limited, when the probe was incubated with higher protein concentrations (Figure 17). The supershift observed might result from the assembly of multiple proteins and RNAs. To verify this hypothesis I crosslinked the samples after the incubation in the EMSA conditions. Two bands, designed as high and low molecular weight complex, were detected on the gel (Figure 18, bottom panel). At the highest protein concentration, the high molecular weight band was more prominent than the low molecular weight one (Figure 18, bottom panel, lane 8). The ratio was inverted at a reduced protein concentration (Figure 18, bottom panel, lane 7). Considering that the increase of the protein concentration led to a higher molecular weight complex, either more S-long proteins bind the same small RNA, or S-long is able to multimerize. The

hypothesis of multimerization is supported by recent results suggesting that EJC form a multimeric complex binding other EJCs and additional factors (Singh et al., 2012).

In vivo binding

The SELOR domain mediates binding of BTZ to the mRNA *in vivo* (Ballut et al., 2005; Degot et al., 2004). Here, I showed that this binding is EJC-dependent, whereas S-long W218D and F188D showed greatly reduced RNA binding compared to the wt S-long (Figure 19A). Similarly, concerning the NMD inducing ability of the SELOR domain (Figure 10), it seems that the SELOR domain requires being part of the EJC to exert its functions. To extend the identification of the RNA interacting with BTZ, not only via the SELOR domain, I planned to crosslink the full-length protein. The crosslinking efficiency of BTZ FL to the mRNA was unfortunately much reduced. Increasing UV intensities, aimed to improve the crosslinking efficiency, resulted in very limited mRNA binding, probably due to protein degradation (Figure 19B). When comparing the crosslinking efficiency of BTZ FL, S-long and NSC-short, it was possible to define the contribution of different BTZ region to RNA binding. BTZ NSC-short had a higher affinity for RNA compared to S-long (Figure 19C). This indicated a positive contribution of the N-terminal region to mRNA binding. Likewise, I observed that the N-terminal region enhanced NMD activity of the SELOR domain (Figure 12C). BTZ FL instead was weakly crosslinked to the mRNA, suggesting that the C-terminal domain had a negative effect on mRNA binding (Figure 19B). From these observations it is possible to argue that binding to the mRNA correlates with NMD function: while the NMD competent regions of BTZ (SELOR and N-terminus) favour this binding, the NMD incompetent region (C-terminus) has a negative effect. BTZ NSC-short showed the strongest RNA binding. However, a signal was also detected in the lane of the samples that were not crosslinked. This might be due to phosphorilation of the protein during incubation with the PNK mix. To circumvent this problem, I planned to identify the RNA bound to S-long. However, several attempts to obtain a library of mRNA bound by S-long failed. A possible reason for this is the weak mRNA affinity of S-long for the mRNA. Comparing the signal for hnRNP C and S-long, I observed that

S-long was less efficiently crosslinked to the mRNA (Figure 19 D). Nevertheless, I cannot exclude other reasons for the failure of the protocol. Only the detection of the complex protein RNA (similar to Figure 19) can be verified until the final step of the library generation. In 2012 Singh and colleagues defined the RNA interactome of the EJC. Upon identification of those mRNAs, they delineated six processes the EJC was involved in: RNA processing, chromatin organization, mitosis, cell cycle, positive regulation of transcription, negative regulation of transcription and cell division. Considering that BTZ binding to the mRNA is dependent on the EJC, I speculate that the RNA interactome of BTZ will be similar to the EJC interactome.

5.3 SELOR protein interactome

To gain a better insight in BTZ functions, I performed a MS analysis of FLAG-tagged S-long expressing Flp-In T-REx 293 cell lines. The wt protein interactome was compared to the corresponding W218D mutant, to distinguish between EJC-dependent and –independent binding partners. The specificity of the approach was confirmed by the identification of the EJC core components (MAGOH, Y14 and eIF4A3), which were among the most enriched proteins. Similarly, several members of the SRSF proteins were enriched in the IP of S-long, in good agreement with the identification of these proteins in the EJC-interactome (Singh et al., 2012) (Figure 20A). Singh and colleagues proposed that the SRSF proteins were stochiometric to the EJC. Nonetheless, I could not detect any other protein with the same intensity of the EJC components on a silver-stained gel (Figure 21A). Therefore, the precise abundance and stochiometry of SRSF proteins in purified EJCs needs to be re-evaluated. Among the other factors enriched, I identified the NMD proteins UPF3b and UPF2, components of the splicing-associated nineteen complex (NTC) and components of the U2 snRNP (Figure 20A). Interestingly, eIF3 was enriched in S-long W218D IP. Chazal and colleagues observed that BTZ is the only EJC component whose overexpression correlates with increased translation. This effect was mainly observed for intron-containing mRNAs and only weakly for intronless mRNAs. This would support the hypothesis that BTZ mediates the translation stimulation function of the EJC. On

the other hand, they observed that an EJC-impaired BTZ mutant was not able to increase the translation of spliced mRNAs, and modestly stimulated the translation of unspliced mRNAs (Chazal et al., 2013). This supports the hypothesis that BTZ could also work independently of the EJC. The identification of eIF3 in the EJC-impaired S-long mutant IP corroborates this hypothesis. Nevertheless, more studies will be required to elucidate the correlation between translation induction and BTZ.

In general, the proteins enriched in S-long IP were components of the EJC or engaged in processes BTZ was already described to be part of, such as NMD, transcription, splicing, spliceosome assembly and RNA processing. Moreover, proteins involved in apoptosis and skin development were identified in S-long IP. BTZ was originally identified as a protein implicated in cancer progression, thus the interaction with apoptotic and skin development factor might be reconduct to its role in cancer (Arriola et al., 2008; Tomasetto et al., 1995) (Figure 20B).

5.3.1 Functions of S-long binding partners

The SRSF proteins were highly enriched among the identified S-long binding partners. The role of SRSF1 in degradation of PTC-containing mRNAs was already described (Zhang and Krainer, 2004), but no systematical analysis to study the other members of the SRSF proteins family has been performed. Upon tethering to the mRNA, I observed a reduction of the mRNA levels for SRSF1, SRSF2, SRSF6 and SRSF7. This degradation was NMD-dependent, consistent with the observations that SRSF1 and SRSF7 seemed to activate an UPF2-dependent and endocleavage-based pathway (Figure 21 C and D). Currently, it is not possible to confirm the importance of the SRSF proteins in NMD activation or if BTZ-induced NMD requires the presence of SRSF proteins. Similarly, it is not clear if the SRSF proteins can independently activate NMD or if they require other components, such as the EJC. Nevertheless, the effect of the SRSF proteins was different from the spliceosomal components of the NTC complex (BCAS2 and PLRG1) and the NTC-related SKIP complex protein SNW1, also enriched in S-long IP. The reduction of the reporter level when tethering these components was not affected upon UPF2 KD, suggesting that they were UPF2-independent

(Figure 22 B, C and D). In addition, no 3′ fragment was generated upon their tethering in XRN1 depleted cells (Figure 22E). This raises the question of how the proteins of the NTC complex and SNW1 induce the reduction of the reporter mRNA, although they do not activate NMD. Indeed, they are UPF2-independent and the degradation of the reporter is not endocleavage-mediated. In addition, I could exclude that the reduction of the reporter at the steady state level is due to exonucleolytic cleavage, because UPF1 activation is required for this pathway (Loh et al., 2013; Unterholzner and Izaurralde, 2004). Instead, the effect of the tethered PLRG1 and SNW1 was not affected by UPF1 depletion (Supplementary figure S4). To define the mechanism that induces the reduction of the reporter level upon PLRG1, BCAS2 and SNW1 tethering, a more in depth analysis will be required.

Two groups of proteins were identified in the S-long IP, SRSF family and NTC complex, which, even though both induced reporter degradation, behaved in two different ways. The SRSF proteins are involved in NMD, while the NTC complex proteins function through a different mechanism. This is a clear indication that the effect observed for these proteins is specific of the protein itself and not just a consequence of the binding with S-long.

5.5 Conclusions

In this thesis work I identified new functions of the N-terminal and C-terminal regions of BTZ. Moreover, I established that two different regions of BTZ activate UPF2-dependent and –independent NMD and that the C-terminal region seems to be involved in polyadenylation. The N-terminal region of BTZ activates NMD in an EJC-independent way. However, it was not elucidated how BTZ N interacts with UPF2 in order to activate the UPF2-dependent NMD. A preliminary experiment to understand better BTZ N-terminus function could be the identification of its binding partners via MS. Similarly, I could not elucidate how the C-terminal region is involved in polyadenylation. In this respect, it is interesting to note that a protein reported to localize to the cytoplasm has instead a nuclear function. BTZ was reported to join the EJC in the cytoplasm.

