INAUGURAL - DISSERTATION

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich-Mathematischen Gesamtfakultät

der

Ruprecht - Karls - Universität

Heidelberg

vorgelegt von

Styliani Lamprinaki (Diploma in Molecular Biology and Genetics)

aus Thessaloniki, Griechenland

Tag der mündlichen Prüfung: 23/10/2009

DISSERTATION

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the

Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Styliani Lamprinaki (Diploma in Molecular Biology and Genetics)

From Thessaloniki, Greece

Date of oral examination: 23/10/2009

T h e m a

Studying splicing and the formation of mRNPs

Gutachter: 1. Dr. Stephen Cusack

2. Prof. Dr. Ed Hurt

Summary

Splicing is a fundamental step in eukaryotic gene expression. During this process, introns are excised from pre-mRNAs and exons are ligated to form a continuous reading frame. As a consequence of splicing, a number of proteins are deposited on the mRNA product. These proteins can influence a number of downstream processes of the mRNA metabolism, such as export from the nucleus, translation efficiency and stability in the cytoplasm. It becomes evident that splicing not only plays an important role in the maturation of an mRNA in the nucleus, but also influences distant downstream processes affecting the ultimate fate of mRNAs.

The exon junction complex (EJC) is assembled ~20-24 nucleotides upstream of splice junctions in a splicing-dependent, but sequence-independent manner (Le Hir et al., 2000). Trying to elucidate what influences the composition of the EJC, our hypothesis was that the intron downstream of the assembled EJC could influence its composition. It has been reported that NMD-competent mRNPs can be generated by distinct routes (Gehring et al., 2005). The question we wanted to address was whether this heterogeneity of NMD-activating complexes is reflected into EJC assemblies. In order to address this question we wanted to affinity select mRNPs spliced *in vitro*. We spliced NMD substrates *in vitro* and established a GRNA affinity purification method. The protein composition of the affinity purified mRNPs could not be studied with mass spectrometry, possibly due to the low amount of proteins. Immunoblot analysis revealed the presence of eIF4A3 and UAP56 within the affinity selected mRNPs.

EJC has been assembled *in vitro* from recombinant proteins on a RNA substrate in the presence of ATP (Ballut et al., 2005). Therefore, its assembly can occur in the absence of the physiological deposition machinery, the spliceosome. However, under these experimental conditions, its physiological assembly hierarchy was not determined. One of the goals of this thesis was to study EJC deposition during splicing. The stepwise assembly of EJC during splicing, in parallel to the spliceosome assembly was investigated. We used an *in vitro* splicing system that faithfully recapitulates the splicing-dependent deposition of EJC proteins. We found that eIF4A3 and MAGOH-Y14 formed a pre-EJC before exon ligation and that deposition of eIF4A3, MAGOH and Y14 required splicing, while Barentsz and UPF3b did not require splicing to bind at

the EJC. These results show that the minimal EJC (eIF4A3 and MAGOH-Y14) assembles on mRNA prior to exon ligation, while Barentsz and UPF3b join the complex later, after the completion of splicing to form an NMD-competent core EJC.

While studying the splicing-mediated deposition of EJCs, we noticed that MAGOH mutants lacking the interaction with PYM subtly but reproducibly co-immunoprecipitated more spliced MINX mRNA than wild type MAGOH. Therefore, we hypothesized that EJCs containing such mutants are more stable than EJCs with wild type MAGOH. To directly test if PYM influences the EJC, we purified recombinant PYM (rPYM) from bacterial expression cultures and added increasing amounts of rPYM to splicing reactions containing FLAG-tagged EJC proteins. The investigation led to the following observations: PYM reduced the amount of EJCs bound to spliced mRNAs and that the N-terminus of PYM was required for EJC disassembly. In addition, it was shown that PYM dissociated assembled EJCs but did not inhibit EJC assembly. These results reveal that PYM is an EJC disassembly factor *in vitro* that antagonizes important EJC functions.

A number of protein complexes assemble on the mRNA in the nucleus, such as EJC and TREX. Evidence exists in the literature that EJC and TREX are connected. Some of the TREX complex components have been included into the EJC components (Gatfield et al., 2001; Le Hir et al., 2001b; Le Hir et al., 2000). It has yet to be demonstrated whether TREX complex deposition is EJC-dependent. We investigated the TREX complex deposition using the *in vitro* splicing system described before. The role of DDX39, a UAP56 paralogue protein, was addressed for the first time and compared to UAP56. The results revealed that TREX deposition was splicing-independent and in particular that the recruitment of UAP56 and DDX39 was EJC-independent, while ALY/REF interacted with the EJC. In addition, functional characterization of ALY/REF and UAP56 revealed that the ALY/REF residues 204-241 were required for binding to the cap fragment, but not to the EJC fragment and that UAP56 mutants did not precipitate unspliced pre-mRNA or spliced mRNA fragments. The splicing-independent TREX complex deposition suggests that splicing is not necessary for export, but it may enhance export of spliced mRNAs.

Zusammenfassung

Das Spleißen ist einer der grundlegenden Schritte der Genexpression in Eukaryonten. Durch den Spleißprozess werden Introns aus der pre-mRNA ausgeschnitten und die Exons zu einem durchgängigen offenen Leserahmen verknüpft. Als Folge des Spleißprozesses binden eine Reihe von Proteinen an die gespleißte mRNA. Diese Proteine beeinflussen andere Schritte des mRNA Stoffwechsels, z. B. den Export der mRNA aus dem Zellkern und die Effizienz der Translation und Stabilität im Zytoplasma. Das Spleißen spielt also nicht nur eine wichtige Rolle bei der Reifung von mRNAs im Zellkern, sondern bestimmt bei späteren Prozessen die endgültige Bestimmung von mRNAs.

Der Exon Junction Complex (EJC) bindet an einer Position 20-24 Nukleotide vor Spleißverbindungen an die mRNA. Diese Bindung erfolgt spleißabhängig, ist jedoch vollkommen sequenzunabhängig (Le Hir et al., 2000). Wir wollten untersuchen, ob die Zusammensetzung von EJCs durch das von der Bindungsstelle stromab gelegene Intron beeinflusst werden kann. Frühere Arbeiten legten nahe, dass NMD kompetente mRNP Komplexe auf verschiedenen Wegen generiert werden können (Gehring et al., 2005). Ob die unterschiedlichen NMD kompetenten Komplexe mit unterschiedlich zusammengesetzten EJCs korrelieren war eines der zentralen Fragen unserer Untersuchungen. Hierfür etablierten wir die Affinitätsaufreinigung von mRNPs, die mittels Spleißen in vitro generiert wurden. Solche in vitro mRNPs wurden mittels GRNA Chromatographie aufgereinigt. Die Zusammensetzung der Proteine in den isolierten mRNPs konnte aber nicht mit Massenspektrometrie aufklärt werden, da die benötigte Menge an Protein nicht erreicht wurde. In Immunoblots konnten jedoch in den aufgereinigten mRNPs die Proteine eIF4A3 und UAP56 nachgewiesen werden.

Der EJC kann *in vitro* aus seinen vier rekombinanten Proteinkomponenten in Anwesenheit von ATP auf einem RNA Substrat zusammengesetzt werden (Ballut et al., 2005). Dies bedeutet, dass er auch in Abwesenheit der physiologischen Spleißmaschinerie assemblieren kann. Der normale Ablauf des Zusammenbaus von EJCs konnte unter diesen Bedingungen aber nicht untersucht werden. Eines der Ziele

dieser Doktorarbeit war die Untersuchung des EJC Zusammenbaus während des Spleißvorganges. Hierzu wurde der schrittweise Aufbau des EJC parallel zum Spleißvorgang untersucht. Mit Hilfe eines experimentellen Systems, das die spleißabhängige Assemblierung von EJC-Proteinen darstellen kann, fanden wir, dass eIF4A3 und MAGOH-Y14 bereits vor der Exon-Ligation einen pre-EJC bilden. Obwohl eIF4A3 und MAGOH-Y14 während des Spleißens zusammengesetzt werden, binden Barentsz und UPF3b später an den bereits assemblierten pre-EJC. Dies bedeutet, dass erst durch die Bindung von Barentsz und UPF3b nach dem Spleißen ein NMD-kompetenter EJC Kernkomplex gebildet wird.

Während unserer Untersuchungen zur spleißabhängigen Assemblierung des EJC beobachteten wir, dass MAGOH Mutanten, die nicht mehr mit dem Protein PYM interagieren konnten, reproduzierbar mehr gespleißte mRNA immunpräzipitierten als der MAGOH Wildtyp. Wir stellten daher die Hypothese auf, dass EJCs, die diese Mutanten enthalten, stabiler sind als EJCs mit dem MAGOH Wildtyp. Um herauszufinden, ob PYM die Stabilität des EJC direkt beeinflussen kann, haben wir rekombinantes PYM (rPYM) bakteriell produziert und rPYM in Spleißreaktionen eingesetzt. Wir konnten zeigen, dass PYM die Menge von RNA-gebundenen EJCs verringert, und dass der N-Terminus von PYM für dieses Auseinanderbauen von EJCs benötigt wird. Wir fanden ausserdem, dass PYM zwar reife EJCs dissoziieren kann, nicht jedoch die Assemblierung des EJC während des Spleißens behindert. PYM ist also ein Faktor, der EJCs *in vitro* disassemblieren kann und der deshalb wichtigen Funktionen des EJC entgegenwirken kann.

Eine Reihe von Proteinkomplexen bindet im Zellkern an die mRNA, z.B. der EJC oder der TREX-Komplex. Frühere Arbeiten konnten eine Verbindung zwischen dem EJC und dem TREX Komplex nachweisen. Einige der TREX Komponenten wurden ursprünglich als EJC Proteine angesehen (Gatfield et al., 2001; Le Hir et al., 2001b; Le Hir et al., 2000). Dennoch wurde bislang noch nicht gezeigt, ob die Rekrutierung des TREX Komplex in Abhängigkeit des EJC erfolgt. Wir haben die Bindung des TREX Komplexes an mRNA mit unserem *in vitro* Spleißsystem untersucht. Die Funktion des Proteins DDX39, ein Homolog von UAP56, wurde dabei ebenfalls analysiert und mit UAP56 verglichen. Unsere Arbeiten konnten zeigen, dass

die Rekrutierung von UAP56 und DDX39 unabhängig vom EJC erfolgte, wohingegen ALY/REF mit dem EJC interagierte. Funktionelle Analysen zeigten, dass die Aminosäuren 204-241 von ALY/REF für die Bindung der mRNA Kappe benötigt wurden, jedoch nicht für die Interaktion mit dem EJC. UAP56 Mutanten banden weder an gespleißte, noch an ungespleißte RNAs. Die spleißunabhängige Bindung des TREX Komplex könnte darauf hindeuten, dass der Spleißvorgang nicht notwendig für den mRNA Export ist, diesen jedoch stimulieren kann.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

The process of eukaryotic gene expression involves a number of interconnected steps with messenger RNA (mRNA) being the key intermediate in this process. During gene expression, the nascent mRNA transcripts are subjected to a number of processes in the nucleus, including capping at the 5'end, splicing to remove introns, and polyadenylation at the 3'end, before they are exported to the cytoplasm to be translated. A number of quality control mechanisms, such as nonsense-mediated decay (NMD), have evolved to recognize mistakes that occur during molecular processes in gene expression and to prevent the production of faulty gene products. Distinct multi-component complexes carry out each of the above mentioned steps. Most of the processes and the complexes are physically and functionally coupled to one another. The mRNA does not travel alone throughout the cell; it is associated with proteins forming mRNPs (messenger ribonucleoprotein complexes).

A very important feature of eukaryotic genes is the intron and a central process in the eukaryotic cells is splicing. Splicing removes the intervening sequences of the genes and unites the coding regions. It is carried out by a multi-component protein complex, the spliceosome. During splicing in the nucleus of metazoan cells, the spliceosome deposits the exon junction complex (EJC) on the nascent mRNA. The EJC modulates important nuclear and cytoplasmic events in gene expression such as nuclear export, nonsense-mediated mRNA decay, polysome association and translation. Components of the EJC are also involved in the localization of specific transcripts during development.

A large number of different proteins bind to mRNA as it progresses to form a mature mRNP in the nucleus. Many of these proteins are associated with certain steps of the gene-expression pathway, such as transcription, splicing or polyadenylation, and are generally removed before nuclear export. Apart from those, many proteins remain bound to the mRNA and accompany mRNPs to the cytoplasm. A very appropriate example is the exon junction complex (EJC). Its components are thought to recruit the

member of another protein complex, the TREX complex (TRanscription EXport complex).

1.1 The steps of gene expression

In eukaryotic cells, the expression of most genes is a multi-step process that starts with the activation of transcription of the respective gene by RNA polymerases, leads to the production of proteins and ceases with the degradation of the products. At an activated locus, the RNA polymerase II transcription machinery copies the genetic information into the pre-mRNA. Eukaryotic RNAs undergo significant co- and post-transcriptional modifications. The 5' ends of eukaryotic pre-mRNAs are modified co-transcriptionally by the addition of a protective m^7GpppN cap structure that is also required for transcriptional elongation, splicing and mRNA export. The 3'end of an mRNA is cleaved and the poly(A)-tail is added (for a review, see (Shatkin et al., 2000)). A typical mRNA is shown in figure 1 and the different steps of gene expression are depicted schematically in figure 2. Genes in eukaryotes often contain introns (intervening sequences) that must be removed during gene expression. This process is known as splicing and represents one of the major steps in the control of eukaryotic gene expression (Jurica et al., 2003). Splicing will be described in detail in the next section. The regulated mechanism of splicing and splice site recognition can result in alternative splicing, allowing multiple different proteins to be translated from a single RNA transcript and hence is a means to modulate the final outcome of gene expression (Smith et al., 1989). Since these processes are tightly coupled and utilize common factors, transcripts that have failed to progress through a maturation step access more downstream processes less efficiently. Therefore, pre-mRNAs that are not efficiently processed may be retained in the nucleus and degraded by the nuclear exosome (Hilleren et al., 1999). After their proper processing, however, mRNAs are exported via the nuclear pore from the nucleus to the cytoplasm of the eukaryotic cell (for a review, see (Conti et al., 2001)). In the cytoplasm, translation, the deciphering of the mRNA

into a polypeptide, takes place. The multi-step process of translation is subdivided into the main steps of initiation, elongation and termination and is catalyzed by the ribosome. Each of these phases requires a distinct subset of protein factors that function in conjunction with the ribosome such as initiation factors, elongation factors and release factors. Once the mRNA has completed its function or if it is defective, it will be degraded by various RNA degradation processes after a characteristic time unique to each mRNA (Jacobson et al., 1996). In principle, the mRNA levels in a cell are influenced both by the rate of transcription and the rate of mRNA turnover (Jacob et al., 1961). Bulk mRNA turnover can be considered as the key mechanism controlling the expression of many genes (Wilusz et al., 2004).

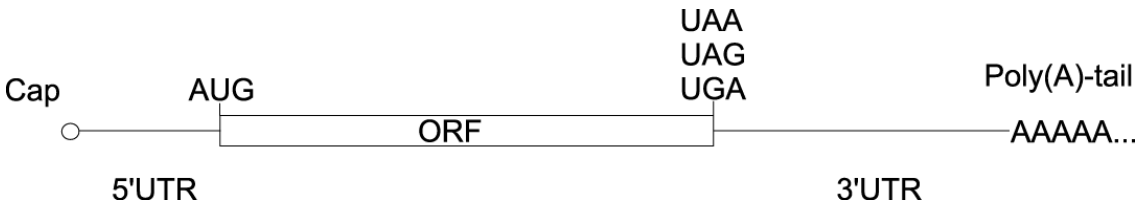


Figure 1. Schematic representation of an mRNA. A typical mRNA consists of a 5'and 3'untranslated region (UTR) and open reading frame (ORF). The 5'end has a protective methylated cap (m^7GpppG) and the 3'end a poly(A)-tail. The codon AUG is the initiation codon and codons UAA, UAG and UGA are termination codons.

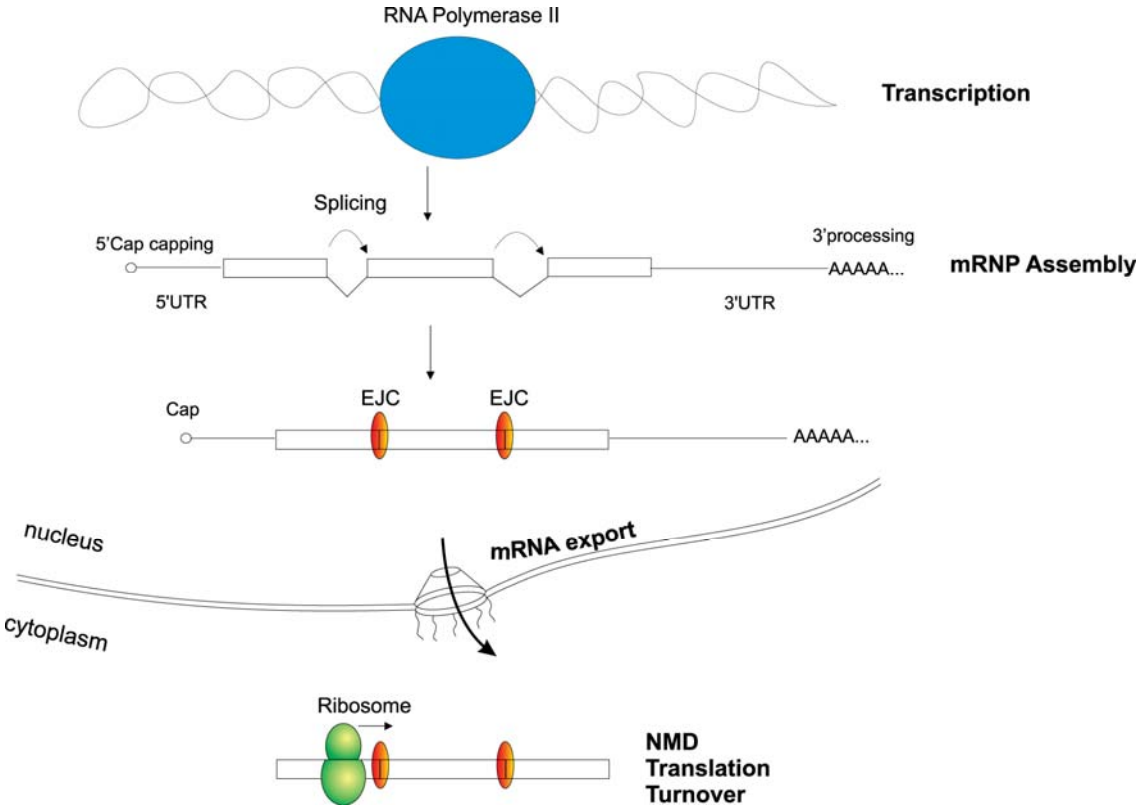


Figure 2. Simplified scheme of eukaryotic gene expression. The figure shows the different steps of gene expression, from transcription and nuclear mRNP processing steps (such as 5'end capping, splicing, mRNP assembly, 3'- end processing) to mRNP surveillance and RNA export. All these steps are coupled with or coordinated by transcription. EJC-exon junction complex; NMD-nonsense-mediated decay.