Considering the identification of a nuclear function and with the support of more studies, the model of BTZ joining to the EJC might be revised.

6. Materials and Methods

6.1 Materials

6.1.1 Cell lines

Flp-In T-REx 293 cell lines (Life Technologies), containing stably inserted Tetracycline-inducible expression cassette were generated and cultured according to the manufacturer's recommendations. All transfections were performed in HeLa Tet-off (HTO) cells (Clontech). All the cells were grown at 37 °C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) GlutaMAX (Life Technologies), supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin.

6.1.2 Bacterial strains

For cloning purposes, XL1-blue *E.coli* (Stratagene) were used. For bacterial expression of recombinant proteins, BL21 star (DE3) (Invitrogen) were used. Cells were grown in LB medium or plates with the appropriate antibiotic.

6.1.3 Plasmids

Plasmid constructs β-globin wt, pCI-FLAG, pCI-MS2, pCI-PP7, pCI-λNV5, pCI-Venus, β-globin 4MS2, β-globin 4boxB, β-globin 4PP7, β-globin wt300+e3, pCI-TPI and expression vectors for BTZ were described previously (Gehring et al., 2009a) (V Boehm, N Haberman, F Ottens, J Ule and NH Gehring, in press).

6.1.4 Antibodies

Antibody	From	Dilution
UPF2	Dr. Jens Lykke-Andersen	1:2000
GFP	Abcam (ab290)	1:3000
FLAG	Sigma-Aldrich (F7425)	1:3000
V5	QED Bioscinscience (18870)	1:3000
Actin	Sigma-Aldrich (A5441)	1:5000
RMB8 (Y14)	Sigma-Aldrich (R07935)	1:1500

XRN1	Bethyl (A300-443A)	1:3000
MAGOH	Santa Cruz Biotechnology (sc-271365)	1:2000
eIF4A3[1]	Genscript	1:1000
Mouse-HRP	Jackson ImmunoResearch	1:5000
Rabbit-HRP	Jackson ImmunoResearch	1:5000

[1]raised in rabbits and affinity-purified by Genscript with a KLH-conjugated peptide (ATSGSARKRLLKEED)

6.1.5 siRNAs and Primers

siRNA name	siRNA sequence 5′-3′
Luciferase	AACGTACGCGGAATACTTCGA
XRN1	AAAGATGAACTTACCGTAGAA
UPF2	CACGTTGTGGATGGAGTGTTA
UPF1	AAGATGCAGTTCCGCTCCATT

Primer name	Primer sequence 5′-3′
BTZ 480 as	TTTTTTTGCGGCCGCTTACTTGCCCCGGACTTCCAA
BTZ 160 as	TTTTTGCGGCCGCTTAGTTCTCCACAGGCTCT
BTZ 137 as	TTTTTGCGGCCGCTTAGTCAGGCTTTTCTTCTCCC
BTZ 372 as	TTTTTTTGCGGCCGCTTAGCCAAGCACTGGAGCA
BTZ 110 se	TTTTCTCGAGTCCAAAGTGAGCTGAAA
pCI se	GGTGTCCACTCCCAGTTCA
pCI as	TGCATTCTAGTTGTGGTTTGTC
BTZ 286 as	TTTTTTTGCGGCCGCTTAACCCTGGTGGCGATG
BTZ 137 se	TTTTCTCGAGGACACCAAAAAGCACTGTGA
BTZ 255 se	TTTTCTCGAGACCGAAGAATCCGG
BTZ 286 se	TTTTCTCGAGGACCTTGGGGGCACCTACCAC
SNW1 se	TTTTCTCGAGATGGCGCTCACCAGCT
SNW1 as	TTTTTTTGCGGCCGCCTATTCCTTCCT
SRSF6 se	TTTTCTCGAGATGCCGCGCGTCTACA
SRSF6 as	TTTTTTTGCGGCCGCTTAATCTCTGGA

BCAS2 se	TTTTCTCGAGATGGCGGGCACAGGT
BCAS2 as	TTTTTTTTGCGGCCGCTCAGAAGTCTTG
PLRG1 se	TTTTCTCGAGATGGTCGAGGAGGTACA
PLRG1 as	TTTTTTTGCGGCCGCTTAAAATCTCTTT
SRSF7 se	TTTTCTCGAGATGATGTCGCGTTACGGGC
SRSF7 as	TTTTTTTGCGGCCGCTCAGTCCATTCT
CSTF1 se	CTTTGCTGCTGGGACTCCAGGACAGCCG
CSTF1 as	TTTTCTCGAGATGTACAGAACCAAAGTGG
CSTF3_1800 as	TTTTTTTGCGGCCGCTTATGGAGGTGCTA
CSTF3_60 se	TTTTCTCGAGGCGGAAAAGAAATTAGAA

All the primers in this table contain XhoI restriction site in the 5'and NotI in the 3'

Primer name	Primer sequence 5'-3'
oRP1628	CGAGCACAGAATTAATACGACTTTTTTTTTTTTTTTTTTTV[1]
oRP1630	CGAGCACAGAATTAATACGAC

[1]V, could be A, C, G.

6.1.6 Buffers

Bacteria culture

<u>Autoinduction medium LBM5052</u>

480 ml LB medium

10 ml 50xM

1 ml MgSO4

10 ml 50x5052

0.1 ml 1000xTrace metals

<u>50xM</u>

134 g Na2HPO4 x 7 H2O

68 g KH2PO4

53.5 g NH4Cl

14.2 g Na2SO4

H2O up to 400 ml

<u>LB medium</u>

10 g Tryptone

10 g NaCl

5 g Yeast extract

H20 up to 1L

<u>NP lyses buffer</u>

100 µg/ml DNAse

125 µg/ml Lysozyme

250 µg/ml RNAse

Protease Inhibitor dilution 1:250

NP buffer

50 mM NaH2PO4

300 mM NaCl

pH 8.0

50x5052

125 ml Glycerol

12.5 g Glucose

50 g α-D-Lactose

H2O up to 500 ml

1000xTrace metals

25 ml 50 mM FeCl3

500 μl 20 mM CaCl2

250 μl 10 mM MnCl2

250 μl 10 mM ZnSO4

H20 up to 50 ml

Transfection

2x BBS

50 mM BES

1.5 mM Na2HPO4

280 mM NaCl

pH 6.96

Northern Blot

RNA loading dye

50 % glycerol

1 mM EDTA

0.25 % brome-phenol blue

10x MOPS buffer

0.2 M MOPS

80 sodium acetate

10 mM EDTA

pH 7.0

RNA sample buffer

500 μl formamide (deionized)

200 μl 37% formaldehyde

100 μl 10x MOPS buffer

1 drop of 10 mg/ml

ethidiumbromide

20x SSC

3 M NaCl

0.3 M Trisodium-citrate

pH 7.0

Church buffer

0.5 M NaH2PO4

1 mM EDTA

7 % SDS

pH 7.2

Northern wash buffer I

2x SSC

0.1 % SDS

Northern wash buffer II

0.2x SSC

0.1 % SDS

Western Blot

1x Transfer buffer
25 mM Tris
192 mM glycine
0.1 % SDS
10 % MeOH

10x TBS
500 mM Tris pH 7.4
1500 mM NaCl

1x TTBS
1x TBS
0.2 % Tween-20

Blocking solution
1x TBST
5 % low fat milk powder

SDS Page

6x SDS-PAGE loading buffer
0.35 M Tris/HCl pH 6.8
10.28 % (w/v) SDS
36 % (v/v) glycerol
0.6 M dithiothreitol
0.2 % brome-phenol blue

4x Separating gel buffer
1.5 M Tris/HCl pH 8.8
0.4 % SDS
4x Stacking gel buffer
0.5 M Tris/HCl pH 6.8
0.4 % SDS

10 % Separating gel mix
42 ml H2O
33 ml 30 % acrylamide (37.5 : 1)
25 ml 4x separating gel buffer

Stacking gel mix
57.8 ml H2O
17.2 ml 30 % acrylamide (37.5 : 1)
25 ml 4x stacking gel buffer

1x SDS running buffer
25 mM Tris
192 mM glycine
0.1 % SDS

Immunoprecipitation

Lysis buffer
50 mM Tris [pH 7.2]
150 mM NaCl
0.5% Tween
50 μg/ml RNase A
complete protease inhibitor [Roche]

Wash buffer
50 mM Tris [pH 7.2]
150 mM NaCl
 0.5% Tween

CLIP

PNK Buffer
20 mM Tris-HCl, pH 7.4
10 mM MgCl2
0.2% Tween-20

Loading buffer

5 µl LDS (Novex)

13 µl H20

2 µl reducing reagent (Novex)

PNK mix

0.4 µl PNK

0.8 µl γ^{32}P-ATP

0.8 µl PNK buffer A

6 µl H20

High salt wash

50 mM

Tris-HCl, pH 7.4

1 M NaCl

1 mM EDTA

1% Igepal CA-630 (Sigma I8896)

0.1% SDS

0.5% sodium deoxycholate

Immunofluorescence

Permeabilising solution

0.5% Triton

PBS up to 1 ml

EMSA

Reaction buffer

100mM KCl

50 mM Hepes

10% glycerol

2mM EDTA

Native gel

1.3 ml acrylamide (37.5 : 1)

1ml TBE 10%

10 µl Temed

100 µl APS

H2O up to 10 ml

10X TBE

108 g of Tris base

55 g of boric acid

7.5 g of EDTA

H2O up to 1L

6.2 Methods

6.2.1 Cloning

pCI-FLAG, pCI-MS2, pCI-PP7, pCI-λNV5 plasmids were already available and they were used as vector. BTZ full-length wt and all the point mutants were as well already available. BTZ full-length was used as template DNA for the PCRs required to amplify the entire set of length mutants. SRSFs, CSTF1, CSTF3, SNW1, BCAS2 and PLRG1 DNAs were PCR-amplified from HeLa cDNA using Phusion

DNA polymerase (Finnzymes). XhoI and NotI restriction site were inserted with the primers. The primers used for the amplification are listed in the paragraph 6.1.5. After separating the PCR product by agarose gel electrophoresis, the fragment was cut from the gel and purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). Subsequent digestion with XhoI and NotI (NEB) was performed o/n at 37 °C and the resulting fragment was again purified and ligated with XhoI/NotI cut pCI-FLAG vector. After 4h of ligation, transformation in chemically competent XL1-blue *E.coli* was performed and the bacteria were grown o/n at 37 °C on LB plate with the appropriate antibiotic. Single colonies containing the intended insert were identified by colony PCR and sequenced (GATC Biotech). The colony containing the insert was then grown in LB medium and DNA obtained with Midi prep kit (Macherey Nagel) and sequenced (GATC Biotech).

6.2.2 Plasmid transfections

For tethering experiments HeLa cells were transfected with 1.5 µg tethering construct, 2 µg 4boxB or 4PP7 construct, 0.6 µg control construct (wt 300+e3), 0.5 µg GFP expression vector and 0.5 µg of rescue construct in the rescue experiments. For the detection of the 3′ fragment, HeLa cells were transfected with 1.5 µg tethering construct, 2 µg TPI 4MS2, 0.75 µg control construct (β-globin wt), 0.5 µg GFP expression vector. Transfections for immunoprecipitations were done with 1.5 µg of the FLAG-expression plasmid and 0.5 µg of GFP- and 0.5 µg V5-expression. For tethering experiments HeLa cells were seeded in 6 well plates, 240.000 cells/well one day prior transfection. For immunoprecipitations experiments HeLa cells were seeded in 6 cm dished 70.000 cell/dish. 4h prior transfection the medium was changed. For 6-well plates, the indicated amount of plasmids was filled up with ultrapure H_2O to 90 µl, 10 µl 2.5 M $CaCl_2$ and 100 µl 2x BBS. For 6 cm dishes, the indicated amount of plasmids was filled up with ultrapure H_2O to 180 µl, 20 µl 2.5 M $CaCl_2$ and 200 µl 2x BBS. The mix was incubated at RT for 15 min before applying to the cells.

6.2.3 siRNA and plasmid transfection

For UPF2 siRNA transfections, 10^6 HeLa cells were grown over night in 10 cm plates in minus antibiotic medium and transiently transfected with 600 pmol

siRNA using Lipofectamine RNAiMAX (Life Technologies). The day post transfection the cells were seeded in minus antibiotic medium in 6 well plates and 8 hours later transfected again with 150 pmol siRNA. 24 h later they were transfected with plasmid constructs. For XRN1 siRNA transfections, $5x10^5$ HeLa cells were grown over night in 6 cm plates in minus antibiotic medium and transiently transfected with 300 pmol siRNA. 24 h post transfection the cells were split 1:2 in 10 cm plates in minus antibiotic medium and the day after transfected again with 600 pmol siRNA. 24 h later the cells were transferred to 6 well plates and the day after transfected with DNA constructs. For UPF1 siRNA transfection, $10x10^5$ HeLa cells were grown over night in 10 cm plates and transiently transfected with 600 pmol siRNA. The day post transfection cells were seed in 6 well plates and the 8 hours later transfected with DNA constructs.

6.2.4 RNA extraction and Northern blot

Total RNA was extracted with TRIzol (Life Technologies) according to the manual. The samples were analyzed by northern blot. 3 µl of RNA were mixed with 16 µl of RNA sample buffer and incubated 15 minutes at 65 °C. Then 2µl of RNA loading dye were added. The samples were loaded on the Northern gel (1% agarose, 1x MOPS buffer and 7.5 % formaldehyde) and run at 33V. When the run was completed, the gel was washed twice in water, once in NaOH 50 mM and incubated in 20X SSC for 40 minutes. For the capillary transfer a neutral nylon membrane (Hybond-NX [GE Healthcare] or Roti®-Nylon 0.2 [Carl-Roth]) was used. Subsequently, RNA was cross-linked by UV irradiation at 254 nm to the membrane by applying 120 mJ/cm2 in a UV crosslinker, incubated 1h in Church Buffer and then with the *in vitro* transcribed α-^{32}P-GTP labeled RNA probe o/n at 65 °C. The membrane was washed twice for 15 minutes in Northern wash I and twice for 15 minutes in Northern wash II. It was dried for 40 minutes and exposed to a phosphoimager screen (Fuji BAS-MP IP) for about 4 h. Signals were quantified using a Typhoon Trio (GE Healthcare).

6.2.5 Tethering assay

The tethering assay is a method used to study NMD inducing proteins. The protein of interest is modified with the addition of a specific tag that recognizes

its binding sequence in the 3′ UTR of the reporter mRNA. In this work several tag were used that recognized specific binding sequences.

TAG	Reporter mRNA
MS2	TPI 4MS2-SMG5
λN	β-globin 4boxB
PP7	β-globin 4PP7

Figure 23 is a schematic representation of the tethering assay. The 3′UTR of the reporter is modified with the addition of a specific binding sequence (1). The tagged protein binds this sequence downstream the normal stop codon (2). In case of an EJC protein this mimics the situation of an EJC protein downstream of a PTC. The protein then can exhibit an effect on the mRNA itself or by recruiting additional factors (3). Transfections are performed as indicated in paragraph 6.2.2. Two days later the cells are harvested in TRIzol (Life Technologies). RNA extraction and northern blot are performed as indicated in paragraph 6.2.4.
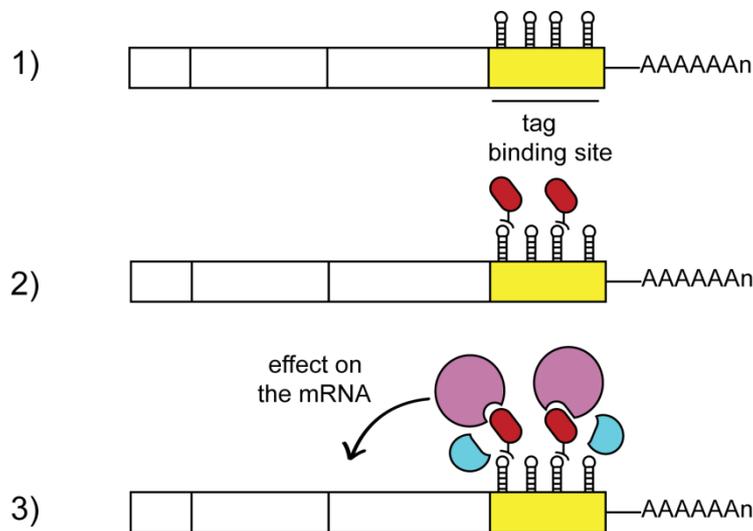
Figure 23. Schematic representation of the tethering assay 1) Modified reporter mRNA. 2) Binding of the tagged protein. 3) Effect on the mRNA

6.2.6 Immunoprecipitation

FLAG-complexes were immunoprecipitated from HeLa cell lysates. Cells were harvested 2 days after transfection (see paragraph 6.2.2) in 400 µl of lysis buffer. The samples were frozen for at least 2h and then centrifuged for 10 minutes at 10000 x g. 40 µl of each sample were set aside as input for the western blot. The rest of the sample was mixed with 12 µl of M2 anti-FLAG magnetic beads (Sigma) pre-washed in lysis buffer and incubated at 4 °C over night. The next day the beads were washed three times (each of 5 minutes) with 600 µl wash buffer and complexes were eluted with 40 µl of SDS-sample buffer.