1.2 Splicing: at the center of gene expression

In most of the eukaryotic genes the coding sequences (exons) are interrupted by intervening regions (introns). During splicing, the introns are removed and the exons are joined together, to generate a functional mRNA. Splicing consists of two subsequent transesterification reactions. During the first reaction, the 2'-hydroxyl of a conserved adenosine residue (the adenosine in the branch site) carries out a nucleophilic attack on the upstream boundary of the intron (the 5'splice site). The outcome of this reaction is the release of the upstream exon and the formation of a lariat intron. The 3'-hydroxyl group of the newly released exon performs a second nucleophilic attack on the downstream boundary of the intron (the 3'splice site). This reaction results in the release of the lariat intron and the ligation of the two exons. The biochemical steps are depicted schematically in figure 3.

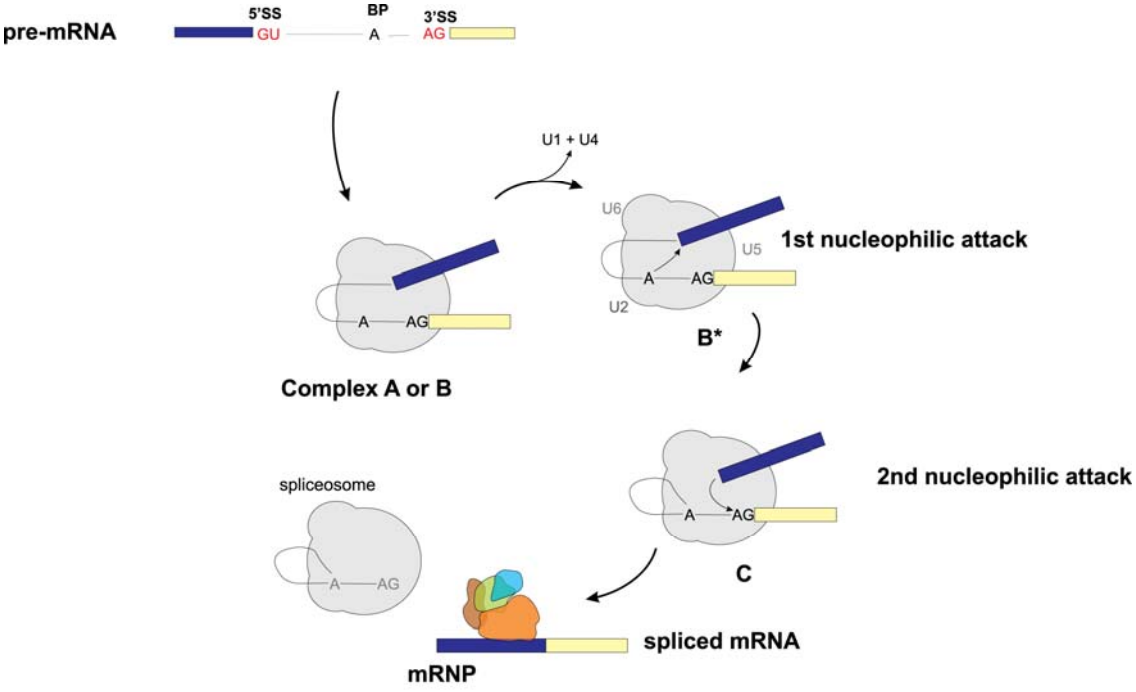


Figure 3. Diagram illustrating the two biochemical steps of splicing. The grey cloud represents the spliceosome. The coloured shapes on the spliced mRNA represent the proteins associated with the mRNA during splicing.

Splicing is carried out by a sophisticated machinery, the spliceosome. It is composed of RNAs and proteins, more specifically by five spliceosomal small nuclear ribonucleoprotein complexes (U1, U2, U4, U5 and U6 snRNPs) termed UsnRNPs (uridine-rich small ribonuclear particles) and numerous non-small ribonucleoproteins (non-snRNPs). Each UsnRNP consists of an UsnRNA associated with a set of seven Sm or Sm-like proteins and several particle-specific factors. U1, U2, U4, U5 and U6 are the major UsnRNPs and they are responsible for splicing the majority of introns (U2-type introns). U11, U12, U4atac and U6atac are less abundant snRNPs that are responsible for splicing, together with U5, of a less common class of introns, called U12-type introns.

The spliceosome assembly begins with the interaction of U1 with the 5'splice site and U2 with the branch site, leading to complex A. Base-paired U4 and U6 bind U5 form a tri-snRNP that joins complex A to give complex B. Structural changes that involve the dissociation of the intermolecular U4-U6 RNA base-pairing and the formation of interactions among U6, U2 and the pre-mRNA, catalytically activate the spliceosome. Complex B undergoes the first splicing reaction generating complex C. Subsequently, complex C undergoes the second catalytic step. After the second step, the complex is disassembled and the mRNA product is released. Figure 4 outlines the splicing assembly.

Constitutive splicing takes place at strong splice sites and in the absence of repression factors. When these strong optimal splice sites are suppressed, alternative suboptimal sites are used and this process is called alternative splicing. Most alternatively spliced introns are controlled by several splicing enhancers and silencers.

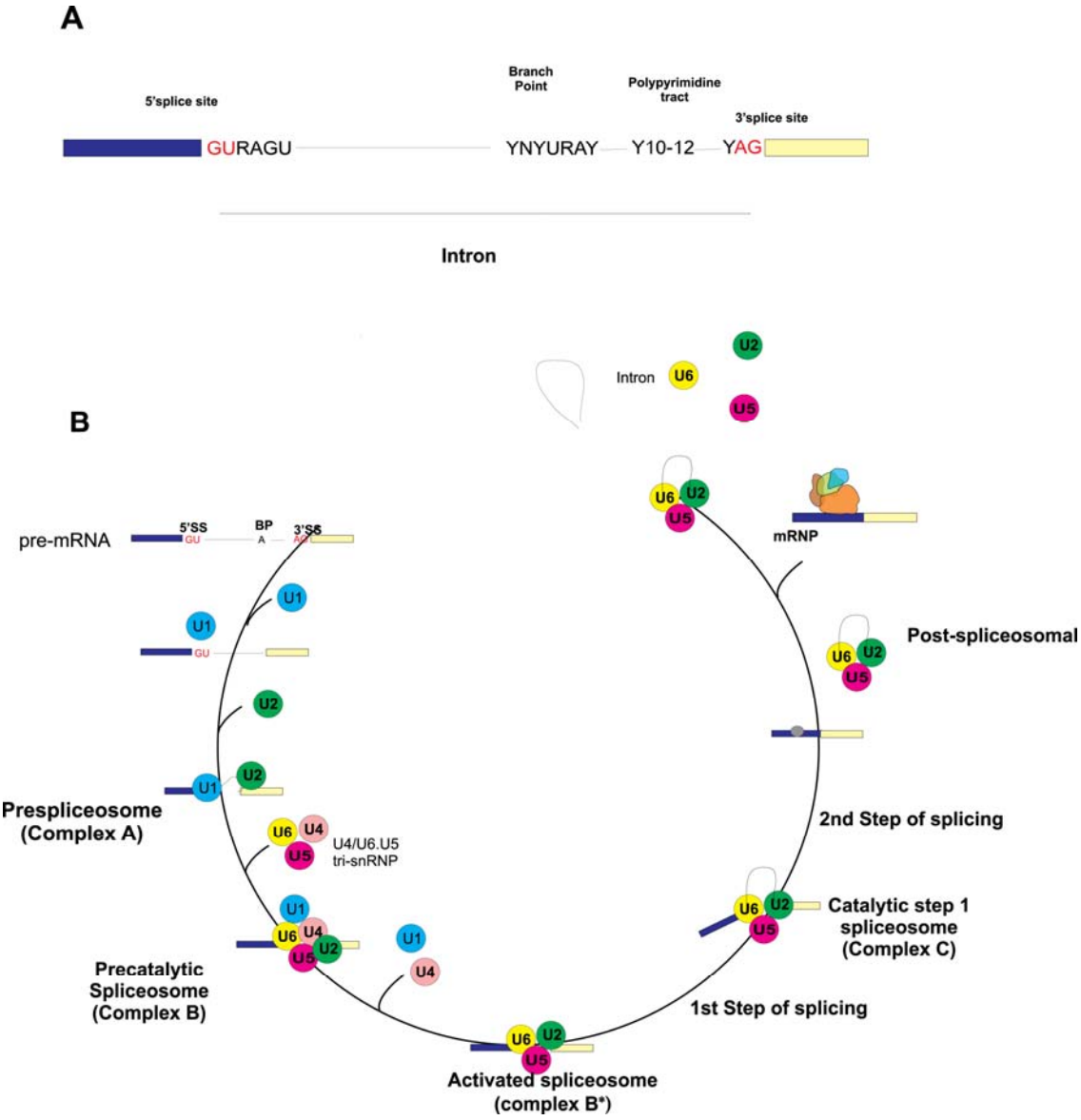


Figure 4. Spliceosome Assembly. (A) The sequence features of metazoan pre-mRNAs are depicted schematically. The exons are depicted in a box shape and the intron as a line. In the scheme, the consensus sequences [5'splice site (SS), branch point sequence (BPS) and the 3'splice site (SS)] are also indicated. In addition, another sequence, the polypyrimidine tract, which is located between the BPS and 3'SS, is shown. N stands for any nucleotide, R for a purine and Y for a pyrimidine. **(B)** Assembly and disassembly of the major spliceosome. The spliceosomal snRNPS are depicted in coloured circles.

1.2.1 Exon Junction Complex (EJC)

The Exon Junction Complex is a multiprotein complex that is assembled ~20-24 nucleotides upstream of splice junctions in a splicing-dependent manner (Le Hir et al., 2001b), but sequence-independent. Since its discovery in 2000, a number of proteins have been named as EJC components.

The composition of the EJC is thought to evolve as the mRNA travels from the nucleus to the cytoplasm (Degot et al., 2004; Kataoka et al., 2001; Kim et al., 2001b; Le Hir et al., 2001a; Le Hir et al., 2001b). The EJC components could be divided into the core components that remain bound to mRNAs in the nucleus and in the cytoplasm and a group of more transiently associated proteins that associate with the mRNAs at different steps. The core complex is comprised of MLN51 (Metastatic Lymph Node 51) or Barentsz, the heterodimer MAGOH:Y14 and eIF4A3 that constitutes a platform onto which the rest of the EJC is assembled (Ballut et al., 2005; Chan et al., 2004; Degot et al., 2004; Ferraiuolo et al., 2004; Kataoka et al., 2000; Le Hir et al., 2001a; Le Hir et al., 2001b; Palacios et al., 2004; Tange et al., 2005).

The second set of factors, those that are transiently associated, includes SRm160, RNPS1, Acinus, SAP18 and Pinin (related to splicing), UAP56, ALY/REF and TAP-p15 (related to mRNA export) and the proteins UPF3 and UPF2 (NMD factors) (Le Hir et al., 2008a; Le Hir et al., 2008b; Tange et al., 2004). RNPS1 and UPF3 remain associated to the mRNA both in the nucleus and the cytoplasm.

The EJC has been implicated in a number of nuclear and cytoplasmic processes. It increases the efficiency of mRNA export in *Xenopus laevis* oocytes (Le Hir et al., 2001b) and enhances translation in both *Xenopus* oocytes and mammalian cells (Nott et al., 2004; Wiegand et al., 2003). In addition, several EJC protein components are required for the correct localization of *oskar* mRNA and antero-posterior axis formation during early embryonic development in *Drosophila melanogaster* (Hachet et al., 2001; Mohr et al., 2001; Palacios et al., 2004). All the accumulated evidence define EJC as a key regulator of mRNA function in higher eukaryotes (Le Hir et al., 2003; Tange et al., 2004).

The EJC has been reconstituted *in vitro* from recombinant proteins (Ballut et al., 2005). It was shown that the four proteins, MAGOH-Y14, eIF4A3 and the SELOR fragment of MLN51 (Barentsz) form a highly stable complex that requires the presence of RNA and ATP.

The crystal structure of the EJC core shows the organization of the four proteins around the RNA (Figure 5) (Andersen et al., 2006; Bono et al., 2006). From this structure, important insights can be obtained on how the MAGOH-Y14 inhibits eIF4A3 helicase activity thus locking the EJC on the RNA. Moreover, it becomes apparent that MLN51 stimulates the activity of eIF4A3.

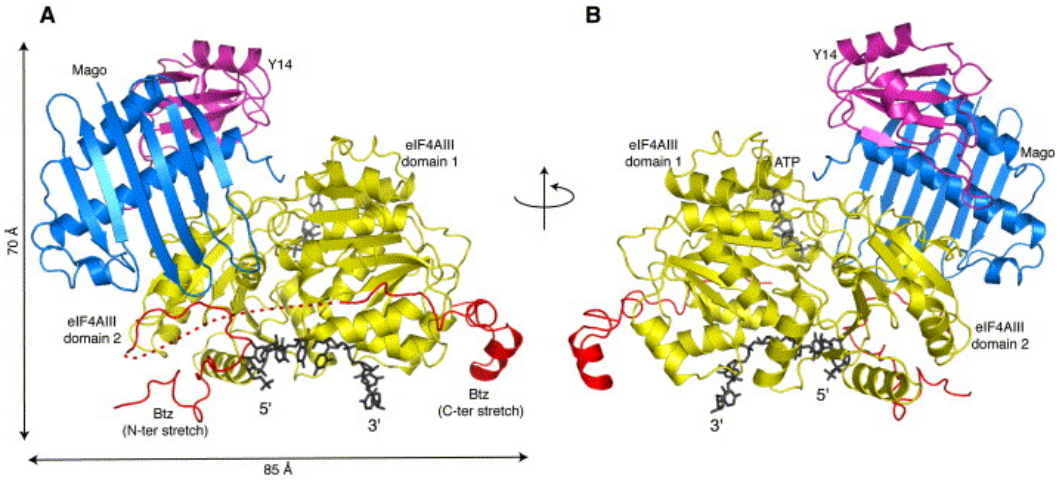


Figure 5. Crystal structure of the EJC. In this figure, the human EJC is shown in two orientations (the structure shown in (A) is rotated 180° around a vertical axis). The Barentsz protein (Btz; shown in red) stretches around the DEAD-box helicase eIF4A3 (in yellow). Both proteins interact with RNA (in black), which is bound at a cleft formed between the two RecA-like domains of eIF4A3. ATP (in gray) binds at an interface between the two domains of eIF4A3, distinct from the RNA binding cleft. The other two protein components of the EJC, Mago (blue), and Y14 (magenta), bind mainly to domain 2 of eIF4A3, but the interaction surface also extends over to the interface with domain 1. The figure was taken from (Bono et al., 2006). All ribbon drawings were rendered using PyMOL (DeLano, W.L., 2002, http://www.pymol.org).

Trying to elucidate the hierarchy of the EJC assembly, the analysis of purified H, C, and mRNP complexes revealed that at least one EJC component, ALY/REF, can interact with pre-mRNA prior to spliceosome assembly. In mammals the H complex is nonspecific, assembles first, it does not require splice sites or ATP and consists of hnRNP proteins (Bennett et al., 1992; Konarska et al., 1986). On the other hand, Y14, MAGOH, RNPS1, UAP56, and SRm160 are found in intermediate-containing spliceosomes. The proteins ALY/REF, Y14 and MAGOH remain stably bound to spliced mRNA, while the association of RNPS1, UAP56, and SRm160 is destabilized at the step of exon ligation (Reichert et al., 2002). Kataoka and colleagues developed a simple whole cell lysate system for *in vitro* splicing, to study the EJC assembly during splicing. Three of the EJC components, ALY/REF, RNPS1 and SRm160, are found on pre-mRNA by the time the spliceosome is formed, whereas Upf3b associates with splicing intermediates during or immediately after the first catalytic step of the splicing reaction (cleavage of exon 1 and intron-lariat formation). In contrast, Y14 and MAGOH, which remain stably associated with mRNA after export to the cytoplasm, join the EJC during or after completion of exon-exon ligation. These findings give the first hint that EJC formation is an ordered pathway, that involves a stepwise association of components and is coupled to specific intermediates of the splicing reaction (Kataoka et al., 2004). The exact assembly pathway of the core EJC, during the two distinct reactions involved in splicing, remains to be clarified.

A number of questions about the EJC is still unanswered. One of the most interesting questions is when/how the EJC is removed from the mRNA and what is the destiny of the stripped-off mRNA factors, possibly recycled to be used at the next cycle. It has been shown that at least Y14 remains associated with mRNAs in the cytoplasm until they are translated. According to the current model in literature, translation is responsible for the removal of Y14 from mRNAs, and thus of the EJC (Dostie et al., 2002).