6.2.7 Western blot

Protein samples mixed with SDS-PAGE loading buffer were loaded on SDS-polyacrylamide gels. The proteins were then transferred on Hybond-ECL nitrocellulose membrane (GE Healthcare) using 1x transfer buffer and the semi-dry system. The transfer was performed for 90-120 minutes at 10 V. The membrane was incubated in blocking solution for 1h at RT and incubated o/n at 4 °C with the primary antibody diluted in the blocking solution. After 3 washings of 10 minutes in TTBS, the membrane was incubated for 1h RT in the secondary antibody diluted in the blocking. After other 3 washings of 10 minutes in TTBS, the membrane was incubated with ECL western blotting detection reagent (GE Healthcare) and developed with ImageQuant LAS 4000.

6.2.8 Silver stain

Silver staining was performed with the SilverQuest Silver Staining Kit (Invitrogen) according to the manufacturer's protocol.

6.2.9 Crosslinking and immunoprecipitation (CLIP)

Flp-In T-REx 293 cell lines were induced for 3 days with 1 µg/ml of tetracycline. After medium removal, cells were washed once with PBS (Gibco) and crosslinked in 1 ml of PBS with variable intensities of UV light at 254 nm, harvested and frozen as pellet. The day of use, the pellet was resuspendend in 1m of lyses buffer (with protease inhibitor added freshly). The sample was sonicated 2 times with 10s bursts at 5 decibels. 2µl of Turbo DNAse (Ambion) and 10 µl of RNAse If (NEB) (1:50 (low, L) or 1:500 (high, H) dilution) were added and samples were

shaken at 37 °C for 3 minutes at 1100 rpm followed by 3 minutes on ice. When low RNAse dilution is added to the sample, the RNA bound to the protein is cutted in small fragments that minimally affect the molecular weight of the protein and its running behavior; this condition is used to control the specificity of the IP. The high RNA dilution instead generates bigger RNA fragments that affect the molecular weight of the protein, which will appear a smear on the gel. Subsequently the samples were centrifuged at max speed for 20 minutes. 40 μl of each sample were set aside as input for the western blot, the rest of the sample was incubated with 12 μl of M2 anti-FLAG magnetic beads (Sigma) pre-washed in lyses buffer. After 2h, the beads were washed twice with high salt buffer and twice with PNK buffer. 1/10 of the last wash was set aside and eluted with 20 μl of SDS-sample buffer and used for western blot together with the input to check protein expression and IP. The last wash was removed from the beads that were incubated with 8 μl PNK mix for 5 minutes at 37 °C and shaken at 1100 rpm. The PNK mix was removed and the samples were eluted in 20 μl of loading buffer at 70 °C for 5 minutes. The entire sample was load on a 4-12% pre-cast gel (Novex) and run according to the producer. The gel was subsequently dried and exposed to a phosphoimager screen (Fuji BAS-MP IP) for about 2h. Signals were quantified using a Typhoon Trio (GE Healthcare).

6.2.10 Mass spectrometry

FLAG-complexes were immunoprecipitated from lysates of Flp-In T-REx 293 cell lines, which were induced for 3 days with tetracycline, using M2 anti-FLAG magnetic beads (Sigma) at 4 °C over night in western blot lyses buffer and analysed with MALDI TOF/TOF Analyzer (AB SCIEX).

6.2.11 EMSA assay

The oligo used as probe was 5′end labelled for 1h at 37 °C according to the following reaction :

20 pmol oligo

1 μl PNk

4 μl PNK buffer

1 μl γ^{32}P-ATP

35 µl H20.

The probe was then precipitated o/n with the addition of 40 µl of NaoAc pH 5.5 and 1 ml EtOH and diluted in H20.

For the EMSA assay, the desired amount of protein and 5nM of oligo probe were incubated with the reaction buffer and 1mM DDT for 20 minutes at 20 °C. Native gel electrophoresis (4% acrylamide gel in 0.5 X TBE running buffer) was used to separate the free and bound probe.

6.2.12 *in vitro* protein purification

PABP RRM1-4 was already available in the lab as GST fusion tag in pGEX6p3 vector (GE Healtcare), while S-long was cloned in pETDuet-1 vector (Novagen) and fused to a N-terminal FLAG tag and C-terminal Strep tag. BL21 star DE3 *E.coli* strain was used for the plasmid transformation. A single bacteria colony was pre-inoculate at 37 °C o/n in 10 ml of LB medium containing Ampicillin. The pre culture was then added to 500 ml of autoinduction LBM5052 medium and grown o/n at 28 °C. The following day the bacteria were harvested by centrifugation at 4700 rpm for 15 minutes, resuspended in NP lyses buffer and stored at -20 °C. Prior to purification the suspension was thawed and let rotate at 4°C for 2h. Afterwards, bacteria cells were lysed by sonication (4x 1 minute sonication at 50% duty cycle and output set to 5) using a Branson Sonifier 250. The sample was centrifuged at 20 000 rpm for 30 minutes to separate cell debris from crude extract. The crude extract was loaded on StrepTactin Superflow Plus Cartridge (QIAGEN) and purified with the Äkta FPCL (GE Healthcare). The column was washed with 10 ml of NP buffer and the protein eluted with NP supplemented with 2.5 mM desthiobiotin (Sigma-Aldrich). The column was regenerated by washing with 15 ml of NP buffer supplemented with 1 mM HABA (2-(4- Hydroxyphenylazo)benzoic acid) and washed again with 15 ml of NP buffer. The purified protein concentration was measured with the NanoDrop (Thermo Fisher).

6.2.13 Immunofluorescence

Cells were transfected in 6 well plates as indicated in paragraph 6.2.2, with the difference that only 100 ng of the Venus-tagged protein were transfcted. The day

after transfection, the cells were splitted 1:2 on cover slip in 6well plate. The second day after transfection the medium was removed, the cells washed with PBS and fixed for 15 minutes in fresh prepared 3% paraformaldehide. After 3 washes of 10 minutes in PBS, the cells were incubated for 5 minutes on ice in permeabilising solution and washed again. After the last wash DAPI 1:10000 diluted in PBS was added to the cells for 10 minutes at RT in the dark. The cells were washed again PBS and the cover slip were mounted on the microscope slide (Roth).