1.3 mRNA nuclear export

In eukaryotic cells, the processes of the nucleus are separated from those in the cytoplasm by a double membrane, the nuclear envelope. The transport between the nucleus and the cytoplasm is mediated by receptors that recognize specific cargoes and carry them through the nuclear pore complex (NPC). All known receptors are members of the karyopherin/importin-β family of proteins that function by interacting with the small GTPase Ran (Mattaj et al., 1998). The karyopherin/importin-β family members and Ran do not seem to be involved in the export of spliced mRNA. mRNA export is employing the TAP/p15 heterodimer (Mex:Mtr2 in yeast) (Braun et al., 2001; Guzik et al., 2001; Katahira et al., 1999).

Nuclear export of mRNA is tightly associated with the controlled remodelling of mRNP complexes (Cole et al., 2006a, b). A great number of different proteins bind to mRNA as it progresses to form a mature mRNP in the nucleus. Many of these proteins are associated with certain steps of the gene-expression pathway, such as transcription, splicing or polyadenylation, and are generally removed before nuclear export. Apart from those, many proteins remain bound to the mRNA and accompany mRNPs to the cytoplasm, although not all are involved directly in nuclear export, such as the exon junction complex (EJC). The precise way in which completion of mRNP maturation is recognized is still unclear. Binding of Mex67:Mtr2 is necessary for export (Braun et al., 2001; Guzik et al., 2001; Katahira et al., 1999). Apart from Mex67, which has an mRNA-binding domain, adaptor proteins such as REF/Aly/Sub2, EJC components and SR (Serine/Arg rich) proteins, also promote its binding to mRNPs (Huang et al., 2005; Moore, 2005) .

Translocation of mRNP complexes is thought to be mediated primarily by weak interactions between the TAP/p15 heterodimer and the FG-nucleoporins, NPC proteins that harbour Phe-Gly repeats. Transport complexes generally move through the NPC by simple diffusion. The function of the carrier is to facilitate equilibrium of the transport complex between the inner and the outer compartments by a passive mechanism that prevents the return to the donor compartment (Cole et al., 2006a). The mechanism by which TAP/p15 is removed from the mRNP is still unclear. In budding yeast, it has

been observed that the DEAD-box helicase Dbp5 remodels mRNPs at the cytoplasmic NPC side by removing Mex67 (the yeast homologue of TAP). Its ATPase activity is activated by Gle1, an essential mRNA export factor and an inositol hexaphosphate (IP_6) (Alcazar-Roman et al., 2006; Weirich et al., 2006).

1.3.1 Human TREX (TRanscription-EXport) Complex

TREX is a protein complex conserved from yeast to metazoans (Gatfield et al., 2001; Jensen et al., 2001; Luo et al., 2001; Strasser et al., 2000, 2001; Stutz et al., 2000; Zhou et al., 2000). This conserved complex provides one of the best examples of coupling different mechanisms in yeast versus metazoans, as it is associated with the transcription machinery in yeast and the splicing machinery in humans (Reed et al., 2005).

The human TREX complex contains the proteins UAP56 and ALY/REF (known as Sub2 and Yra1 in yeast) as well as the multisubunit THO complex (Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002). The human THO complex is comprised of hTho2, hHpr 1, f SAP79, f SAP35 and f SAP24. The components of the TREX complex in yeast, *Drosophila melanogaster* and mammals are shown in detail in table 1. In yeast, Sub2 and Yra1 are co-transcriptionally recruited to nascent transcripts by the THO complex (Abruzzi et al., 2004). In contrast, the human TREX complex has been suggested to be loaded onto mRNA during splicing (Masuda et al., 2005). RNA interference and biochemical studies in metazoans and genetic analyses in yeast indicate that the conserved TREX complex functions in mRNA export (Gatfield et al., 2003).

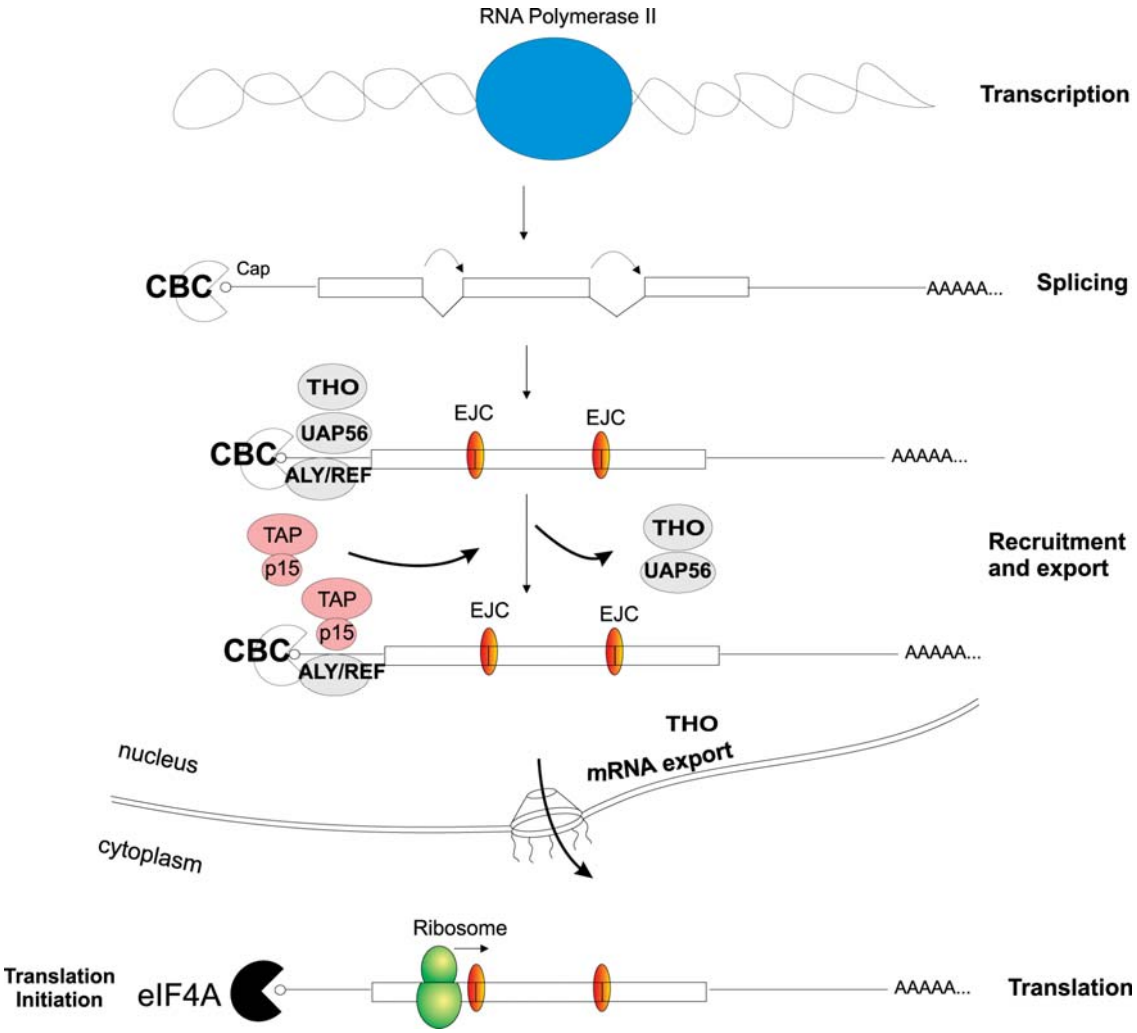


Figure 6. Splicing-coupled mRNA export in metazoans. For metazoan, several mRNA export models have been described. In this figure, the splicing- and cap-dependent modes of human TREX recruitment to the mRNP are depicted. Export receptors are indicated in pink and adaptors in grey.

The deposition of the human TREX complex has been addressed in a number of studies. In previous work, UAP56 and ALY/REF were identified as components of the exon junction complex (EJC) (Gatfield et al., 2001; Le Hir et al., 2001b; Le Hir et al., 2000). The EJC was also reported to contain the mRNA export receptor TAP/p15 (Le Hir et al., 2001b; Tange et al., 2004).

Masuda and coworkers identified the human THO complex and showed that it is associated with spliced mRNA, but not with unspliced pre-mRNA *in vitro*. The human THO complex co-localizes with splicing factors in nuclear speckle domains *in vivo*. Considering that splicing occurs co-transcriptionally in humans, their data indicate that recruitment of the human TREX complex to spliced mRNA is not directly coupled to transcription, but is instead coupled to transcription indirectly through splicing. Studying the requirement of the ALY/REF-UAP56 interaction, they found that ALY/REF interacts with UAP56-C terminus as efficiently as with full-length UAP56, indicating that the C terminus is sufficient for ALY/REF binding. Both UAP56 and the hTHO complex bind at the C terminus of ALY/REF (Masuda et al., 2005).

In a more recent study, Cheng and co-workers sought to determine whether the human THO complex is also deposited on the mRNA as part of the EJC. This analysis revealed that the human TREX complex is recruited in a splicing- and cap-dependent manner to a region near the 5'end of the mRNA. Their data indicate that ALY/REF binds to the mRNA closest to the cap and THO complex, while UAP56 binds downstream of ALY/REF but upstream of the EJC. In their model, the recruitment of the TREX complex to mRNA occurs via an interaction between the TREX component-ALY/REF and the cap-binding protein CBP80. While the TREX complex is bound only on the first exon, the EJC is recruited at every exon-exon junction independently of the 5' cap/CBP80. In addition, the above paper suggests that the cap has a role in mRNA export only for spliced mRNAs and not for cDNA transcripts or random RNAs (Cheng et al., 2006). Moreover, analysis of ALY/REF mutants identified a mutant that lacks 16 C-terminal amino acids, totally disrupting the binding to UAP56. This mutant still binds to CBP80. The recruitment of TREX at the 5′end of mRNA provides a biochemical explanation for electron microscopic studies showing that the mRNPs of the giant Balbiani rings of *Chironomus tentans* translocate through the nuclear pore in a 5′ to 3′ direction (Daneholt, 1997; Mehlin et al., 1992; Visa et al., 1996a). CBC is detected on the 5′end of the Balbiani ring mRNP and is co-exported with it to the cytoplasm (Visa et al., 1996a; Visa et al., 1996b). The Balbiani ring data also indicate that the ribosome associates with the mRNA as it emerges from the nuclear pore (Daneholt, 1997; Mehlin et al., 1992; Visa et al., 1996a). It is already known that

cap/CBC is present on unspliced pre-mRNA (Visa et al., 1996b), yet TREX is not recruited to unspliced pre-mRNA. The cap/CBC is also present on cDNA transcripts, yet TREX is not efficiently recruited to them either. Interesting questions that remain to be answered are whether the cap plays a role in export of naturally intronless mRNAs and whether TREX is recruited to the 5′end. It is thus interesting to investigate the contribution of the EJC to the deposition of TREX complex.

Yeast TREX Complex	***Drosophila* TREX Complex**	**Human TREX Complex**
Sub 2	UAP56	UAP56
Yra 1	REF	ALY/REF
Tex1	TEX1	hTex1
Yeast THO Complex:	***Drosophila* THO Complex:**	**Human THO Complex:**
Mft1	-	-
Thp2	-	-
Tho2	THO2	hTho2
Hpr1	HPR1	hHpr1
-	THOC5	f SAP79
-	THOC6	f SAP35
-	THOC7	f SAP24

Table 1: Members of the conserved TREX complex in yeast, *Drosophila melanogaster* and human are listed.

1.4 mRNA first round of translation: decision to translate or to degrade

In the nucleus of mammalian cells, the mRNA 5'cap is recognized by the cap binding complex, which is a heterodimer consisting of a small (CBP20) and a big (CBP80) subunit (Mazza et al., 2001). The 3'poly-A tail is associated to the nuclear poly-A-binding protein (PABPN1). Once the mRNA is exported to the cytoplasm, CBC interacts with translation initiation factor 4G (eIF4G), which is required for the recruitment of the small ribosomal subunit. When the small ribosomal subunit is recruited, it initiates the 5' to 3'scanning of the 5'UTR, searching for an AUG start codon (Lejeune et al., 2004). Once the start codon is identified, the large ribosomal subunit is recruited and the 80S translation-competent complex is formed. During translation, a number of nuclear-acquired factors associate at the mRNA ORF (open reading frame) are dissociated. In addition to the remodeling, premature translation termination codons (PTCs) are recognized. Once a termination codon is found, the presence of an EJC downstream is assessed. If both, a termination codon and an EJC are present, the mRNA is degraded by a mechanism called nonsense-mediated mRNA decay (NMD).

Contrary to the first round of translation, steady-state translation carries out the bulk of cellular protein synthesis. The mRNAs do not carry EJCs at their exon-exon junctions, CBC is replaced with the major cytoplasmic cap-binding protein eIF4E and PABPN1 is replaced by the cytoplasmic poly-A-binding protein PABP.

1.5 mRNA quality control and nonsense-mediated mRNA decay

Mutations in a transcript can lead to severe consequences in the regulatory network of mRNA metabolism. Eukaryotic cells have developed a number of quality control mechanisms that recognize defective mRNAs and eliminate them. The crucial role of these surveillance mechanisms is underlined by the finding that several human diseases are linked to defective mRNA metabolism (Hentze et al., 1999).

1.5.1 Nonsense-mediated mRNA decay limits the expression of truncated polypeptides

Nonsense-mediated mRNA decay (NMD) is an RNA surveillance pathway that recognizes and degrades mRNAs carrying a premature translation termination codon (PTCs). In order for a termination codon to be perceived as premature, it has to occur >50-55 nucleotides upstream of the last exon-exon boundary (Figure 7). In mammalian cells, NMD is linked to splicing. For the activation of NMD in mammals two signals are required. The first signal is the existence of a PTC. The physiological termination codon and the PTC have the same sequence; therefore a way needs to be found to distinguish the two types of stop codons. Once a termination codon is recognized, the current hypothesis states, that the ribosome scans the mRNA in order to find the exon junction complex (EJC) at a downstream exon-exon junction. If the EJC is present, NMD is activated and the mRNA is degraded (Figure 8).

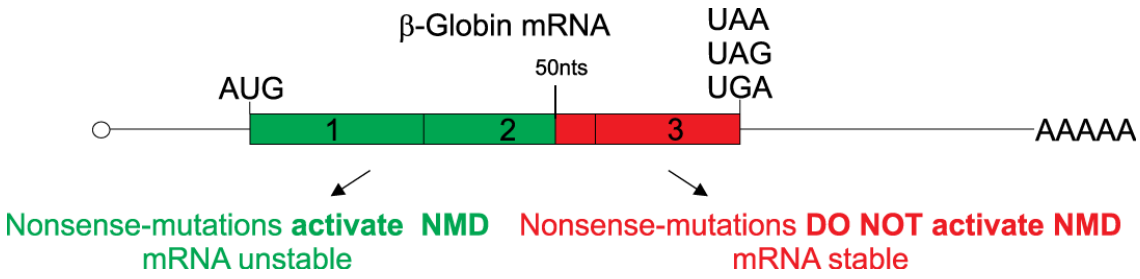


Figure 7. PTC definition in mammals. The figure depicts a mRNA with three exons. According to this "50-55 nt rule", a stop codon located in the green region would trigger NMD, while a stop codon situated in the red region would be considered physiological and would not trigger NMD.

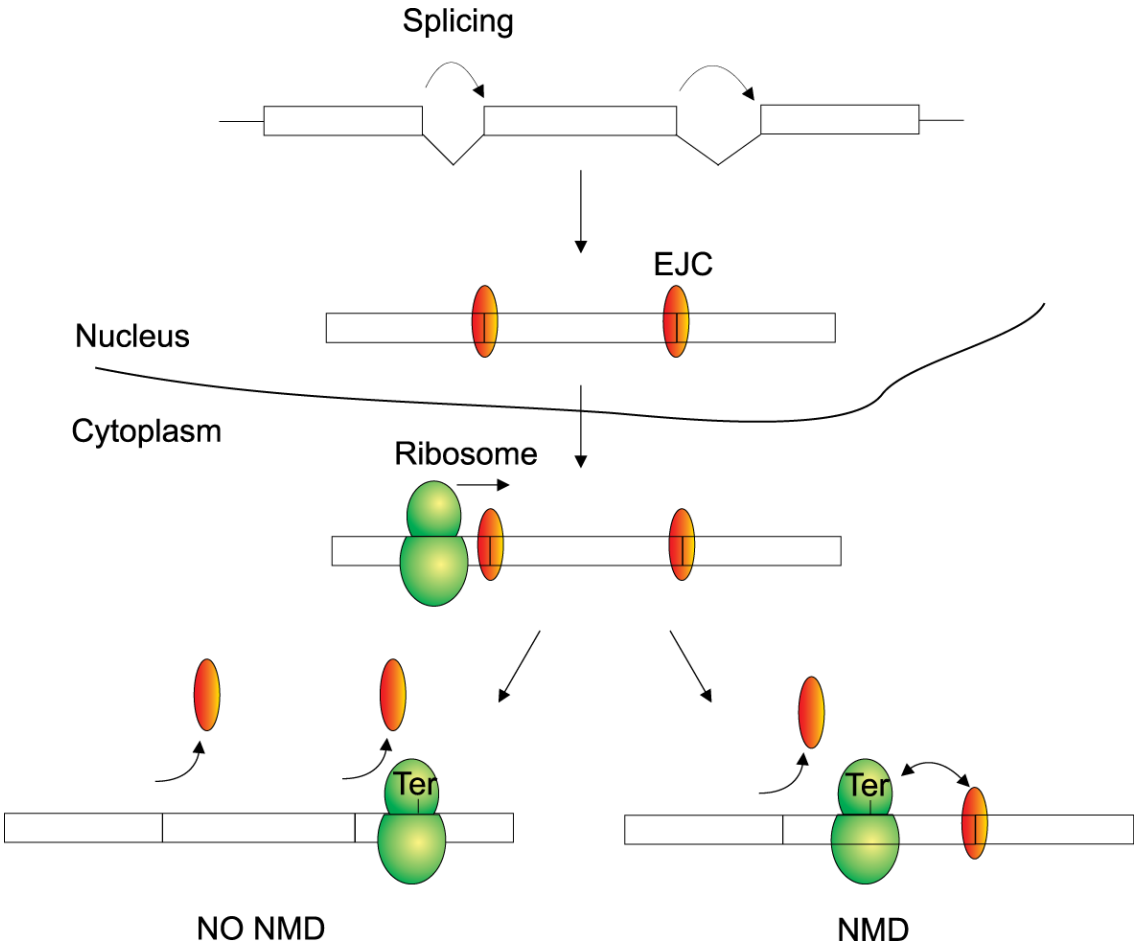


Figure 8. Model illustrating mammalian NMD. Splicing in the nucleus deposits a protein complex 20 to 24 nts upstream of each exon-exon junction, the exon junction complex (depicted with the orange oval shape). The exon junction complex is thought to recruit further NMD factors. Upon transport through the nuclear pore, the EJC is remodeled and several factors are released. During the first round of translation, EJCs are removed by the elongating ribosome. If the translating ribosome encounters a PTC (premature termination codon) at least 50-55 nts upstream of an exon-exon junction, NMD is triggered. Otherwise, the ribosome removes the termination codons and NMD is not activated.

1.6 Aims of the Study

Profound knowledge of the molecular mechanisms of protein deposition on the RNA during splicing will improve our understanding of mRNP formation. In particular, the elucidation of the deposition of the EJC and TREX complex will help to establish the connection between splicing and downstream processing mechanisms of mature RNA in mammals.

Although the EJC is central in the mechanism of mammalian NMD, little is known about the exact assembly line of it. Another important question is whether EJCs of differential composition exist.

The aim of this thesis was to study the protein composition of the mRNPs focusing on the EJC and the TREX complexes. The first goal was to affinity purify different spliced mRNPs and investigate the composition of the EJC. To reach this goal, the GRNA affinity purification method was chosen as an experimental system. The aim was to affinity purify mRNPs of different substrates spliced *in vitro*. This approach will be used to compare the potential different EJC subcomplexes assembled on different introns, such as MINX and β-globin.

The second goal was to study the assembly and disassembly of the EJC. The third goal was to investigate the deposition of TREX. Several models of mRNA export and TREX deposition exist in the literature. There are multiple suggestions of what dictates the TREX recruitment, for instance cap-dependency, splicing-dependency or EJC-dependency. I particularly wanted to test the contribution of splicing and the EJC to the recruitment of TREX factors.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals

All salts and standard chemicals were purchased from Roth (Karlsruhe, Germany) and Sigma-Aldrich (Taufkirchen, Germany). Cell culture media were purchased from Gibco BRL div. of Invitrogen (Karlsruhe, Germany). Exceptions to this statement are the following:

Agarose	Biozym Scientific (Oldendorf, Germany)
Bacto-agar	Becton, Dickinson and Co. (Sparks, USA)
Bromophenolblue	Serva (Heidelberg. Germany)
$CaCl_2$ (calcium chloride)	ICN Biomedicals (Ohio, USA))
Complete EDTA-free protease inhibitor	Roche (Mannheim, Germany)
DTT (dithiothreitol)	Promega (Madison, USA)
Formaldehyde	Merck (Darmstadt, Germany)
NaAc (sodium acetate)	Merck (Darmstadt, Germany)
Nucleotides (NTPs) and deoxy-nucleotides (dNTPs)	MBI Fermentas (Burlington, Canada) or Promega (Madison , USA)
Trizol (TriReagent)	MRC (Cincinnati, USA)
Tris-base	Merck (Darmstadt, Germany)

2.1.2 Enzymes

Restriction enzymes were purchased from New England Biolabs (NEB) (Ipswich, MA, USA), together with the appropriate buffers. SP6, T7 and T3 RNA-polymerases were purchased from Roche (Mannheim, Germany). PCR reactions were performed with Phusion Polymerase (Finnzymes, Espoo, Finland) or Taq Polymerase (NEB).

2.1.3 DNA oligonucleotides

DNA oligonucleotides were ordered from Biomers (Ulm, Germany). PCR primers were obtained in desalted, lyophilized form, and were diluted in H_2O.

2.1.4 Antibodies

The antibodies and corresponding dilutions that were used in this study are listed below.

Antibody	**Properties**	**Supplier**	**Dilution**
αBARENTSZ	Polyclonal, rabbit	Produced by us	1:1000
αFLAG	Polyclonal, rabbit	Sigma-Aldrich	1:2000
αMAGOH	Monoclonal, mouse	AbCam, UK	1:2000
αeIF4A3	Polyclonal, rabbit	Produced by us	1:1000
αRNPS1	Polyclonal, rabbit	Krainer laboratory	1:4000
αUPF3b	Polyclonal, rabbit	Lykke-Andersen laboratory	1:5000
αY14	Polyclonal, rabbit	Izaurralde laboratory	1:1000

αUAP56	Polyclonal, rabbit	AbCam, UK	1:2000
αTHOC4	Polyclonal, mouse	AbCam, UK	1:1000
α mouse IgG peroxidase coupled	Polyclonal, goat	Sigma-Aldrich	1:5000
α rabbit IgG peroxidase coupled	Polyclonal, goat	Sigma-Aldrich	1:5000

2.1.5 Kits

The following kits were used throughout the study. Unless stated otherwise, the manufacturer's recommendations were followed.

Application	**Name**	**Supplier**
Plasmid DNA extraction from agarose gels, PCR reactions or enzymatic digests	QIAquick™ Gel Extraction Kit	Qiagen, Hilden, Germany
DNA preparation from *E. coli*	QIAprep™(mini, midi and maxi) Kit	Qiagen
In vitro synthesis of large amounts of capped RNA	mMESSAGE mMACHINE®	Ambion
Silver Staining	SilverQuest™ Silver Staining Kit	Invitrogen
Affinity Resin for antibodies	AminoLink® Plus Immobilization Kit	Thermo Scientific Pierce

2.1.6 Instrumental material

Agarose gel electrophoresis equipment	Peqlab Biotechnologie (Erlangen, Germany)
Acrylamide gel (SDS-PAGE) electrophoresis equipment	Whatman-Biometra (Goettingen, Germany)
Acrylamide gel (with urea) electrophoresis equipment	Owl Separation Systems (Portsmouth,USA)
Blotting (Trans-Blot SD semi-dry transfer cell)	Bio-Rad (Hercules, USA)
Photometer (Biophotometer)	Eppendorf (Hamburg, Germany)
Ultrasonic processor Model VC130PB (Sonicator)	Sonics & Materials Inc. (Newtown, USA)
Thermocycler (T3000)	Whatman-Biometra (Goettingen, Germany)
PhospoImager FLA-3000	FujiFilm-LifeScience (Düsseldorf, Germany)

2.1.7 Commonly used buffers, solutions and media

All solutions were prepared with double deionized water.

Antibiotic solutions (1000x)	100 mg/ml ampicillin in H_2O 50 mg/ml kanamycin in H_2O
Anode buffer (Immunoblot)	25 mM Tris
BBS (2x) for transfection	50 mM BES (N,N-bis(Hydroxyethyl)-2-Aminoethansulfonat) 280 mM NaCl 1.5 mM Na_2HPO4 pH 6.96 (adjusted with 10 N NaOH) filter sterilized (0.22 µm)
$CaCl_2$ for transfection	2.5 M $CaCl_2$
Cathode buffer (Immunoblot)	25 mM Tris-HCl 40 mM 6-amino-4-hexanoic acid 0.01% (w/v) SDS
Coomassie stain solution:	For 1 L: 40% (v/v) methanol 10% (v/v) acetic acid 2.5 g Coomassie Brilliant Blue (Sigma-Aldrich)
Coomassie destain solution:	10% (v/v) methanol 10% (v/v) ethanol
DNA loading buffer (6x)	0.9% (w/v) bromophenol blue (Sigma-Aldrich) 60% (v/v) glycerol 60 mM EDTA pH 8.0

Glutathione Buffer (Elution buffer for GRNA Chromatography)	50 mM L-Glutathione reduced minimum 99% 500 mM Tris pH 8.0 1 mM DTT
IP buffer 1 (GRNA Chromatography)	500 mM NaCl 20 mM Tris-HCl pH 7.5 1.5 mM $MgCl_2$ 0.05% (v/v) Nonidet® P40 (Igepal CA-630) 8.7% (v/v) Glycerol 12 µg/ml Heparin
IP buffer 2 (Flag Affinity Purification)	150 mM NaCl 20 mM HEPES pH 7.9 0.1% (v/v) Nonidet® P40 (Igepal CA-630) 1 mM $MgCl_2$ 0.2% (v/v) Triton X 100 0.05% (w/v) deoxycholate
Luria-Bertani (LB) agar (autoclaved)	15 g/l Bacto agar in LB medium
LB medium (autoclaved)	10 g/l Bacto tryptone 5 g/l Bacto yeast extract 5 g/l NaCl pH 7.6 (adjusted with 5 N NaOH)
Ponceau staining solution	0.5% (w/v) Ponceau S dye 1% (v/v) acetic acid
Resolving gel (8% to 12%) solution	375 mM Tris-HCl, pH 8.8 8 to 12 % (w/v) acrylamide/bisacrylamide 0.1% (w/v) SDS

RNase A	3.5 mg/ml RNase A in RNase digestion buffer 50% (v/v) glycerol
RNA loading buffer for *In vitro* splicing assay	80% formamide (Sigma-Aldrich) 0.1% (w/v) xylene cyanol 0.1% (w/v) bromophenol blue 2 mM EDTA, pH 8.0
SDS Sample Buffer (6x)	350 mM Tris-HCl, pH 6.8 600 mM DTT 30% (v/v) glycerol 10% (w/v) SDS 0.012% (w/v) bromophenol blue
SDS Running Buffer (Laemmli buffer) (5x)	1.5 % (w/v) Tris 7.2% (w/v) Glycine 0.5% (w/v) SDS
Stacking gel (5%) solution	125 mM Tris-Cl, pH 6.8 5 % (w/v) acrylamide/bisacrylamide 0.1 % (w/v) SDS
TBE (10x)	0.89 M Tris base, pH 8.3 0.89 M boric acid 20 mM EDTA, pH 8.0
TBS	20 mM Tris-HCl, pH 7.6 137.5 mM NaCl
TBS-Tween	20 mM Tris-HCl pH 7.6 137.5 mM NaCl 0.05% (v/v) Tween 20

Total cell lysis buffer (for protein isolation)	0.5 % Triton X-100 1x Complete protease inhibitor (Roche) 30 µg/µl RNase A In TBS
TE buffer	10 mM Tris-Cl 1 mM EDTA pH 7.5

2.1.8 HeLa Nuclear Extract

The HeLa 5.0 x 10^9 nuclear extract used for the *in vitro* splicing reactions was purchased from CilBiotech (Mons, Belgium).

2.1.9 Eukaryotic cell lines

In the present study, two cell lines were used HeLa and HEK 293 cell lines. The HeLa cell line originates from a human cervical carcinoma. The HEK 293 cells were purchased from Invitrogen and their complete name is GripTite™ 293 MSR Cell Line. This cell line is a genetically engineered Human Embryonic Kidney (HEK 293) cell line that expresses the human macrophage scavenger receptor and strongly adheres to standard tissue culture plates for dependable results. Developed from a 293-H subclone, GripTite™ 293 MSR cells show the same fast cell growth, high transfection efficiency, and high-level expression characteristics of the parental 293-H cells.

2.1.10 Plasmids

The following plasmids were used in the present thesis: pCI-neo (Promega), pGEM-4Z (Promega), pGEM7 (Promega) and pBlueskriptSK II+ (Stratagene) are the basic plasmids used to clone all the constructs used in the present study. Each construct was generated as shown in the table below. All PCR amplified inserts were validated by sequencing (GATC Biotech).

Name	Features
pGEM4- MINX	MINX (Bell et al., 2002) was cloned EcoRI/BamHI into pGEM4.
pGEM4- MINX (15)	Generated by PCR-mutagenesis.
pGEM4- MINX Δintron	Generated by PCR-mutagenesis.
pGEM4- MINX GG	Generated by PCR-mutagenesis.
pGEM4- MINX GG (15)	Generated by PCR-mutagenesis.
pGEM4- MINX+30	Generated by PCR. The primer apart from the restriction enzyme site included a 30nts-sequence of β-globin.
pGEM4- MINX Δi+30	Generated by PCR. The primer apart from the restriction enzyme site included a 30nts-sequence of β-globin.
pGEM7- MINX	MINX sequence was amplified from pGEM4-MINX and cloned Sac I/Xma I.
pGEM7- MINX-4BoxBs	The BoxBs sequence was inserted Xma I/ Xho I to the above construct.
pCI neo-Flag	Created by inserting the Flag-sequence into the NheI/XhoI sites of pCI-neo vector (Gehring et al., 2003).

pCIneo-FlagY14	The sequence of Y14 was subcloned from a pCI-λN-Y14 construct into the XhoI site of pCIneo-Flag (Gehring et al., 2003).
pCIneo-FlagRNPS1	RNPS1 sequence was amplified from HeLa cDNA and cloned like Y14 (Gehring et al., 2003).
pCIneo-FlageIF4A3	Gehring et al., 2009
pCIneo-FlagMAGOH	Gehring et al., 2009
pCIneo-FlagBarentsz	Gehring et al., 2009
pCIneo-FlagCBP80	CBP80 sequence was amplified from HeLa cDNA and cloned into pCI-neo FLAG.
pCIneo-FlagUAP56	Cloned by PCR amplification of the respective ORFs from HeLa cDNA and ligation to pCIneo-Flag construct.
pCIneo-FlagUAP56 ΔC171	Generated by site directed PCR mutagenesis.
pCIneo-FlagUAP56 ΔN251	Generated by site directed PCR mutagenesis.
pCIneo-FlagUAP56K95N	Generated by site directed PCR mutagenesis.
pCIneo-FlagDDX39	Cloned by PCR amplification of the respective ORFs from HeLa cDNA and ligation to pCIneo-Flag construct.
pCIneo-FlagALY/REF	Cloned by PCR amplification of the respective ORFs from a plasmid obtained from Origene and ligation to pCIneo-Flag construct.
pCIneo-FlagALY/REF ΔC53	Generated by site directed PCR mutagenesis.
pCIneo-FlagALY/REF ΔC16	Generated by site directed PCR mutagenesis.
pCIneo-FlagALY/REF ΔN80	Generated by site directed PCR mutagenesis.

2.1.11 Sequences of oligonucleotides used in PCR reactions

DNA oligonucleotides were ordered from Biomers (Ulm, Germany). PCR primers were obtained in desalted, lyophilised form and were dissolved in water. In addition, those primers used to amplify pre-mRNA were HPLC-purified.

Name	Sequence	Restriction Site
Minx-Upstream	TTT TTC GAG CTC CCA CTC TTG GAT CGG AAA CCC GTC GGC CTC CGA ACG GT	Sac I
Minx-Downstream	TTC CCC CGG GAG AGT TTG TCC TCA ACC GCG AGC TGT GGA AAA AAA A	Xma I
β-globin Upstream	TTT TTC GAG CTC GCA AGG TGA ACG TGG ATG AAG TTG GTG GTG AGG CCC TG	Sac I
β-globin Downstream	TTT TTC CCC CGG GCC TCT GGG TCC AAG GGT AGA CCA CCA GCA GCC T	Xma I
MS2Box Upstream	CCC CCC GGG GGG GTT CCC TAA GCT CGA CCA AAG G	Xma I
MS2Box Downstream	CCC TCG AGG GAG TTT AGT ACT CGA GGT CGA GTG CA	Xho I
MINX+30 Upstream and MINX Δi+30 Upstream	TTTTGAATTCTGTGTTCACTAGCAAC CTCAAACAGACACCGAGCTCGCCCA CTCTT	EcoRI
MINX+30 Downstream and MINX Δi+30 Downstream	TTTTGGATCCTAATACGACATCACTA TAGGGC	BamHI

2.2 Methods

2.2.1 Bacterial Techniques

2.2.1.1 Bacterial strains

For the propagation of plasmid DNA, the *E. coli* strain XL1-Blue from Stratagene (La Jolla, CA, USA) was used. The genotype of this strain is the following: hsdR17, supE44, recA1, endA1, gyrA96, thi, relA1, lac/F´[proAB^{+} lacI^{a}, lacZΔM15:Tn10(Tet^{R})].

2.2.1.2 Preparation of transformation-competent *E.coli*

For the preparation of transformation competent *E.coli* XL1-Blue, bacteria were streaked onto LB agar and incubated at 37°C over night. A single colony was inoculated into 5 ml LB medium and allowed to proliferate over night at 37°C. One milliliter of the bacterial suspension was transferred to 100 ml LB medium and incubated at 37°C until the suspension reached an OD600 of 0.6. The culture was cooled on ice, and the bacterial cells were pelleted by centrifugation at 1,200 g for 5 minutes in a pre-cooled rotor. The bacterial pellet was resuspended in 10 ml of ice-cold TSS, aliquoted in pre-cooled reaction tubes and stored at -80°C. Aliquots were tested by transformation with serial dilutions of plasmid DNA for transformation efficiency of at least 106 colonies per μg of plasmid DNA.

2.2.1.3 Transformation of bacteria

One hundred nanograms of plasmid DNA or 10 µl of ligation reaction were mixed with 100 µl of transformation-competent *E.coli*, incubated on ice for 20 minutes and heat-shocked for 60 seconds at 42°C. The suspension was chilled on ice, 700 µl of LB medium was added and incubated for 40 minutes at 37°C. The transformed bacteria were then plated on LB agar plates containing the appropriate antibiotic(s) and incubated at 37°C over night.