7. Supplemental material
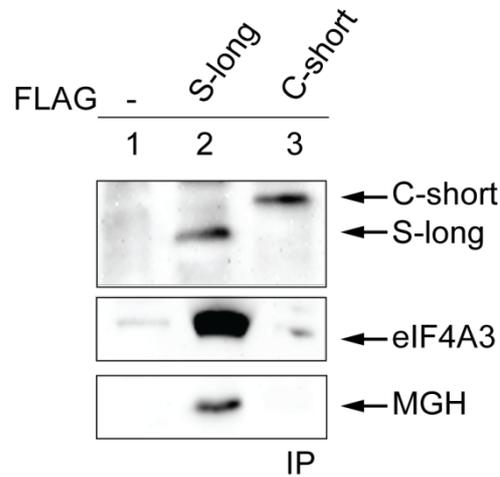
Supplementary figure S1. C-short does not bind the EJC. Western blot of samples derived from Flp-In T-REx 293 stable cell lines, analyzed as in figure 11C.
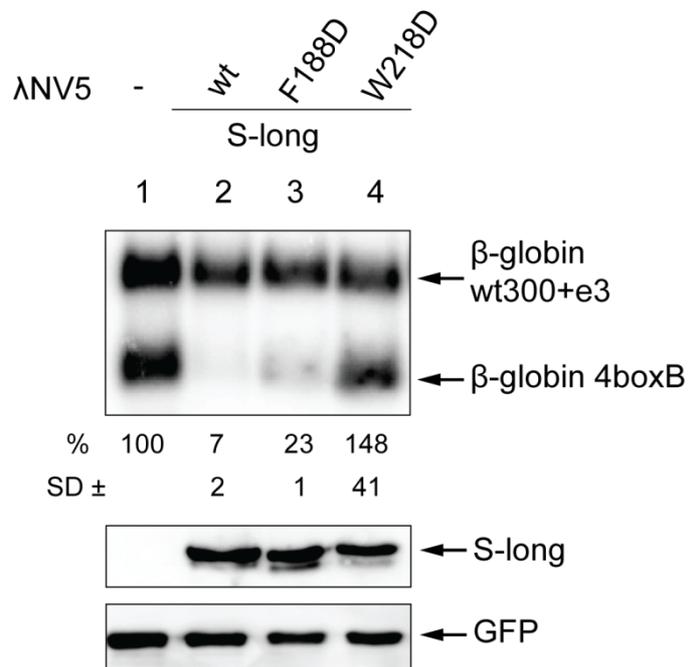
Supplementary figure S2. S-long induced NMD. Northern blot analysis of RNA samples and western blot analysis of proteins derived from HeLa cells as described in Figure 9B.
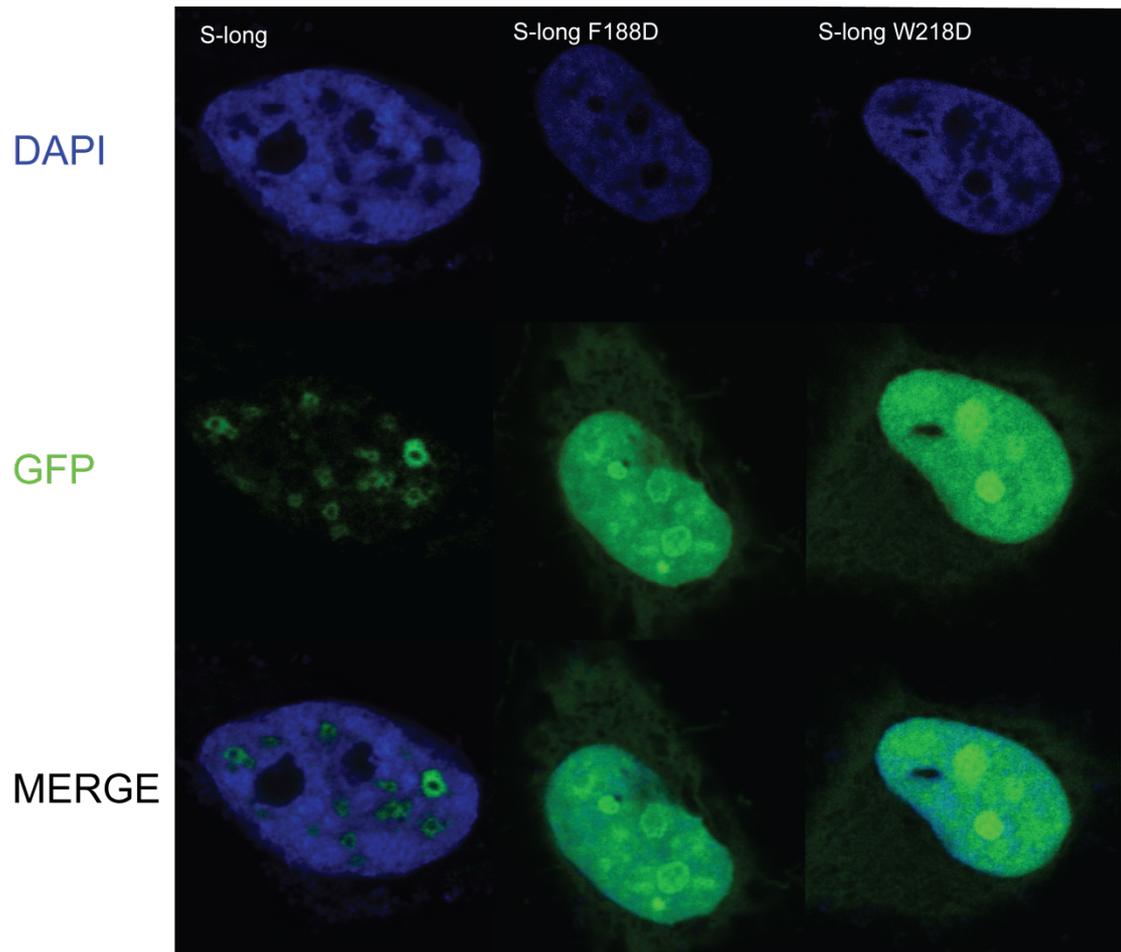
Supplementary figure S3. S-long mutants localization. GFP tagged proteins were transfected in Hek 293 cells. Dapi is used to stain the nucleus. . Images were taken using FV1000 confocal microscope (Olympus).
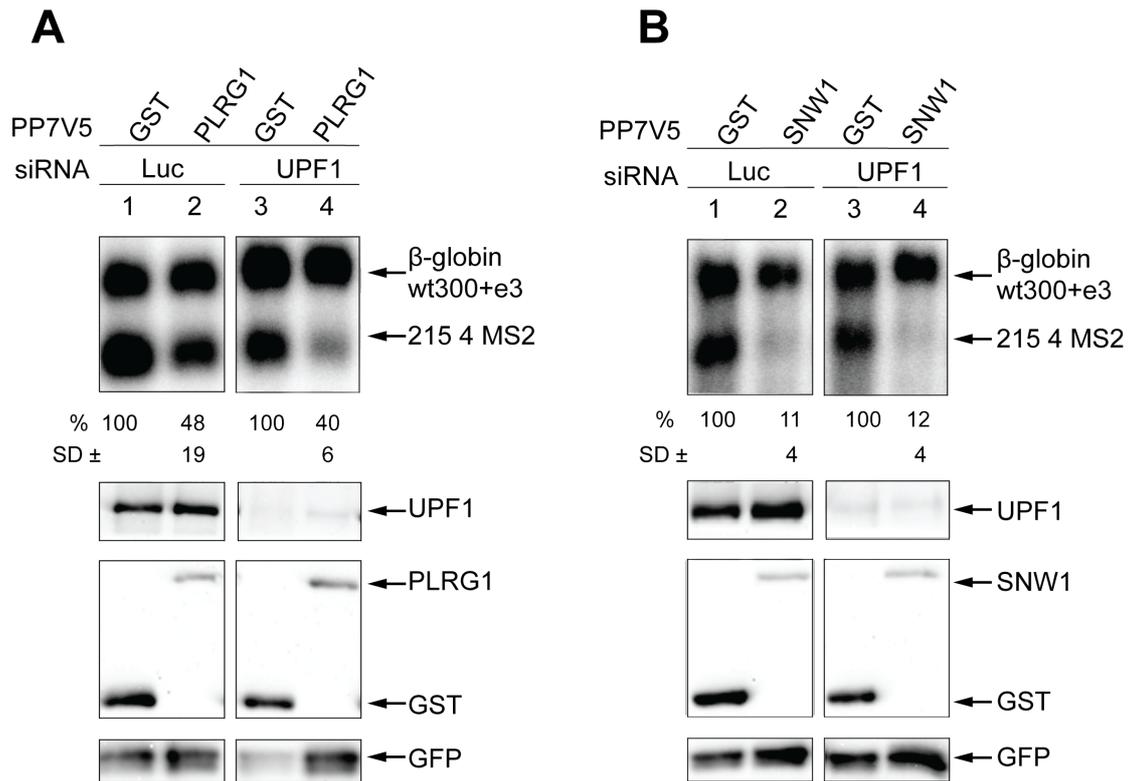
Supplementary figure S4. PLRG1 and SNW1 are UPF1-independent. (A, B) Northern blot (upper panel) of RNA samples extracted from HeLa cells transfected with the indicated siRNAs and constructs. Co-transfected β-globin wt300+e3 served as control mRNA and β-globin 4 boxB as reporter mRNA (n=3). Western blot analysis (bottom panel) was performed as in Figure 9B.

8. Bibliography

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nat Struct Mol Biol *19*, 568-576.

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature *432*, 112-118.

Amrani, N., Sachs, M.S., and Jacobson, A. (2006). Early nonsense: mRNA decay solves a translational problem. Nat Rev Mol Cell Biol *7*, 415-425.

Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science *313*, 1968-1972.

Anderson, P., and Kedersha, N. (2006). RNA granules. J Cell Biol *172*, 803-808.

Arriola, E., Marchio, C., Tan, D.S., Drury, S.C., Lambros, M.B., Natrajan, R., Rodriguez-Pinilla, S.M., Mackay, A., Tamber, N., Fenwick, K.*, et al.* (2008). Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines. Lab Invest *88*, 491-503.

Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrancois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P.*, et al.* (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251-262.

Baguet, A., Degot, S., Cougot, N., Bertrand, E., Chenard, M.P., Wendling, C., Kessler, P., Le Hir, H., Rio, M.C., and Tomasetto, C. (2007). The exon-junction-complex-component metastatic lymph node 51 functions in stress-granule assembly. J Cell Sci *120*, 2774-2784.

Bai, Y., Auperin, T.C., Chou, C.Y., Chang, G.G., Manley, J.L., and Tong, L. (2007). Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol Cell *25*, 863-875.

Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B., and Le Hir, H. (2005). The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. Nat Struct Mol Biol *12*, 861-869.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Sauliere, J., Wittkopp, N., and Izaurralde, E. (2007). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett *581*, 2845-2853.

Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified

human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. RNA *16*, 2384-2403.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846-850.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713-725.

Bono, F., Ebert, J., Unterholzner, L., Guttler, T., Izaurralde, E., and Conti, E. (2004). Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. EMBO Rep *5*, 304-310.

Bono, F., and Gehring, N.H. (2011). Assembly, disassembly and recycling: the dynamics of exon junction complexes. RNA Biol *8*, 24-29.

Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im. Mol Cell *12*, 1467-1476.
Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. Curr Opin Struct Biol *12*, 123-133.

Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol *15*, 85-93.

Chan, S., Choi, E.A., and Shi, Y. (2011). Pre-mRNA 3′-end processing complex assembly and function. Wiley Interdiscip Rev RNA *2*, 321-335.

Chanarat, S., and Strasser, K. (2013). Splicing and beyond: the many faces of the Prp19 complex. Biochim Biophys Acta *1833*, 2126-2134.

Chazal, P.E., Daguenet, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., and Le Hir, H. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. Proc Natl Acad Sci U S A *110*, 5903-5908.

Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5′ end of mRNA. Cell *127*, 1389-1400.

Cheung, A.C., and Cramer, P. (2012). A movie of RNA polymerase II transcription. Cell *149*, 1431-1437.

Clancy, S. (2008). RNA Splicing: Introns, Exons and Spliceosome. Nature education.

Conti, E., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol *17*, 316-325.

Cosson, B., Berkova, N., Couturier, A., Chabelskaya, S., Philippe, M., and Zhouravleva, G. (2002a). Poly(A)-binding protein and eRF3 are associated in vivo in human and Xenopus cells. Biol Cell *94*, 205-216.

Cosson, B., Couturier, A., Chabelskaya, S., Kiktev, D., Inge-Vechtomov, S., Philippe, M., and Zhouravleva, G. (2002b). Poly(A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation. Mol Cell Biol *22*, 3301-3315.

Culbertson, M.R. (1999). RNA surveillance. Unforeseen consequences for gene expression, inherited genetic disorders and cancer. Trends Genet *15*, 74-80.

Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and Peltz, S.W. (1998). The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. Genes Dev *12*, 1665-1677.

D'Ambrogio, A., Nagaoka, K., and Richter, J.D. (2013). Translational control of cell growth and malignancy by the CPEBs. Nat Rev Cancer *13*, 283-290.

Daguenet, E., Baguet, A., Degot, S., Schmidt, U., Alpy, F., Wendling, C., Spiegelhalter, C., Kessler, P., Rio, M.C., Le Hir, H.*, et al.* (2012). Perispeckles are major assembly sites for the exon junction core complex. Mol Biol Cell *23*, 1765-1782.

Degot, S., Le Hir, H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Seraphin, B., Rio, M.C., and Tomasetto, C. (2004). Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. J Biol Chem *279*, 33702-33715.

Degot, S., Regnier, C.H., Wendling, C., Chenard, M.P., Rio, M.C., and Tomasetto, C. (2002). Metastatic Lymph Node 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. Oncogene *21*, 4422-4434.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835-845.

Diem, M.D., Chan, C.C., Younis, I., and Dreyfuss, G. (2007). PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. Nat Struct Mol Biol *14*, 1173-1179.

Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. Curr Biol *12*, 1060-1067.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol *3*, 195-205.

Eberle, A.B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol *16*, 49-55.

Eberle, A.B., Stalder, L., Mathys, H., Orozco, R.Z., and Muhlemann, O. (2008). Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. PLoS Biol *6*, e92.

Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet *14*, 496-506.

Feoktistova, K., Tuvshintogs, E., Do, A., and Fraser, C.S. (2013). Human eIF4E promotes mRNA restructuring by stimulating eIF4A helicase activity. Proceedings of the National Academy of Sciences of the United States of America *110*, 13339-13344.

Ferraiuolo, M.A., Lee, C.S., Ler, L.W., Hsu, J.L., Costa-Mattioli, M., Luo, M.J., Reed, R., and Sonenberg, N. (2004). A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. Proc Natl Acad Sci U S A *101*, 4118-4123.

Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003). An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat Biotechnol *21*, 89-92.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A *102*, 16176-16181.

Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003). A novel mode of RBD-protein recognition in the Y14-Mago complex. Nat Struct Biol *10*, 433-439.

Frischmeyer, P.A., and Dietz, H.C. (1999). Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet *8*, 1893-1900.

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M., and Izaurralde, E. (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. Curr Biol *11*, 1716-1721.

Gehring, N.H., Hentze, M.W., and Kulozik, A.E. (2008). Tethering assays to investigate nonsense-mediated mRNA decay activating proteins. Methods Enzymol *448*, 467-482.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes

of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell *20*, 65-75.

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol *7*, e1000120.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon junction complexes by PYM. Cell *137*, 536-548.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b form an NMD-activating complex. Mol Cell *11*, 939-949.

Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoecklin, G., Dember, L.M., and Anderson, P. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. Mol Biol Cell *15*, 5383-5398.

Giorgi, C., and Moore, M.J. (2007). The nuclear nurture and cytoplasmic nature of localized mRNPs. Semin Cell Dev Biol *18*, 186-193.

Glanzer, J., Miyashiro, K.Y., Sul, J.Y., Barrett, L., Belt, B., Haydon, P., and Eberwine, J. (2005). RNA splicing capability of live neuronal dendrites. Proc Natl Acad Sci U S A *102*, 16859-16864.

Gorlach, M., Burd, C.G., and Dreyfuss, G. (1994). The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. Exp Cell Res *211*, 400-407.

Gu, M., and Lima, C.D. (2005). Processing the message: structural insights into capping and decapping mRNA. Curr Opin Struct Biol *15*, 99-106.
Guhaniyogi, J., and Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. Gene *265*, 11-23.

Gunkel, N., Yano, T., Markussen, F.H., Olsen, L.C., and Ephrussi, A. (1998). Localization-dependent translation requires a functional interaction between the 5' and 3' ends of oskar mRNA. Genes Dev *12*, 1652-1664.

Hachet, O., and Ephrussi, A. (2001). Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. Curr Biol *11*, 1666-1674.

Hall, G.W., and Thein, S. (1994). Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. Blood *83*, 2031-2037.

Hansen, S., Lane, D.P., and Midgley, C.A. (1998). The N terminus of the murine p53 tumour suppressor is an independent regulatory domain affecting activation and thermostability. J Mol Biol *275*, 575-588.

Harigaya, Y., and Parker, R. (2012). Global analysis of mRNA decay intermediates in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *109*, 11764-11769.

Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H., and Luhrmann, R. (2009). Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. Mol Cell Biol *29*, 281-301.

Hinnebusch, A.G. (2006). eIF3: a versatile scaffold for translation initiation complexes. Trends Biochem Sci *31*, 553-562.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges. Cold Spring Harbor Perspectives in Biology *4*.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics *123*, 301-313.

Hogg, R., McGrail, J.C., and O'Keefe, R.T. (2010). The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. Biochem Soc Trans *38*, 1110-1115.

Holbrook, J.A., Neu-Yilik, G., Hentze, M.W., and Kulozik, A.E. (2004). Nonsense-mediated decay approaches the clinic. Nat Genet *36*, 801-808.

Holzmann, K., Gerner, C., Poltl, A., Schafer, R., Obrist, P., Ensinger, C., Grimm, R., and Sauermann, G. (2000). A human common nuclear matrix protein homologous to eukaryotic translation initiation factor 4A. Biochem Biophys Res Commun *267*, 339-344.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res *37*, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44-57.