2.2.1.4 Isolation of plasmid DNA from bacteria

Plasmids were propagated in *E.coli* XL1-blue. Single colonies were picked from an agar plate and incubated in 5 ml LB medium containing the appropriate antibiotic for mini-preparation or 100 ml LB medium for maxi-preparation of plasmids. Plasmid DNA was isolated using Qiagen plasmid purification kits according to the manufacturer's instructions. The DNA concentration was spectrophotometrically measured at 260 nm (A260) in water. An A260 value of 1 corresponds to a double-stranded DNA concentration of 50 µg/ml. The plasmids were also visualised by a BrEt-stained agarose gel.

2.2.3 DNA techniques

2.2.3.1 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction was used to generate DNA fragments for cloning. A typical reaction contained 100 ng of template DNA, 0.4 µM of the forward and reverse primers, dNTPs (Roth) at a final concentration of 200 µM each, as well as 1 unit of

Phusion™ High Fidelity Polymerase (Finnzymes), 1x Phusion HF reaction buffer, in a final volume of 50 µl. PCR thermocycle protocol was as follows: initial denaturation for 5 minutes at 95°C, followed by 25 cycles of 20 seconds denaturation at 95°C, 20 seconds annealing at 57-60º C (depending on primers' Tm) and 30 seconds/kb extension at 72°C. A final extension was performed for 7 minutes at 72°C. Reactions were performed in a T3 Thermocycler (Whatman Biometra, Göttingen, Germany).

2.2.3.2 DNA Extraction from Agarose Gels

DNA fragments generated by PCR or by restriction digestion were purified depending on their sizes on 0.8% to 1.5% agarose gels. Subsequently, excised DNA fragments were extracted using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. DNA was eluted from the silica-based purification column with 40 µl H_2O and stored at -20°C.

2.2.3.3 Restriction Digests

Preparative digests of agarose gel-purified PCR products to generate suitable ends for cloning were performed at 37°C over night using 2 U of the appropriate restriction enzymes in a suitable restriction buffer (NEB) supplemented with 1xBSA (2 µg in a reaction volume of 20 µl) where required. Preparative digest of plasmid DNA was performed at 37°C for 3 hours with 2 µg of DNA and 1 U of each restriction enzyme in its corresponding buffer. Analytical digests of plasmid DNA were performed at 37°C for 1.5 hours with 0.5 µg of DNA and 0.5 U of the appropriate enzymes with their corresponding buffer. Restriction fragments were visualized by agarose gel electrophoresis and gel-purified as described in 2.2.3.2.

2.2.3.4 Ligation

Restriction digested vectors and inserts (either PCR generated or excised from a plasmid) were ligated in a molar ratio of 1:5 with 1 U of T4 DNA Ligase (Promega) in 1x ligase buffer (Promega). The ligation reaction was incubated for 2 hours at room temperature, and transformed into competent *E.coli* as described in 2.2.1.3.

2.2.3.5 DNA Sequencing

DNA sequencing was performed by the company GATC Biotech. The following primers were used for sequencing:

Name	**Sequence**
SP6	CATACATTTAGGTGACACTATAG
T7	TAATACGACATCACTATAGGGC
T3	CTTATCATGTCTGCTCGAAGC

2.2.4 RNA techniques - RNA analysis by « *in vitro* splicing»

2.2.4.1 *In vitro* Transcription

Radiolabeled mRNAs for *in vitro* splicing were obtained by *in vitro* transcription. One microgram of the linearized plasmid was transcribed with the SP6 RNA-polymerase (Roche) with 50 μCi [α^{32}-P]GTP (800 mCi/mmol; NEN) for 45 minutes at 40°C in the presence of nucleotides (500 μM ATP, CTP, UTP and 50 μM GTP) and 1 mM DTT (Promega). 1 mM of cap analog was added to the reaction (Promega). After the 45

minutes of incubation, DNase I digestion was performed for 15 minutes at 37 °C. The RNA was extracted using Trizol and chloroform.

2.2.4.2 *In vitro* Splicing

SP6 transcripts were prepared as described in the respective protocol. The pre-mix (ingredients used for the recycling of ATP) contained 0.1µl of 270 mM $MgCl_2$, 0.1 µl of 10 mM ATP, 0.4 µl of 500 mM phosphocreatine and 0.2 µl RNAsin (Promega). The next step was the preparation of splice mix, which included 0.88 µl of pre-mix, 5.0 µl of HeLa Nuclear Extract (NE), 2.0 µl of 12% PVA, 0.55µl of 400 mM Hepes buffer with pH 7.3 and 0.55 µl of RNAsin. To each reaction of final volume 10 µl, 1.0 µl (1µl = 10,000 CPM) of radiolabeled transcript was added. The samples were kept on ice at all times during the preparation. The samples were incubated at 30°C for 2 to 3 hours (depending on the transcript used). The RNA was extracted using Trizol and chloroform.

2.2.4.3 RNase H digestion of spliced products

For RNase H digestions, MINX and MINX Δintron (Δi) transcripts were incubated under splicing conditions (at 30^{o}C) for 120 minutes. For this assay MINX and MINX Δi transcripts with a 30 nts longer first exon were used (MINX+30 and MINX Δi+30). Endogenous RNase H was activated by adding a cDNA oligonucleotide (1 mM) complementary to the -36 position to the splicing reactions and incubating them for 15 minutes at 30^{o}C. RNAs were extracted and analyzed as described above. The sequence of the cDNA oligonucleotide was 5'-GAGTGGGCGAGC-3'. The oligonucleotide binds from position -42 to -31.

2.2.5 Protein Techniques

2.2.5.1 Protein extraction

For the preparation of HeLa extract, 1,400,000 HeLa cells were plated in a 10 cm dish. The next day, each plate was transfected with 4 μg of selected plasmid and 2 μg of GFP. 12-16 hours later, the transfection efficiency was checked using the fluorescent microscope and the cells were washed twice with TBS and incubated for another day at 37 ^{o} C. The cells were harvested in the afternoon of the following day. Each plate was washed once with PBS and 800 μl of PBS were used to collect the scraped cells from the dish. The cells were pelleted by centrifugation of 2 minutes at 1,400 rpms. The supernatant was discarded and the pellet was diluted in 600 μl of lysis buffer (0.5% Triton X100, complete protease inhibitor and 30 μg/ μl RNase A in TBS). The resulting samples were vortexed and placed at -80 ^{o} C overnight.

2.2.5.2 Protein concentration measurement

The protein concentration of extracts was determined with the protein assay supplied by BioRad in a spectrophotometer (BioPhotometer, Eppendorf, Hamburg, Germany) at a wavelength of 595 nm. The protein concentrations were calculated using a BSA standard curve.

2.2.5.3 SDS-Polyacrylamide Gel Electrophoresis

Proteins were separated according to their size by analytical sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a minigel system (BioRad). To this end, samples were heated with the detergent SDS and the reducing agent DTT. The sample passed the upper stacking gel to concentrate the proteins before entering the lower resolving gel. Stacking and 8-, 10- or 12% resolving gels were polymerized by mixing the respective solutions with 0.1% ammonium peroxodisulfate (APS) and 0.04% (resolving gel) or 0.1% (stacking gel) N,N,N',N'-tetramethyl-ethylendiamine (TEMED). Protein samples were mixed with 6x SDS sample buffer and heated for 5 minutes at 95°C prior to loading. Electrophoresis was carried out in 1x Laemmli buffer at constant current of 30 mA until the desired resolution was reached as judged by the separation of molecular weight markers (NEB and Fermentas). Subsequently the gel was processed for immunoblotting.

2.2.5.4 Immunoprecipitation of proteins

The samples, from the protein extraction, were thawed and the insoluble material was pelleted by centrifuging at 10,000 rpms for 10 minutes. From the resulting supernatant, an aliquot of 60 µl was kept for loading control; the rest of the supernatant was used for immunoprecipitation. For the Flag-Immunoprecipitation, 30 µl of FLAG beads (ANTI-FLAG M2-Agarose from mouse from Sigma) were washed with 500 µl of Lysis Buffer. As a next step, the 600 µl of HeLa extract was added to the agarose and the samples were rotated for 2 hours at 4 ^{o} C. The samples were washed three times with 900 µl of wash buffer (0.5% Triton X100 in TBS). After the 3rd wash, the supernatant buffer was removed and 800 µl of wash buffer were added to the "dry" beads. The solution of beads and wash buffer was transfered to the spin columns (Micro Bio-Spin Chromatography Columns from BIORAD). The flow-through was discarded and the bound protein complexes were eluted with 60 µl of 1x SDS loading buffer by

centrifuging 1 minute at 1,000 g. Samples were then analyzed by SDS-PAGE and immunoblot.

2.2.5.5 Immunological Detection of Proteins on a Solid Support - Immunoblot

Immunoblot analysis was performed using 10-30 µg of extracts for 8-12% SDS-polyacrylamide gel electrophoresis. Samples were blotted with a semidry method onto Westran nitrocellulose membrane (Whatman Schleicher & Schuell) pre-equilibrated in methanol and sandwiched together with the gel between 8 Whatmann 3MM papers equilibrated in either cathode or anode buffer. This sandwich (depicted in the following scheme, figure 9) was then mounted into a transblot semidry transfer chamber (BioRad) and a current (1 mA/cm^2 of gel) was applied for 1-1.5 hours. The membrane was blocked for unspecific binding with 5% non-fat skimmed milk in TBS-Tween (0.1%) for 1 hour on a shaking platform. Next, the blot was incubated overnight with shaking at 4°C with the primary antibody diluted in TBS-T containing 5% milk. The blot was washed three times (20 minutes each) with TBS-T, incubated for 1 hour with secondary, horseradish peroxidase coupled antibody (Sigma-Aldrich) diluted in TBS-T containing 5% milk. Again, the blot was washed three times (45 minutes each) with an excess of TBS-T. Blots were developed with ECL Western Blotting Detection Reagent or ECL plus Western Blotting Detection Reagent (Amersham Pharmacia Biotech). The specific primary antibodies and secondary antibodies used in this study are listed with their corresponding dilutions in section 2.1.4. Signals were visualized on autoradiograph ECL film (Amersham Pharmacia Biotech).

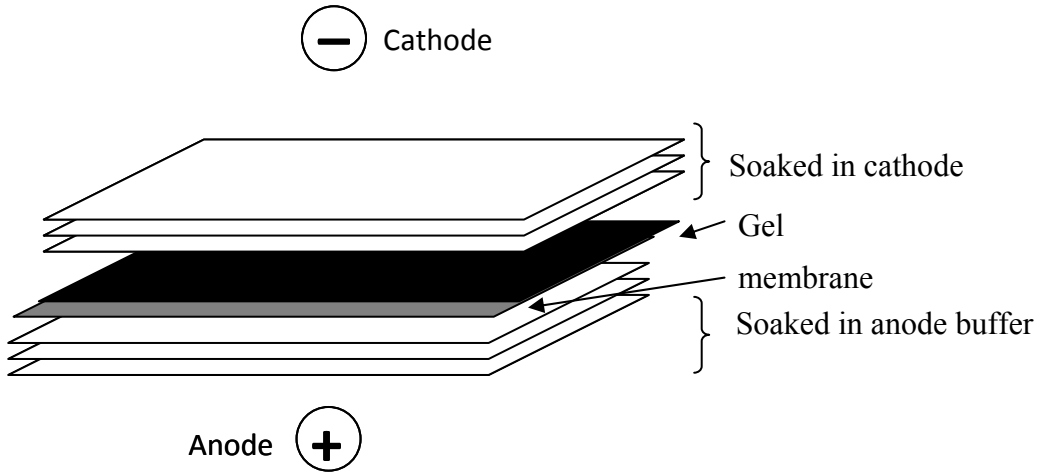


Figure 9: Transfer assembly for Immunoblot

2.2.5.6 Protein Preparation and Dialysis

This protocol describes a four-day procedure. During the first day, a 5 ml LB culture with the necessary antibiotics was inoculated with a single bacterial colony. This small colony was grown overnight at 37^{o} C with shaking. The next day, a big culture (450 mL) with ZYM 5052 Medium, containing the appropriate antibiotics and the 5 mL culture was started. It was left to grow at 37^{o} C for 4 hours shaking. Once the 4 hours had elapsed, the culture was transferred to an incubator of 25^{o} C and it was left to grow overnight shaking. The third day, the big culture was divided into 4 aliquots. These aliquots were centrifuged for 15 minutes and the pellet was kept. If the pellet was pure, the next step was followed. In case the pellet seemed inpure, it was washed with 10 mL 1X PBS, placed in a 50 mL falcon and centrifuged for 10 minutes. The PBS was discarded and 10 mL of lysis buffer (20 % General Buffer (100 mM NaCl in 1X PBS), 0.002% Triton X 100, 0.2 mM PMSF and 0.8 mM DTT) were added to the resulting pellet. The samples were vortexed and incubated on ice for 30 minutes. While incubating on ice, the samples were vortexed from time to time. The pellets were

sonicated for 5 minutes at settings 7/30/pulse and centrifuged for 15 minutes. The resulting supernatants were kept for the following steps, while the pellets were discarded. In the meantime, the protein purification column was prepared by adding 1mL Glutathione Sepharose, the slurry was left to flow through by gravity. Then 1 mL of 1X PBS was added and the resulting solution was mixed by pipetting. The column was allowed to empty by gravity and this washing step was repeated once more. The bottom lid of the column was placed, then the column was loaded with the supernatant (approximately 10 ml), and at the end the top lid was used to close the column. The column was rotated (during this step the protein binds to the sepharose) at 4 ^{o} C for 1 hour 30 minutes. The sepharose was washed twice with wash buffer (80% of General Buffer, 0.008% Triton X 100 and 0.8 mM DTT). The column was allowed to empty by gravity and for the elution 1 mL of glutathione buffer was added. Once the elution buffer was added, the column was mixed by inverting it and it was incubated for 5-10 minutes at room temperature. The eluate was collected in an eppendorf tube; the elution step was repeated 5-6 times. The first 2-3 eluates are aliquots with the highest protein concentration. At this point, an SDS PAGE gel was used to check the protein's presence and quality. Once this quality control check point was passed, the protein was dialysed overnight at 4 ^{o} C in the following dialysis buffer (500 mM NaCl and 5% glycerol diluted in 1X PBS). For the dialysis, the D-Tube^{TM} Dialyzer Maxi, MWCO 6-8 kDa from Novagen was used. On the last day, if an extra purification step is required, 2 ml of Ni-NTA Agarose (Qiagen) was added to a purification column. The resin was washed with 3 ml of 1X PBS. The sample from 1st dialysis was added to the column and rotated at 4 ^{o} C for 2 hours. An aliquot of the 1st dialysis before loading sample to the column was kept. After 2 hours of rotation, the resin was washed with 50 ml wash buffer (50 mM NaH_2PO_4, 300 mM NaCl and 20 mM imidazole: Adjust pH value to 8.0 using NaOH). The sample was eluted with 4 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole: Adjust pH value to 8.0 using NaOH). The eluate was collected and dialysed overnight with the same buffer as before. On the 5th Day the twice-dialysed protein was obtained.

2.2.5.7. Coomassie Staining

Disassemble the gel. Place gel in a box and cover with Coomassie stain solution. The coomassie stain solution includes: 40% (v/v) methanol, 10% (v/v) acetic acid and 2.5 g Coomasie Brilliant Blue (per liter). Place the box on a rocker and leave for 1 hour at room temperature. Discard staining solution and replace with destaining solution. Destaining solution includes 10% (v/v) methanol and 10% (v/v) acetic acid. Destaining lasts about 1 hour, shaking at room temperature. After destaining, cover gel with water and shake at room temperature till bands are clear.

2.2.6 Cell culture techniques

2.2.6.1 Propagation of human cell lines

HeLa cells, which originate from a human cervical carcinoma, and HEK 293 deriving from human embryonic kidney cells were maintained in monolayer in 175 cm^2 tissue culture flasks (Sarsted, Nümbrecht, Germany) at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) (Gibco BRL) supplemented with 10% fetal calf serum (Biochrom, Berlin, Germany), 100 U/ml penicillin and 100 µg/ml streptomycin (PAA laboratories, Pasching, Austria). For subculturing, the cells were washed once with PBS and 0.05% trypsin was added and the cells were allowed to detach. 10 ml of cell culture medium was added and the cells were diluted 1:4 with fresh culture medium every 2-3 days. For the HEK 293 cells the addition of Versene (Gibco) was required for the detachment of the cells from the flask.

2.2.6.2 Transfection of Eukaryotic Cells using the BBS/calciumphosphate method

Cells were transfected with Flag-tagged expression plasmids using the calcium phosphate method. HeLa cells were seeded 24 hours before transfection at a density of $1x10^6$ cells per 10 cm cell culture dish in 10 ml medium. Cells were transfected with 4 µg (HeLa cells) or 8 µg (HEK 293 cells) of Flag-protein plasmids, and 2 µg of GFP-expression vector for visual estimation of transfection efficiency. Approximately two hours before transfection, the medium was changed. For each sample, the appropriate DNA mixture was supplemented with water to 450 µl and mixed with 55 µl of 2.5 M $CaCl_2$. Five hundred microliters of 2xBBS were added, mixed by vortexing, incubated at room temperature for 15-20 minutes and applied drop by drop to the cells. The cells were incubated with the precipitate for 16-18 hours at 37°C and 3% CO_2. After removal of the precipitate with TBS, fresh medium was added to the cells. Cells were incubated for additional 24 hours in 5% CO_2 at 37°C, washed with PBS and harvested for analysis.