Huang, M., Rech, J.E., Northington, S.J., Flicker, P.F., Mayeda, A., Krainer, A.R., and LeStourgeon, W.M. (1994). The C-protein tetramer binds 230 to 240 nucleotides of pre-mRNA and nucleates the assembly of 40S heterogeneous nuclear ribonucleoprotein particles. Mol Cell Biol *14*, 518-533.

Huang, Y., and Steitz, J.A. (2005). SRprises along a messenger's journey. Mol Cell *17*, 613-615.

Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA *14*, 2609-2617.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., Konig, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274-287.

Hurt, E., Strasser, K., Segref, A., Bailer, S., Schlaich, N., Presutti, C., Tollervey, D., and Jansen, R. (2000). Mex67p mediates nuclear export of a variety of RNA polymerase II transcripts. J Biol Chem *275*, 8361-8368.

Isken, O., and Maquat, L.E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. Nat Rev Genet *9*, 699-712.

Ivanov, P.V., Gehring, N.H., Kunz, J.B., Hentze, M.W., and Kulozik, A.E. (2008). Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. EMBO J *27*, 736-747.

Ivshina, M., Lasko, P., and Richter, J.D. (2014). Cytoplasmic Polyadenylation Element Binding Proteins in Development, Health, and Disease. Annu Rev Cell Dev Biol.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127.
Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2012). Termination and post-termination events in eukaryotic translation. Adv Protein Chem Struct Biol *86*, 45-93.

Johnstone, O., and Lasko, P. (2001). Translational regulation and RNA localization in Drosophila oocytes and embryos. Annu Rev Genet *35*, 365-406.

Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N. (2005a). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes & Development *19*, 104-113.

Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N. (2005b). Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. Genes Dev *19*, 104-113.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev *20*, 355-367.

Kataoka, N., Diem, M.D., Kim, V.N., Yong, J., and Dreyfuss, G. (2001). Magoh, a human homolog of Drosophila mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. EMBO J *20*, 6424-6433.

Kataoka, N., Yong, J., Kim, V.N., Velazquez, F., Perkinson, R.A., Wang, F., and Dreyfuss, G. (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. Mol Cell *6*, 673-682.

Kay, B.K., Williamson, M.P., and Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J *14*, 231-241.

Kedersha, N., and Anderson, P. (2002). Stress granules: sites of mRNA triage that regulate mRNA stability and translatability. Biochem Soc Trans *30*, 963-969.

Kedersha, N., Chen, S., Gilks, N., Li, W., Miller, I.J., Stahl, J., and Anderson, P. (2002). Evidence that ternary complex (eIF2-GTP-tRNA(i)(Met))-deficient preinitiation complexes are core constituents of mammalian stress granules. Mol Biol Cell *13*, 195-210.

Kim, V.N., Kataoka, N., and Dreyfuss, G. (2001). Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. Science *293*, 1832-1836.

Kim-Ha, J., Kerr, K., and Macdonald, P.M. (1995). Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. Cell *81*, 403-412.

Kinniburgh, A.J., Maquat, L.E., Schedl, T., Rachmilewitz, E., and Ross, J. (1982). mRNA-deficient beta o-thalassemia results from a single nucleotide deletion. Nucleic Acids Res *10*, 5421-5427.

Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. J Biol Chem *275*, 6530-6536.

Konarska, M.M., Grabowski, P.J., Padgett, R.A., and Sharp, P.A. (1985). Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature *313*, 552-557.

Kornberg, R.D. (1999). Eukaryotic transcriptional control. Trends Cell Biol *9*, M46-49.

Krause, S., Fakan, S., Weis, K., and Wahle, E. (1994). Immunodetection of poly(A) binding protein II in the cell nucleus. Exp Cell Res *214*, 75-82.

Lau, C.K., Diem, M.D., Dreyfuss, G., and Van Duyne, G.D. (2003). Structure of the Y14-Magoh core of the exon junction complex. Curr Biol *13*, 933-941.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J *20*, 4987-4997.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000a). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J *19*, 6860-6869.

Le Hir, H., Moore, M.J., and Maquat, L.E. (2000b). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. Genes Dev *14*, 1098-1108.

Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. Genes Dev *5*, 2303-2314.

Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Mol Cell Biol *12*, 2165-2177.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. EMBO J *21*, 3536-3545.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A *100*, 189-192.

Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. Eur J Biochem *247*, 461-469.

Li, C., Lin, R.I., Lai, M.C., Ouyang, P., and Tarn, W.Y. (2003). Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1. Mol Cell Biol *23*, 7363-7376.

Lin, C.L., Huang, Y.T., and Richter, J.D. (2012). Transient CPEB dimerization and translational control. RNA *18*, 1050-1061.

Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. Genes Dev *27*, 2125-2138.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem J *417*, 15-27.

Longman, D., Plasterk, R.H., Johnstone, I.L., and Caceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans NMD pathway. Genes Dev *21*, 1075-1085.

Lu, S., and Cullen, B.R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. RNA *9*, 618-630.

Luna, R., Rondon, A.G., and Aguilera, A. (2012). New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. Biochim Biophys Acta *1819*, 514-520.

Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. Proc Natl Acad Sci U S A *96*, 14937-14942.

Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. Nature *413*, 644-647.

Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. (2000). Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. Cell *103*, 1121-1131.

Lykke-Andersen, S., and Jensen, T.H. (2007). Overlapping pathways dictate termination of RNA polymerase II transcription. Biochimie *89*, 1177-1182.

Macchi, P., Kroening, S., Palacios, I.M., Baldassa, S., Grunewald, B., Ambrosino, C., Goetze, B., Lupas, A., St Johnston, D., and Kiebler, M. (2003). Barentsz, a new component of the Staufen-containing ribonucleoprotein particles in mammalian cells, interacts with Staufen in an RNA-dependent manner. J Neurosci *23*, 5778-5788.

MacDonald, C.C., and Redondo, J.L. (2002). Reexamining the polyadenylation signal: were we wrong about AAUAAA? Mol Cell Endocrinol *190*, 1-8.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci *65*, 1099-1122.

Mao, X., Schwer, B., and Shuman, S. (1995). Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene. Mol Cell Biol *15*, 4167-4174.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev *19*, 1512-1517.

Matsumoto, K., Wassarman, K.M., and Wolffe, A.P. (1998). Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. EMBO J *17*, 2107-2121.

Mayer, B.J., Hamaguchi, M., and Hanafusa, H. (1988). A novel viral oncogene with structural similarity to phospholipase C. Nature *332*, 272-275.

Mazza, C., Ohno, M., Segref, A., Mattaj, I.W., and Cusack, S. (2001). Crystal structure of the human nuclear cap binding complex. Mol Cell *8*, 383-396.

Mazza, C., Segref, A., Mattaj, I.W., and Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. EMBO J *21*, 5548-5557.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet *36*, 1073-1078.

Metzstein, M.M., and Krasnow, M.A. (2006). Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. PLoS Genet *2*, e180.

Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., Garneau, D., Gendron, D., Lapointe, E., Couture, S., *et al.* (2012). Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. Mol Cell Biol *32*, 954-967.

Micklem, D.R., Adams, J., Grunert, S., and St Johnston, D. (2000). Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. EMBO J *19*, 1366-1377.

Monshausen, M., Gehring, N.H., and Kosik, K.S. (2004). The mammalian RNA-binding protein Staufen2 links nuclear and cytoplasmic RNA processing pathways in neurons. Neuromolecular Med *6*, 127-144.

Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. Science *309*, 1514-1518.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Moriarty, P.M., Reddy, C.C., and Maquat, L.E. (1998). Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. Mol Cell Biol *18*, 2932-2939.

Muhlemann, O., Eberle, A.B., Stalder, L., and Zamudio Orozco, R. (2008). Recognition and elimination of nonsense mRNA. Biochim Biophys Acta *1779*, 538-549.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci *23*, 198-199.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N., and Muhlemann, O. (2010). Nonsense-mediated mRNA decay in human cells:

mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci *67*, 677-700.

Nielsen, K.H., Chamieh, H., Andersen, C.B., Fredslund, F., Hamborg, K., Le Hir, H., and Andersen, G.R. (2009). Mechanism of ATP turnover inhibition in the EJC. RNA *15*, 67-75.

Nott, A., Le Hir, H., and Moore, M.J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. Genes Dev *18*, 210-222.

Nott, A., Meislin, S.H., and Moore, M.J. (2003). A quantitative analysis of intron effects on mammalian gene expression. RNA *9*, 607-617.

Nurenberg, E., and Tampe, R. (2013). Tying up loose ends: ribosome recycling in eukaryotes and archaea. Trends Biochem Sci *38*, 64-74.

Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell *12*, 1187-1200.

Pal, M., and Luse, D.S. (2002). Strong natural pausing by RNA polymerase II within 10 bases of transcription start may result in repeated slippage and reextension of the nascent RNA. Mol Cell Biol *22*, 30-40.

Palacios, I.M. (2002). RNA processing: splicing and the cytoplasmic localisation of mRNA. Curr Biol *12*, R50-52.

Palacios, I.M., Gatfield, D., St Johnston, D., and Izaurralde, E. (2004). An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. Nature *427*, 753-757.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.

Proudfoot, N.J., and Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211-214.

Reed, R. (2003). Coupling transcription, splicing and mRNA export. Curr Opin Cell Biol *15*, 326-331.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. Genes Dev *23*, 1247-1269.

Roberts, R.M., Cleland, T.J., Gray, P.C., and Ambrosiano, J.J. (2004). Hidden markov model for competitive binding and chain elongation. J Phys Chem B *108*, 6228-6232.

Sachs, A. (1990). The role of poly(A) in the translation and stability of mRNA. Curr Opin Cell Biol *2*, 1092-1098.

Sauliere, J., Haque, N., Harms, S., Barbosa, I., Blanchette, M., and Le Hir, H. (2010). The exon junction complex differentially marks spliced junctions. Nat Struct Mol Biol *17*, 1269-1271.

Schell, T., Kocher, T., Wilm, M., Seraphin, B., Kulozik, A.E., and Hentze, M.W. (2003). Complexes between the nonsense-mediated mRNA decay pathway factor human upf1 (up-frameshift protein 1) and essential nonsense-mediated mRNA decay factors in HeLa cells. Biochem J *373*, 775-783.

Schmidt, U., Im, K.B., Benzing, C., Janjetovic, S., Rippe, K., Lichter, P., and Wachsmuth, M. (2009). Assembly and mobility of exon-exon junction complexes in living cells. RNA *15*, 862-876.

Schreiber, V., Moog-Lutz, C., Regnier, C.H., Chenard, M.P., Boeuf, H., Vonesch, J.L., Tomasetto, C., and Rio, M.C. (1998). Lasp-1, a novel type of actin-binding protein accumulating in cell membrane extensions. Mol Med *4*, 675-687.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Luhrmann, R., and Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. EMBO J *16*, 3256-3271.

Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. Cell *9*, 645-653.

Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. Nat Struct Biol *7*, 838-842.

Shiimori, M., Inoue, K., and Sakamoto, H. (2013). A specific set of exon junction complex subunits is required for the nuclear retention of unspliced RNAs in Caenorhabditis elegans. Mol Cell Biol *33*, 444-456.

Sims, R.J., 3rd, Mandal, S.S., and Reinberg, D. (2004). Recent highlights of RNA-polymerase-II-mediated transcription. Curr Opin Cell Biol *16*, 263-271.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 750-764.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol *6*, e111.

Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. Nat Struct Mol Biol *12*, 645-653.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. Cell *136*, 731-745.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J., *et al.* (1993). SH2 domains recognize specific phosphopeptide sequences. Cell *72*, 767-778.

St Johnston, D. (2005). Moving messages: the intracellular localization of mRNAs. Nat Rev Mol Cell Biol *6*, 363-375.

Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. Cell Rep *2*, 454-461.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proc Natl Acad Sci U S A *93*, 15081-15085.

Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P., and Izaurralde, E. (2000). REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. RNA *6*, 638-650.

Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. EMBO J *20*, 1785-1796.

Tange, T.O., Shibuya, T., Jurica, M.S., and Moore, M.J. (2005). Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. RNA *11*, 1869-1883.

Thein, S.L., Hesketh, C., Taylor, P., Temperley, I.J., Hutchinson, R.M., Old, J.M., Wood, W.G., Clegg, J.B., and Weatherall, D.J. (1990). Molecular basis for dominantly inherited inclusion body beta-thalassemia. Proc Natl Acad Sci U S A *87*, 3924-3928.

Tomasetto, C., Regnier, C., Moog-Lutz, C., Mattei, M.G., Chenard, M.P., Lidereau, R., Basset, P., and Rio, M.C. (1995). Identification of four novel human genes amplified and overexpressed in breast carcinoma and localized to the q11-q21.3 region of chromosome 17. Genomics *28*, 367-376.

Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods *37*, 376-386.

Unterholzner, L., and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. Mol Cell *16*, 587-596.

van Eeden, F.J., Palacios, I.M., Petronczki, M., Weston, M.J., and St Johnston, D. (2001). Barentsz is essential for the posterior localization of oskar mRNA and colocalizes with it to the posterior pole. J Cell Biol *154*, 511-523.

Vessey, J.P., Vaccani, A., Xie, Y., Dahm, R., Karra, D., Kiebler, M.A., and Macchi, P. (2006). Dendritic localization of the translational repressor Pumilio 2 and its contribution to dendritic stress granules. J Neurosci *26*, 6496-6508.

Viphakone, N., Hautbergue, G.M., Walsh, M., Chang, C.T., Holland, A., Folco, E.G., Reed, R., and Wilson, S.A. (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat Commun *3*, 1006.

Wagner, E., and Lykke-Andersen, J. (2002). mRNA surveillance: the perfect persist. J Cell Sci *115*, 3033-3038.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66*, 759-768.

Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J.S., Kristiansen, K., Krogh, A., Wang, J., and Porse, B.T. (2012). Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. Genome Biol *13*, R35.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B. (1998). Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell *2*, 135-140.

Wiegand, H.L., Lu, S., and Cullen, B.R. (2003). Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. Proc Natl Acad Sci U S A *100*, 11327-11332.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. Genes Dev *15*, 2215-2228.

Yang, Q., Coseno, M., Gilmartin, G.M., and Doublie, S. (2011). Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure *19*, 368-377.

Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D., and Shatkin, A.J. (1997). Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. Proc Natl Acad Sci U S A *94*, 12898-12903.

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998a). At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. Mol Cell Biol *18*, 5272-5283.

Zhang, J., Sun, X., Qian, Y., and Maquat, L.E. (1998b). Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. RNA *4*, 801-815.

Zhang, Z., and Krainer, A.R. (2004). Involvement of SR proteins in mRNA surveillance. Mol Cell *16*, 597-607.

Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191-207.

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. Nature *407*, 401-405.

9. Acknowledgement

First of all, I want to thank my supervisor Dr. Niels Gehring, for giving me the opportunity to work and graduate in his laboratory. Thank you for constantly support my project, for the interesting discussions and the motivations received. Thanks even for the not scientific support and for showing me how much passion a person can put in this job.

Additionally I want to thank Prof. Dr. Jürgen Dohmen for taking the time to read and grade my thesis, Prof. Dr Martin Hülskamp for being part of my thesis committee and Martina Remboldt for writing the protocol of my defense.

Thanks to Aga, Tobi, Franzi, Lena, Kusum, Marie, Volker and Ben for the great lab atmosphere, for the work done together and for the good time out of the lab as well. In particular thanks to Volker, Tobi and Franzi for the comments to this thesis and the support with the German language. In addition thanks to Heidi and Juliane for the technical help.

And thanks to all my friends here in Cologne, the ones that were part of these great four years from the beginning and the ones met on the way. Thanks to Martin, Anto, Aga, Sarah, Flo, Nico, Ame and Andrea. You guys made these years unforgettable.

Most of all, thanks to my family, without which I could have not been here.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit–einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von PD Dr. Niels H. Gehring betreut worden.

Köln, 01.10.2014 _____
Simona Ciriello

Simona Ciriello
Born on December, 24, 1986 in Naples
Current address: Gustavstr. 6, 50937 Cologne, Germany

Higher education

2010-2014	**PhD student** Thesis title: Deciphering the role of BTZ during mammalian non-sense mRNA decay Supervisor: Dr N.H. Gehring, Institute of Genetic, University of Cologne, Germany
2008-2010	**M.Sc. in Biotechnology**; overall grade 110/110 cum laude (Excellent) Thesis title: Role of the genes Serpine 1 and Runx1 in the maintenance of ES cells undifferentiated state mediated by Klf5 Supervisor: Silvia Parisi, CEINGE, University of Naples, Italy
2005-2008	**B.Sc. in Biotechnology**; overall grade 110/110 cum laude (Excellent); Thesis title: Application of ELISA-MST assay to the study of the protein SHP1 Supervisor: Vittorio de Franciscis, DBBM, University of Naples, Italy

School education

2000-2005	Salvatore Di Giacomo, San Sebastiano al Vesuvio, Naples, Italy