2.2.6.3 Preparation of Whole Cell Extract (WCE)

1,800,000 HEK 293T cells were plated per 10 cm dish. The next day these cells were transfected with 8 µg of the desired plasmid and 24 hours post-transfectionally were harvested and whole cell extracts were prepared. The transfected cells were washed three times with 10 ml of ice-cold PBS each time. The washed cells were scraped and resuspended in 10 ml of PBS. The cells were pelleted by centrifuging at 3,000 x g for 3 minutes at 4 °C. The supernatant PBS was removed completely without disturbing the cell pellet and the resulting pellet was resuspended in 200 µl of ice-cold Buffer E (20 mM Hepes-KCl, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 10% glycerol, and 1 mM dithiothreitol). The resuspended cells were sonicated using an Ultrasonic processor Model VC130PB (Sonics & Materials Inc., Newtown, PA) at 30% continuous pulse for 5 seconds three times with an 30 seconds incubation on ice between bursts. Then the

samples were centrifuged at 15,000 x *g* for 20 minutes at 4 °C. The supernatant was stored as whole cell extracts at -80 °C. This protocol is an adaptation from the Kataoka and Dreyfuss Protocol (Kataoka et al., 2004).

2.2.7 Affinity Purification

2.2.7.1 GRNA Chromatography

2.2.7.1.1 Production of λN peptide

A small LB culture of 100 ml, containing both ampicillin and kanamycin, was inoculated with a colony (or a bit of the frozen culture) and was grown at 37^{o} C overnight. The half of this culture (50 ml) was used to inoculate a 1 L of LB plus Ampicillin culture. This large culture was incubated at 37^{o} C till the OD reached a concentration of approximately 0.6 (approximately 1 hour). Once the appropriate OD was reached, the culture was induced by adding 0.1 M IPTG and incubating for 4 hours at 30^{o} C. After the induction, the culture was centrifuged at 3,500 rpm for 15 minutes. The pellet was resuspended in 20 ml of lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 0.01% Triton X-100, 2 mM DTT and Proteinase Inhibitor (Roche)-1 tablet for 50 ml). For the disruption of the cells, the lysate was passed through the French Press-EmulsiFlex. The sample was pushed through the press, collected and this step was repeated once more. After passing the lysate through the French press, it was centrifuged at 10,000 rpm for 20 minutes. The resulting supernatant was bound on 3 ml prewashed Glutathione Sepharose beads (Sepharose 4B Amersham), by rotation for 1 hour at 4^{o} C. The beads were subsequently washed 3 times with 10 ml of lysis buffer and centrifuged at 2,000 rpms for 3 minutes. The bound protein was eluted with elution buffer. (25 mM Glutathione in Lysis buffer w/o NaCl). The protein was dialyzed

(Dialysis buffer I: 50 mM Tris p.H. 8.0, 250 mM NaCl, 0.01 TX-100, but no DTT), the buffer was changed 2-3 times with a minimum of 30 minutes each dialysis at 4° C. An aliquot of 2 ml Ni-NTA slurry was washed 3 times with dialysis buffer and the dialyzed protein was incubated with the Ni-NTA beads for 1 hour at 4° C. The beads were washed once with 30 ml of 50 mM Tris pH 8.0, 250 mM NaCl, 0.01% TX-100 and another time with 20 ml of 50 mM Tris pH 8.0, 250 mM NaCl without TX-100. The bound protein was eluted with 8 ml of 400 mM imidazole (Sigma I2399-500G, 68.08) in buffer w/o TX-100. The protein was dialyzed three times against 1 L of dialysis buffer II (50 mM Tris p.H. 7.4, 100 mM NaCl, 20% glycerol and 1 mM DTT) at 4° C. For all the dialysis steps, dialysis membrane spectrum 12-14 kDa from Spectra/Pore Spectrum Laboratories was used. The resulting protein was divided in aliquots of 1 ml and its concentration was measured with the Bradford method. All wash steps were carried out at 2,000 rcf for 5 minutes.

2.2.7.1.2 GRNA Affinity Purification

For the affinity purification of each sample, 30 µl of glutathione sepharose were used and washed 2 times with 1 ml of buffer. After centrifugation at 800 rpm for 1 minute, 150 µl of buffer and 30 µg of λN peptide-GST were added. The samples were rotated at 4 °C for 1 hour. The beads were washed 2 times and then 150 µl of buffer and 30 µl of each splicing sample were added to them. The samples were rotated at 4 °C for 1 hour and 30 minutes. Next, the samples were centrifuged at 800 rpm for 1 minute and the resulting supernatant was discarded. The beads were washed 3 times with 1 ml of buffer, each wash for 10 minutes rotating at 4 °C. After the last wash, the samples were centrifuged and the maximum possible supernatant was removed and discarded. The resulting beads were dissolved in 600 µl of the buffer and loaded on the purification column. The column should empty by gravity. After the column was emptied, another 500 µl of buffer were placed in the column. The columns were centrifuged at 1,200 rpm for 10 seconds. Once all the wash buffer was removed from the column, 400 µl of Trizol were added to the column. The flow through was collected at an eppendorf tube.

At the end, RNA was extracted by adding 80 µl of Chloroform, mixing well and centrifuging for 10 minutes at maximal speed. The resulting upper phase (from TriZol extraction (200 µl) from the centrifugation was mixed with 1.0 volume isopropanol. The samples were left to precipitate for 2 hours or longer till overnight at –20 °C. The next day, the samples were centrifuged at maximal speed for 15 minutes. The supernatant was removed completely. After air drying the formed RNA pellet, the pellet was resuspended in 10µl of RNA loading buffer. Before loading the samples to a 8-10% (depending on the size of the transcript spliced) UREA Gel, they were heated for 5 minutes at 95^{o}C.

2.2.7.2 Flag Affinity Purification

As a first step, splicing reactions of 50 µl volume were assembled. The transcript was spliced in a combination of HeLa nuclear extract and HEK 293 whole cell extract in a ratio of 1:25. The splicing was left to proceed for 2 hours. For the Flag affinity purification, 12 µl of Flag agarose were used for each sample. After washing the Flag agarose one time with 300 µl of Flag IP buffer (20 mM Hepes pH 7.9, 200 mM NaCl, 0.2% TX100, 0.1% NP40, 0.005% deoxycholate and 2 mM $MgCl_2$), 220 µl of Flag IP buffer and 30 µl of the splicing reaction were added to the beads. From the original splicing reaction, an aliquot of 10 µl was kept and RNA was extracted from it (serving as the splicing input, control for the splicing efficiency). The samples were rotated for 1 hour and 30 minutes at 4^{o} C. Then, the agarose was washed with 600 µl of Flag IP buffer for 3 times, each wash lasted for 10 minutes. After the final wash, 500 µl of buffer were used to resuspend the agarose and the solution was loaded to columns (Receiver columns 20 µm from Macherey-Nagel). The column was left to empty by gravity and once the initial liquid went through, another 400 µl of buffer were added to wash the column. After a short centrifugation of 30 seconds at 0.8 rpm, 400 µl of Trizol were added to the column. The flow through was kept and to that 80 µl of chloroform were added for RNA extraction. The samples were stored at -20^{o}C overnight. The next day the samples were centrifuged at maximal speed for 15 minutes, to pellet RNA. The

pellet was resuspended in 10 μl of RNA loading buffer. The samples were heated at 95^{o}C for 5 minutes and then cooled on ice for 1-2 minutes. The samples were loaded on a 8-10% UREA gel (depending on the size of the transcript), to analyze the affinity purification outcome. The gels were dried for 2 hours and 30 minutes and then they were exposed on PhosporImager screens and on autoradiography films to visualize the results.

2.2.8 Production of antibodies against Exon Junction Complex components

2.2.8.1 Expression and Purification of Proteins-Antigens

Name	Tag	Resistance	Product Size (including tag)
1. pET45-MAGOH-His	His	amp	~ 18 kDa
2. pET32-eIF4A3	His	amp	~ 70 kDa
3. pET45-PYM-His	His	amp	~ 30 kDa
4. pGEX6P1-Upf3b300	GST	amp	~ 52 kDa
5. pGEX6P1-BTZ359	GST	amp	~ 55 kDa
6. pGEX6P1-RNPS1 153-245	GST	amp	~ 35 kDa
7. pGEX6P1-Y14	GST	amp	~ 50 kDa

Table 2. Vectors used for protein expression.

The expression vectors were given to the EMBL Protein Expression and Purification Core Facility. The BL21(DE3) and BL21(DE3)RIL cells were transformed with vectors 1-7. The transformed samples were plated on LB-Amp (LB Amp/Cam for RIL-cells) and were incubated at 37°C overnight. The next day, a small culture of 3 ml in LB-Amp (-/+Cam) was initiated and was incubated at 37°C, shaking at 250 rpm until OD600 ~0.4. When the expected OD was reached, the temperature was lowered to 18°C, the culture was induced with 0.2 mM IPTG and was grown overnight at 18°C. The 2 ml of each culture were harvested and resuspended in resuspension buffer (+lysozyme, +DNase) for His- or GST-purification. They were lysed for 3 minutes in sonication bath and were centrifuged for 10 minutes at 14 krpm. The supernatant was incubated with 10 µl of appropriate resin for 30 minutes. The resin was washed three times with wash buffer and the bound protein was eluted with 30 µl of 2x SDS-buffer.

For a larger amount of protein, 25 ml of LB with antibiotic were inoculated and were incubated overnight at 37°C shaking at 250 rpm. A large culture of 2 L LB with appropriate antibiotic was inoculate at 1:100 with pre-culture and was incubated at 37°C shaking at 180 rpm until OD600 ~ 0.4-0.6. When the expected OD was reached, the temperature was lowered to 18°C. The culture was induced with 0.2 mM IPTG and was incubated overnight at 18°C. The cells were harvested and stored at -20°C.

Purification of His6-tagged proteins #1-#3

The pellets were resuspended on ice with lysis buffer (50 mM Tris/HCl, pH 8, 500 mM NaCl, 20 mM imidazole+ Roche complete inhibitor), and were disrupted in the cold room with two passes through the French Press. The samples were centrifuged for 45 minutes at 45 krpm, 4°C. The supernatant was filtered through 0.45 µm filter (pellet was washed and was kept at -20°C), and was loaded on equilibrated HiTrap Chelating HP column (Ni2+, 5ml). The loading speed was 0.3-0.5 ml/min. The column was washed with wash buffer (50 mM Tris/HCl, pH 8, 500 mM NaCl, 20 mM imidazole)

with ~4 column volumes. The bound sample was eluted in elution buffer (50 mM Tris/HCl, pH 8, 500 mM NaCl, 500 mM imidazole).

Purification of GST-tagged proteins #4-#7

The pellets were resuspended on ice with lysis buffer (PBS, pH 7.4 + Roche complete inhibitor) and were disrupted in the cold room with two passes through the French Press. The samples were centrifuged for 45 minutes at 45 krpm, 4°C. The supernatant was filtered through a 0.45 µm filter and was loaded on an equilibrated HiTrap GST-column (5 ml). The loading speed was 0.5 ml/min. The column was washed with wash buffer (PBS) with ~4 column volumes and was eluted in elution buffer (PBS, pH 7.4, 20 mM GSH).

2.2.8.2 Immunisation of the animals

The immunisation of the animals and the bleeding were conducted by Eurogentec S.A. (Liege). The custom polyclonal antibody production program was followed and two New Zealand white rabbits (SPF) were used per protein.

3. RESULTS

3.1 Affinity selection of mRNPs spliced *in vitro* to investigate potential substrate specific differences in the composition of EJCs

The exon junction complex (EJC) is assembled ~20-24 nucleotides upstream of splice junctions in a splicing-dependent, but sequence-independent manner (Le Hir et al., 2000). Little is known about what dictates the composition of the EJC. Trying to elucidate what influences the composition of the EJC, our hypothesis was that the intron downstream of the assembled EJC could influence its composition. Gehring and colleagues tested whether all EJC components share the same NMD cofactor requirements. To address this question they performed tethering assays (tethering of the EJC proteins Y14, MAGOH, UPF3b, RNPS1, or eIF4A3) together with siRNA-mediated depletion of UPF1, UPF2, or Barentsz. Their data suggested that the EJC components Y14, MAGOH, and eIF4A3 can activate NMD in an UPF2-independent manner, while RNPS1-induced NMD requires UPF2. The EJC proteins Y14 and MAGOH depend on Barentsz and eIF4A3 for activation of NMD, whereas RNPS1 does not. Therefore two branches were suggested: the RNPS1/UPF2-dependent and the Barentsz-dependent branch. The question I wanted to address was whether this heterogeneity of NMD-activating mRNPs is reflected into EJC assemblies.

To address this question, the efficient splicing of different NMD substrates and the specific and sensitive detection of EJC components are required. Exon junction complexes can be assembled on suitable mRNA substrates by performing *in vitro* splicing reactions (see section 3.2). The resulting mRNP consists of the spliced mRNA, the EJC and additional RNA-binding proteins. The goal was to affinity purify these potentially differential mRNPs and compare their protein composition, focusing on the EJC.

The first step of the experimental strategy involved *in vitro* splicing of the substrates (MINX and β-globin). A number of optimization steps were required for the establishment of optimal splicing conditions for the substrates.

3.1.1 Establishment of efficient *in vitro* splicing conditions

In vitro splicing represents a key technique required for the experiments described in this thesis. The first experimental goal was to optimize the splicing efficiency of the substrates.

One of the transcripts selected as a substrate for *in vitro* splicing was MINX. This transcript is derived from the human adenovirus late transcription unit and was used to set up the optimal *in vitro* splicing conditions presented in this thesis. It was chosen for the following reasons: 1) it has been shown that it is efficiently spliced *in vitro* (Bell et al., 2002; Merz et al., 2007) and 2) its intron activates NMD *in vivo* when introduced into a mRNA at an appropriate position (Thermann et al., 1998).

To establish the optimal conditions for the *in vitro* splicing of the MINX transcript, the following conditions from the literature served as a starting point; a total reaction volume of 10 µl containing 6 µl of nuclear extract, 2 µl of 12% (w/v) polyvinyl alcohol, 2.4 mM of $MgCl_2$, 0.5 mM of ATP and 20 mM of phosphocreatine was incubated for 90 minutes at 30^{o}C (Tarn et al., 1994). Hepes buffer (220 mM) was added in the splicing reaction (Krainer et al., 1984).

The most important components for an *in vitro* splicing reaction are: an optimal volume of splicing-competent extract, an optimized ionic strength and the reaction incubation time. The reaction was optimized by systematically varying the concentrations of one reaction component, while holding the concentration of the remaining components constant. The titrations included different Hepes (220 mM) pH values ranging from 7.0 to 7.9, $MgCl_2$ concentrations ranging from 0.675 mM to 5.4 mM and phosphocreatine concentrations ranging from 10 mM to 40 mM (Figure 10). When the optimal buffer conditions were found (Figure 10: A lane 3, B lane 3, C lane 4), the amount of extract used in the splicing reaction and the incubation time

parameters were tested. With time, more and more spliced product is accumulated, thus incubation times ranging from 30 to 300 minutes were tested. It was observed that MINX was spliced up to 85% after 120 minutes (Figure 11 A, lane 4). 30% of the transcript input has already been spliced after 30 minutes of incubation at 30°C (Figure 11 A, lane 2). A number of titration experiments were performed to reach the maximum splicing efficiency of this transcript, which reached 90%. This percentage refers to the amount of transcript input converted into spliced product at a given time.

Figure 12 shows that a single reaction ingredient can influence the splicing efficiency of MINX. Splicing reactions were performed using the optimal buffer, incubation time and amount of extract conditions. At each panel, one reaction ingredient (stated in the triangle) is varied (Figure 12).

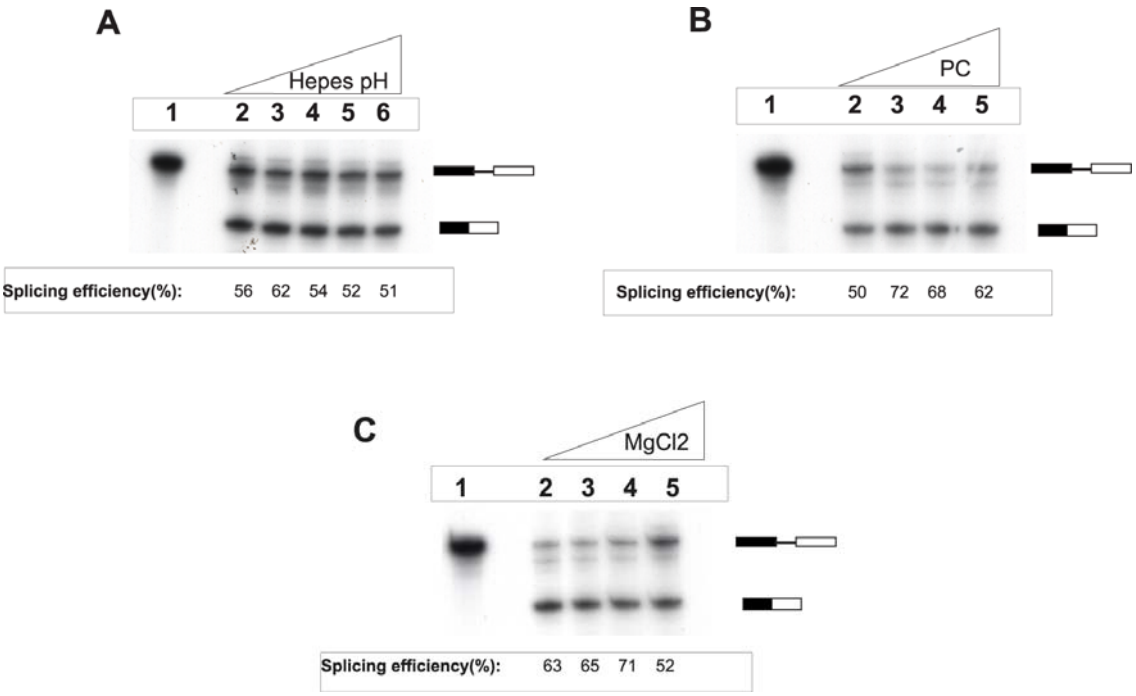


Figure 10. Optimization of buffer conditions for *in vitro* splicing. The figure depicts a set of optimization experiments. The experiments shown in this figure were performed using reaction conditions stated in the literature, and the only parameter varied was the component that was tested (stated in the triangle). **(A)** Splicing reactions were performed in a range of Hepes buffer

(220 mM) pH values: 7.0-7.3-7.5-7.7 and 7.9. The optimal Hepes buffer pH value was 7.3 (lane 3). **(B)** Splicing reactions were performed with increasing phosphocreatine concentrations: 10 mM-20 mM-30 mM and 40 mM. The optimal phosphocreatine concentration was 20 mM (lane 3). **(C)** Splicing reactions were performed with increasing concentrations of $MgCl_2$: 0. 675 mM-1.35 mM- 2.7 mM and 5.4 mM. The optimal $MgCl_2$ concentration was 2.7 mM (lane 4).

Figure 11. MINX is efficiently spliced after 120 minutes of incubation in nuclear extract. Transcript inputs (A and B, lane 1) and different reaction conditions tested (A, lanes 2-4 and B, lanes 2-5) are depicted. **(A)** Splicing reactions were incubated at 30 ^{o}C for 30, 60 and 120 minutes (lanes 2-4 respectively). **(B)** Splicing reactions were performed in 2.5 µl- 5 µl - 7.5 µl and 10 µl of HeLa nuclear extract (lanes 2-5 respectively) for 120 minutes.

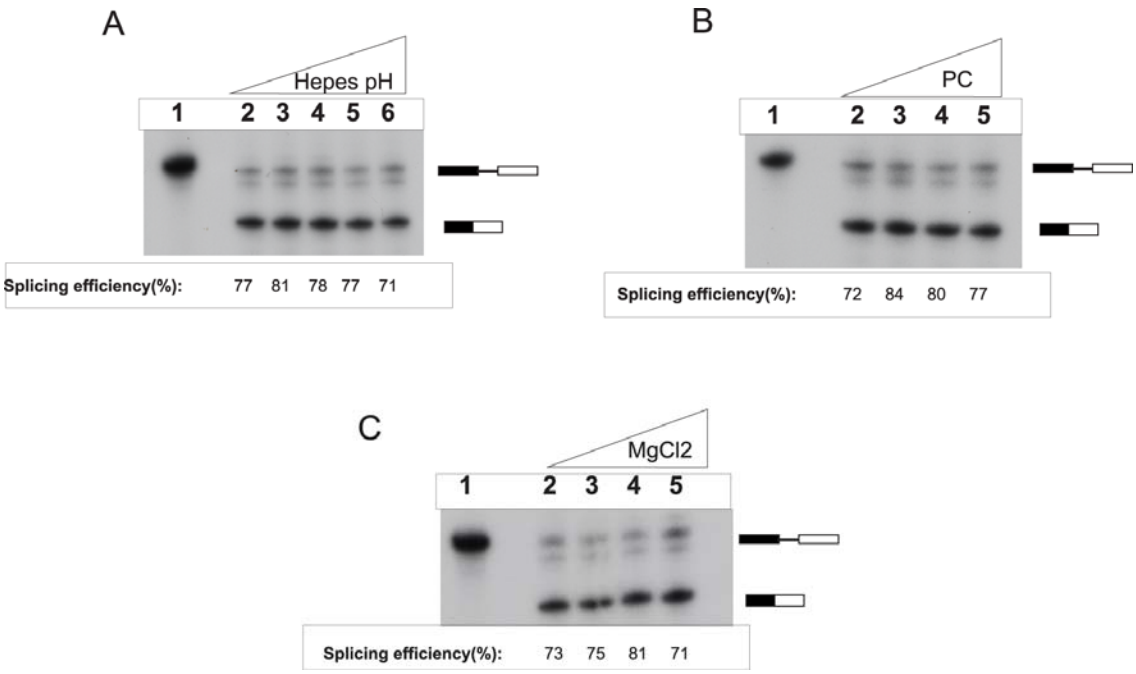


Figure 12. A single ingredient can influence the *in vitro* splicing efficiency of MINX. Radiolabeled MINX transcript is spliced *in vitro* under different reaction conditions. Each buffer component was tested using the optimal reaction conditions described in figure 10 and 11, with one additional variable (Hepes buffer pH value, $MgCl_2$ concentration or PC concentration). **(A)** Splicing reactions were performed in a range of Hepes buffer (220 mM) pH values from 7.0 to 7.9 (7.0-7.3-7.5-7.7 and 7.9). The optimal Hepes buffer pH value was 7.3 (lane 3). **(B)** Splicing reactions were performed in increasing concentrations of phosphocreatine: 10 mM-20 mM-30 mM and 40 mM. The optimal phosphocreatine concentration was 20 mM (lane 3). **(C)** Splicing reactions were performed in increasing $MgCl_2$ concentrations: 0. 675 mM-1.35 mM- 2.7 mM and 5.4 mM. The optimal $MgCl_2$ concentration was 2.7 mM (lane 4).

The optimal reaction conditions for splicing MINX transcript *in vitro* were found to be the following: a total reaction volume of 10 µl containing 5 µl of HeLa nuclear extract, 2.7 mM of $MgCl_2$, 0.1 mM of ATP, 20 mM of phosphocreatine, 2 µl of 12% (w/v) polyvinyl alcohol and 220 mM of Hepes buffer with a pH value of 7.3, which was incubated for 120 minutes at 30^{o}C.

Transcripts / Conditions	Minx	B-globin
Incubation time: 60, 120, 180, 240, 300 minutes	✓	✓
$MgCl_2$ concentration: 0.675 mM, 1.35 mM, 2.7 mM, 5.4 mM	✓	✓
Hepes buffer (220 mM) pH values: 7.0, 7.3, 7.5, 7.7, 7.9	✓	✓
Phosphocreatine concentration: 10 mM, 20 mM, 30 mM, 40 mM	✓	✓
Amount of extract: 2.5 µl, 5 µl, 7.5 µl, 10 µl	✓	✓

Table 3. Summary of conditions tested to achieve optimal *in vitro* splicing reaction conditions. The optimal *in vitro* splicing reaction conditions include: 220 mM of Hepes buffer with a pH value of 7.3, 20 mM of phosphocreatine, 2.7 mM of $MgCl_2$ and 5 µl of splicing-competent nuclear extract.

When the optimal conditions for MINX were achieved, the splicing of the β-globin minigene was pursued. As depicted in figure 13, efficient splicing of β-globin was achieved. The β-globin splicing required longer incubation at 30^{o}C, 180 minutes, compared to the MINX transcript, which required 120 minutes. The human β-globin mRNA is one of the most studied NMD substrates (Neu-Yilik et al., 2001) and it is spliced *in vitro* (Kole et al., 1982). Table 3 summarizes the splicing reaction conditions tested for MINX and β-globin transcripts.

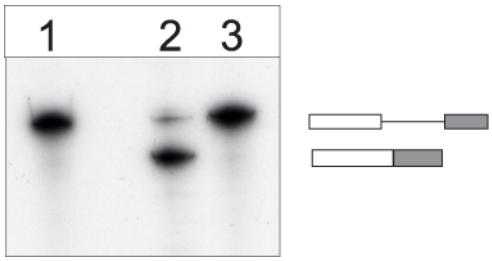


Figure 13. β-globin is efficiently spliced *in vitro*. In lane 1 the transcript input is depicted. β-globin transcript is spliced *in vitro* (lane 2). The splicing reaction was incubated at 30°C for 180 minutes. β-globin transcript was incubated into mock splicing conditions (lane 3). Omitting ATP and phosphocreatine from a splicing reaction abolishes splicing.

3.1.2 GRNA affinity selection of differentially assembled mRNPs spliced *in vitro*

In order to investigate whether the composition of mRNPs varies depending on the RNA substrate, a method to affinity purify these differentially assembled mRNPs from *in vitro* splicing reactions was required. After reaching the optimal *in vitro* splicing conditions for the MINX transcript, the affinity purification method was selected. In order to analyze the protein composition of the EJC, I employed the glutathione-sepharose RNA (GRNA) affinity chromatography that was developed by K. Czaplinski (Czaplinski et al., 2005). Once the optimal conditions for the affinity purification had been found, the protein composition of the MINX and β-globin substrates would be compared. The experimental strategy is shown schematically in figure 14.

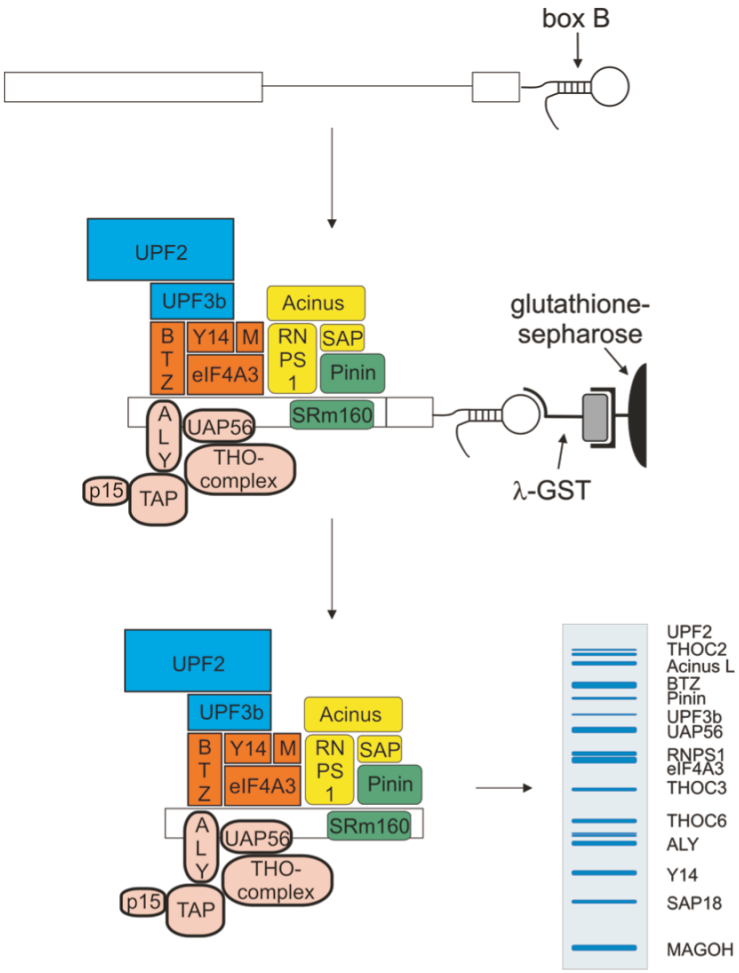


Figure 14. Schematic representation of the experimental strategy used to affinity select mRNPs spliced *in vitro*. The radiolabeled substrate depicted on the top is subjected to *in vitro* splicing in HeLa nuclear extract. The boxB RNA tag is used to affinity purify mRNPs along with the proteins that they carry. As a last step the protein composition can be investigated using either mass spectrometry or immunoblot analysis.

The GRNA technique relies on the insertion of a 19-nucleotide RNA sequence of the λ-phage, called boxB. The boxB sequence is specifically bound by its cognate λN-peptide (a 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N). When this peptide is fused to glutathione-S-sepharose, the resulting module is referred to as λ-GST. λ-GST mediates binding of the RNA to the purification resin (glutathione sepharose) thus allowing affinity purification of the

mRNPs (figure 14, middle panel). After washing, the purified protein complexes are eluted either with an excess of ligand (glutathione) or by digesting the RNA with RNase A. Finally, the eluted proteins can be analyzed by mass spectrometry or immunoblot analysis. The major advantage of this system is derived from the small size of the peptide and its target sequence, which facilitates cloning, its use in biochemical experiments, and diminishes possible interference with the fused protein.

An array of three boxB tags was inserted at the 3'end of the MINX transcript. The specificity of the purification procedure was controlled using a transcript without any boxB tag. The transcripts were transcribed and spliced *in vitro* in HeLa nuclear extract. The affinity selected mRNPs were eluted from the selection resin using glutathione and protein analysis followed. In order to arrive at the final conditions, I extensively optimized the different steps of this complex experimental procedure.

For the initial experiments, as exemplified by the one shown in figure 15, I used unspliced transcripts containing the array of boxB tags, rather than their spliced products.

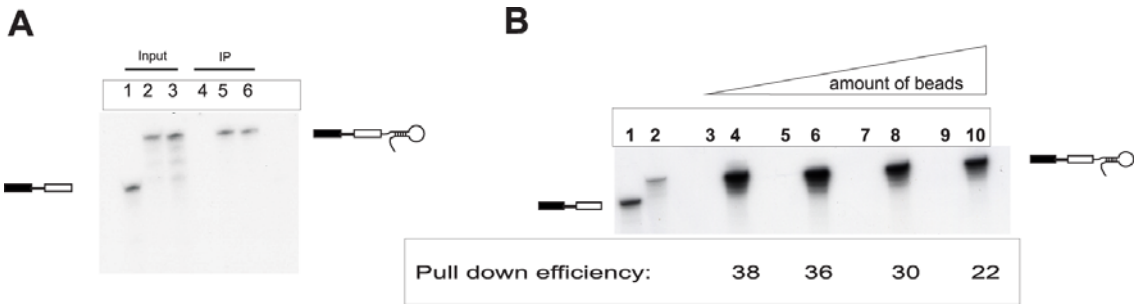


Figure 15. GRNA chromatography specifically affinity purifies MINX pre-mRNAs with cognate boxB tags. (A) An affinity purification experiment performed utilizing unspliced transcripts is shown. Lanes 1-3 depict the transcript inputs, which each contained 1/5 of the amount of transcript that was loaded onto the affinity beads. Lanes 4-6 show the pull-down of the transcripts. From this statement, the pull-down efficiency is estimated to be about 25%. Lanes 2-3 and 5-6 are duplicates of the input and the pull-down of the tagged transcript respectively. **(B)** One of the setting-up experiments is shown. Increasing amounts of affinity resin-beads were tested. Lanes 1 and 2 show the MINX and the MINX carrying an array of boxB tags transcript inputs respectively. Lanes 1, 3, 5, 7 and 9 refer to the MINX transcript and lanes 2, 4, 6, 8 and 10 to the tagged MINX transcript.

One of the initial affinity purifications of unspliced transcripts is shown in figure 15 A. The outcome showed a clear and specific affinity selection of the tagged transcript. The initial pull-down efficiency reached 25%. A transcript without any tag was used as a control. To improve the pull-down efficiency, several setting-up experiments were performed including: titrations of the amount of the affinity resin used, the amount of λN peptide added, incubation times of the resin with the transcripts, number of washing steps and washing buffers. Among the parameters tested to improve the pull-down efficiency was the amount of affinity resin (or as referred to in the figure, beads) (Figure 15 B). As shown previously in figure 15 A, the affinity selection is specific for the tagged transcript (Figure 15 B, lanes 4, 6, 8 and 10). Increasing amounts of affinity resin, 10 µl-20 µl-30 µl and 40 µl, were tested. The pull-down efficiencies are stated at the bottom of the figure 15 B. An amount of 10 µl of affinity resin was able to pull down 38% of the transcript initially loaded onto the beads.

One of the critical parameters of the experimental system was the amount of initial mRNA input that was required to detect specifically bound proteins at the last step of purification. *In vitro* transcription, in the presence of radioactive element α-^{32}P, did not yield high amounts of mRNA. To increase the quantity of pre-mRNA synthesized, the mMESSAGE mMACHINE kit, from the company Ambion, was employed. The high yields of pre-mRNA, resulting from use of this kit, are due to optimized reaction conditions for RNA synthesis in the presence of high nucleotide concentrations. In addition, the enzyme mix, included in the kit, contains an RNase inhibitor that protects the synthesized RNA from degradation. The optimal solution was to splice an amount of pre-mRNA synthesized using the kit, in the presence of a small amount of the radiolabeled transcript [10,000 cpm (counts per minute)]. To achieve the optimal ratio of radiolabeled and non-radiolabeled pre-mRNA, titration experiments were performed. In these experiments, as exemplified by the one shown in figure 16, increasing amounts of non-radiolabeled mRNA were added to splicing reactions where a small amount of radiolabeled transcript was being spliced.

These experiments led to the discovery of the optimal reaction conditions; 200,000 cpm of the radiolabeled transcript and 3µg of non-radiolabeled pre-mRNA were added to a splicing reaction, which had a total volume of 200 µl (these conditions

were used for the experiments shown in figures 17-19 and 21-22). The radiolabeled transcript was used to trace the splicing products, in order to check the splicing and the pull-down efficiencies, using autoradiography.

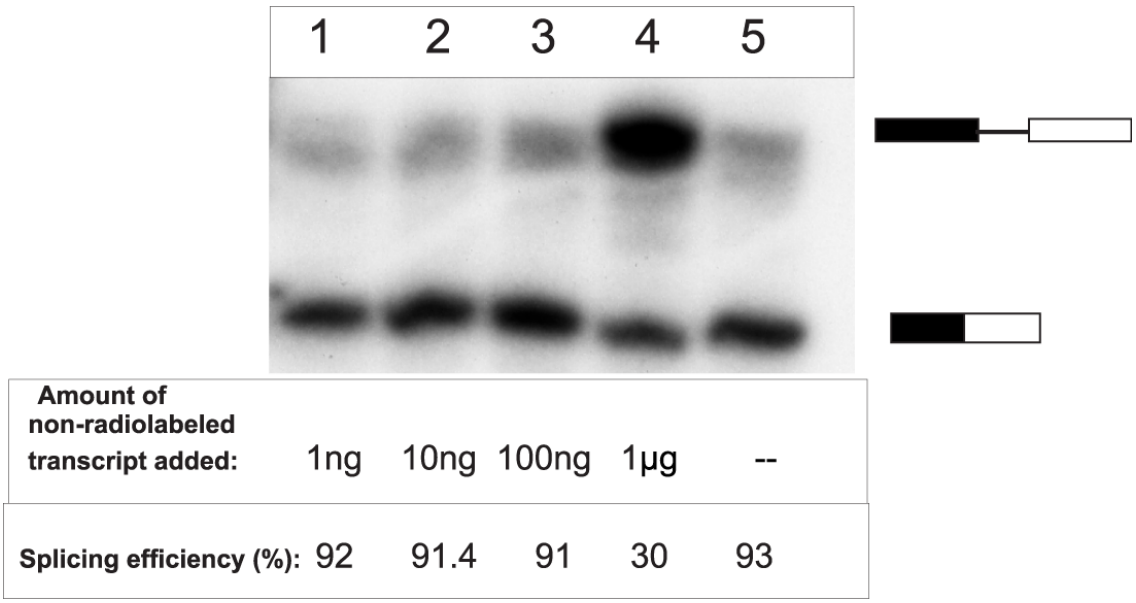


Figure 16. Addition of non-radiolabeled transcript (up to 100 ng) in a splicing reaction does not decrease the splicing efficiency. Increasing amounts (ranging from 1 ng to 1µg) of non-radiolabeled transcript were added to splicing reactions. The total reaction volume was 10 µl.

An important parameter of this purification system is the splicing efficiency of the substrate used. Both the transcript and the spliced product carry the RNA affinity tag, which leads to the affinity selection of a mixture of those two moieties. In order to specifically enrich for the spliced mRNPs, the splicing efficiency must be high. Both MINX and tagged MINX transcripts were efficiently spliced *in vitro*, as shown in figure 17 and the other figures of this section.

Figure 17 shows an experiment validating the affinity selection of mRNPs carrying the boxB tag using GRNA chromatography. The specificity of the system is verified by the results shown in lanes 3 and 6. The spliced tagged MINX product was

specifically precipitated (Figure 17, lane 6), while nothing was precipitated in the case of the untagged transcript (Figure 17, lane 3). Although multiple bands were generated during the splicing process of tagged MINX (Figure 17, lane 5), these bands were not affinity purified (Figure 17, lane 6). GRNA chromatography specifically and reproducibly precipitated the tagged MINX transcript.

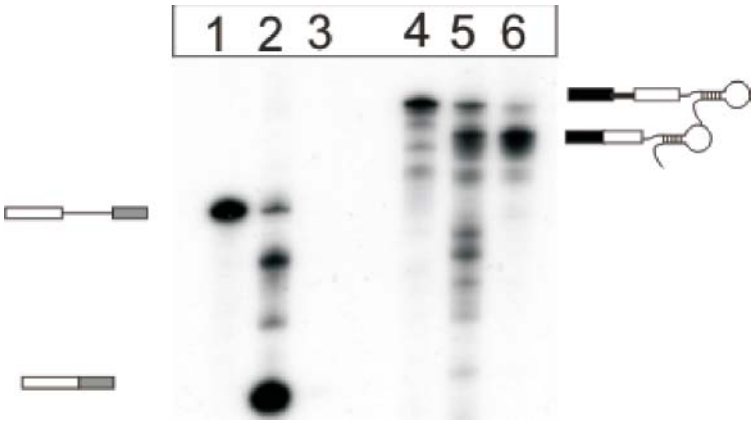


Figure 17. GRNA chromatography specifically precipitates tagged MINX spliced products. Transcript input (lane 1), splicing input (lane 2) and pull-down (lane 3) of the untagged MINX transcript are depicted. Transcript input (lane 4), splicing input (lane 5) and pull-down (lane 6) of the tagged MINX transcript are also shown.

After establishing the affinity purification of mRNAs using GRNA chromatography (Figure 17), the next goal was to investigate the proteins bound onto these mRNAs (mRNPs). Two approaches were employed. The first approach employed, mass spectrometry, can be used for the identification of new proteins. The second method, immunoblot analysis, enables the detection of previously identified proteins, for which specific antibodies are available.

The initial goal was to identify RNA binding proteins by mass spectrometry. Silver staining was performed to visualize the protein content of the affinity purified

mRNPs. The pull down of the spliced substrates is shown in figure 18 A. The pull-down is specific for the mRNAs carrying the affinity tag and efficient (figure 18 A, lanes 3 versus 6 and 8). The protein compositions of untagged and tagged MINX transcripts were compared (Figure 18 B, lanes 2 and 3). MINX transcript without affinity tag served as the negative control for pull-down, which was performed with untagged MINX pre-mRNA in the presence of λN-GST. The silver staining revealed that the protein content of the spliced tagged MINX mRNP and the control pull-down had some similar protein components. The protein bands appeared more intense at the tagged mRNP. This result suggests that the stained proteins are not specific for the tagged transcript, but rather proteins that bind unspecifically to the affinity resin. Several approaches were employed to eliminate this unspecific protein binding. These approaches included increasing the salt concentration and using competitors, such as heparin and tRNA. In addition, different elution methods were employed, such as RNase A for the release of the mRNPs by cutting the unprotected RNA, TCA (trichloroacetic acid) protein precipitation and elution with glutathione. The elution with glutathione was chosen as it was shown to be a very specific and gentle elution method.

In addition to the comparison of tagged and untagged spliced MINX mRNPs, the protein compositions of spliced and unspliced tagged MINX were compared (Figure 18 B, lanes 3 and 4). The protein pattern of the spliced and unspliced affinity purified mRNPs appeared quite similar. Similarities in the protein patterns of the unspliced and the spliced transcript revealed that the binding of the proteins was not splicing specific.

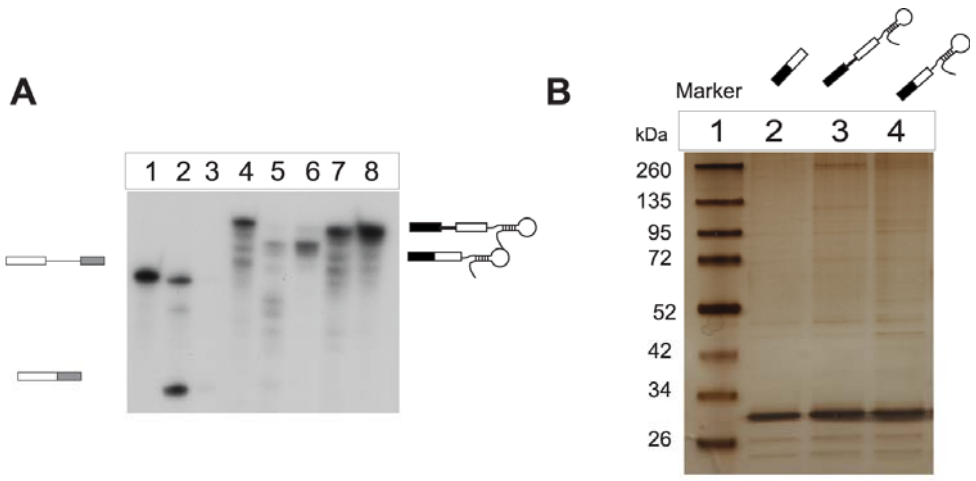


Figure 18. Similar patterns of protein composition for the spliced and unspliced mRNPs. (A) GRNA affinity purification of untagged MINX-control (lanes 1-3) and tagged MINX (lanes 4-8). Transcript inputs of MINX-control (lane 1) and MINX-boxBs (lane 4) are depicted. Splicing input (lane 2) and pull-down (lane 3) of control performed with untagged MINX pre-mRNA in the presence of λN-GST are shown. Splicing input (lane 5) and pull-down (lane 6) of tagged MINX are shown. Splicing input (lane 7) and the pull-down (lane 8) of unspliced tagged MINX. This sample was incubated in mock splicing conditions (namely ATP and phosphocreatine ingredients were eliminated from the splicing mix, leading to the abrogation of splicing). **(B)** The sizes of the molecular weight marker are stated on the left side of the figure (lane 1). Protein composition of control pull-down (lane 2), affinity purified unspliced mRNP (lane 3) and affinity purified spliced mRNP (lane 4).

The mass spectrometry analysis was performed by the EMBL proteomic core facility. The facility required a coomassie stained gel, like the one depicted in figure 19 A. The pull-downs of the spliced substrates is shown in figure 19 C, the efficient and specific affinity selection of spliced mRNAs is the prerequisite for further protein analysis. The goal was to compare the protein composition of a spliced mRNP without any affinity tag, which served as a control pull-down (performed with untagged MINX pre-mRNA in the presence of λN-GST) and a tagged spliced mRNP. The staining revealed a similar protein pattern for both samples. The next step was to identify

proteins bound onto the spliced tagged mRNP, nine protein bands were chosen. These protein bands were very prominent at the affinity purified sample. The bands were selected according to their molecular weight, choosing bands close to the expected size of proteins of interest (known components of the EJC). These bands included: two bands around 34 kDa, three bands around 50 kDa and three bands around 100 kDa. In the group of the smaller proteins, the λN peptide was expected to be identified. Its identification could serve as a positive control for the mass spectrometry analysis. Protein eIF4A3 (47 kDa) was expected to be found in the 50 kDa group of proteins and among the larger proteins, UPF3b was expected to be identified. The selected bands are indicated by arrows in figure 19 B.

An in-gel digestion of proteins with trypsin was used to generate peptide fragments. These peptide fragments were analyzed using matrix-assisted laser desorption/ionization mass spectrometry (MALDI/MS) to determine their exact mass. Resulting peptide masses were then subjected to various database searches to identify the protein through comparison of their masses.

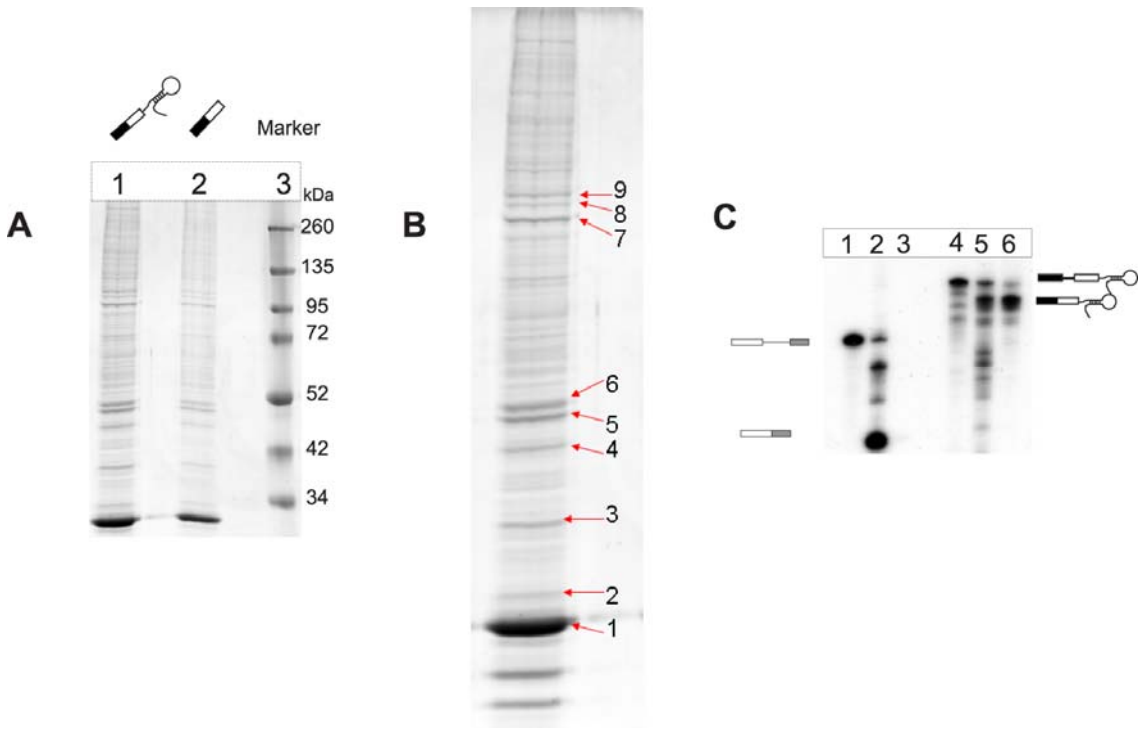


Figure 19. The affinity purified mRNP and the control pull-down show similar protein patterns when visualized with coomassie staining. This figure shows the coomassie stained protein gel given for MS analysis. **(A)** Affinity purified mRNP (lane 1) and the control pull-down performed with untagged MINX pre-mRNA in the presence of λN-GST (lane 2) are depicted. On the right, the molecular weight marker and the respective sizes are shown. **(B)** Magnification of lane 1 from panel (A) is shown. The bands indicated with red arrows were subsequently analyzed by mass spectrometry. The results are summarized in table 4. **(C)** Transcript input (lane 1), splicing input (lane 2) and pull-down (lane 3) of the control are shown. Transcript input (lane 4), splicing input (lane 5) and pull-down (lane 6) of affinity-selected spliced MINX mRNPs are depicted.

The Proteomic facility assigned the bands to specific proteins. The identified proteins are outlined in table 4. The MS did not identify any of the excised gel bands as EJC components. Some RNA binding proteins were identified such as DHX15 and some abundant cellular proteins such as elongation factors, heat shock proteins 90 kDa (Hsp90) and actin. Hsp90 is one of the most abundant proteins in eukaryotic cells, comprising 1–2% of cellular proteins under non-stress conditions (Sreedhar et al., 2004).

As expected, two bands of low molecular weight were identified as the λN peptide. Summarizing these results, the specificity of this purification strategy is not high enough to lead to the identification of EJC components. This may be explained by (1) the low abundance of EJC proteins relative to other RNA binding proteins, (2) sample destruction by the technical procedure.

Gel band excised	Protein name	Protein Description	Sequence Coverage	MW	Peptides matched
1	GST26_SCHJA	Glutathione S-transferase class-mu 26 kDa	50	25482	14
2	GST26_SCHJA	Glutathione S-transferase class-mu 26 kDa isozyme	39	25482	11
3	G3P_HUMAN	Glyceraldehyde-3-phosphate dehydrogenase	41	36030	11
4	ACTB_HUMAN	Actin, cytoplasmic 1	52	41710	17
5	EF1G_HUMAN	Elongation factor 1-gamma	32	50087	17
6	ENOA_HUMAN	Alpha-enolase - Homo sapiens	46	47139	23
7	HS90A_HUMAN	Heat shock protein HSP 90-alpha	49	84607	34
	HS90B_HUMAN	Heat shock protein HSP 90-beta	41	83212	31
	DHX15_HUMAN	Putative pre mRNA-splicing factor ATP-dependent RNA helicase DHX15 - Homo sapiens	35	90875	26
8	EF2_HUMAN	Elongation factor 2	27	95277	24
9	not identified	--	--	--	

Table 4. Mass spectrometry identifies abundant proteins, but not EJC components. The table summarizes the proteins identified from the excised gel bands through mass spectrometry.

These results demonstrated that mass spectrometry could not be used to identify EJC components associated with the spliced RNAs under these experimental conditions. The low abundance of EJC components relative to other RNA binding proteins is a possible explanation for this experimental difficulty. To identify the presence of EJC components, which are in low abundance in post-splicing mRNPs, immunoblot analysis was employed. This approach requires the availability of specific antibodies. Some antibodies (UAP56) were commercially available, while others (RNPS1, Y14, UPF3b) were generously donated by colleagues in the field. In order to obtain a complete set of antibodies, antibodies against EJC proteins (eIF4A3, Barentsz, MAGOH, Y14) were produced. The proteins of interest were expressed in *Escherichia coli* and subsequently purified. These proteins were used as antigens to immunize the rabbits used for the production of the antibodies. The antibodies were produced by the company Eurogentec S.A. (Liege). Due to the fact that these antibodies were not specific, I had to affinity purify eIF4A3 and Barentsz antibodies. A commercially available kit (AminoLink Plus coupling resin) was employed for the affinity purification. This kit is used to covalently attach proteins and other ligands to a beaded agarose support, which results in a stable and reusable column for affinity purification of antibodies, antigens and other bio-molecules. The protein used for the antibody production was attached to the column and then this column was used to affinity purify the antibody produced. In the case of eIF4A3, the affinity purification led to a highly-specific antibody (Figure 20). In figure 20, the antibody (after the affinity purification) specifically immunodetected eIF4A3 protein in HeLa nuclear extract and in HEK 293 whole cell extracts (WCEs). In addition, the antibody detected the Flag- eIF4A3 in HEK 293 WCEs that have been previously transfected with Flag-eIF4A3. In addition to these purified antibodies, commercially available antibodies (UAP56) and antibodies from other laboratories (UPF3b, Y14, RNPS1 and MAGOH) were used to create a panel of antibodies against EJC components.

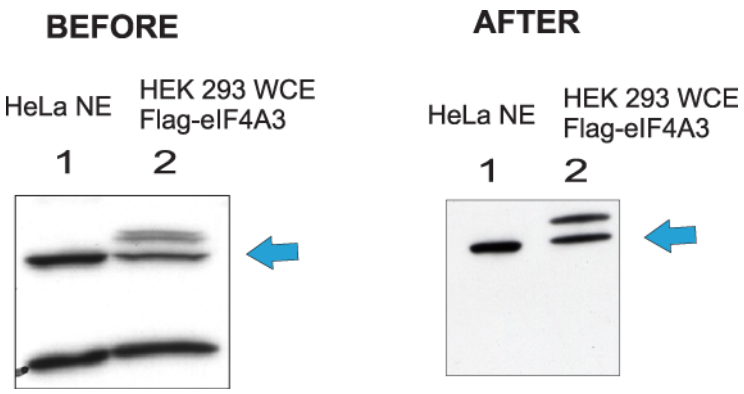


Figure 20. Affinity selection improves the specificity of the produced α-eIF4A3. The figure depicts immunoblots performed using the produced eIF4A3 antibody before and after affinity selection using the AminoLink® Plus Immobilization Kit from Thermo Scientific/Pierce.

Using this highly specific, affinity purified antibody, I subsequently tested whether the mRNPs isolated using GRNA chromatography contained eIF4A3. eIF4A3 was the first protein investigated using immunoblot analysis, since it had been shown to be the platform onto which the rest of the EJC assembles (Shibuya et al., 2004). Therefore finding eIF4A3 co-purified with the affinity purified mRNPs gives an initial hint for the presence of the EJC.

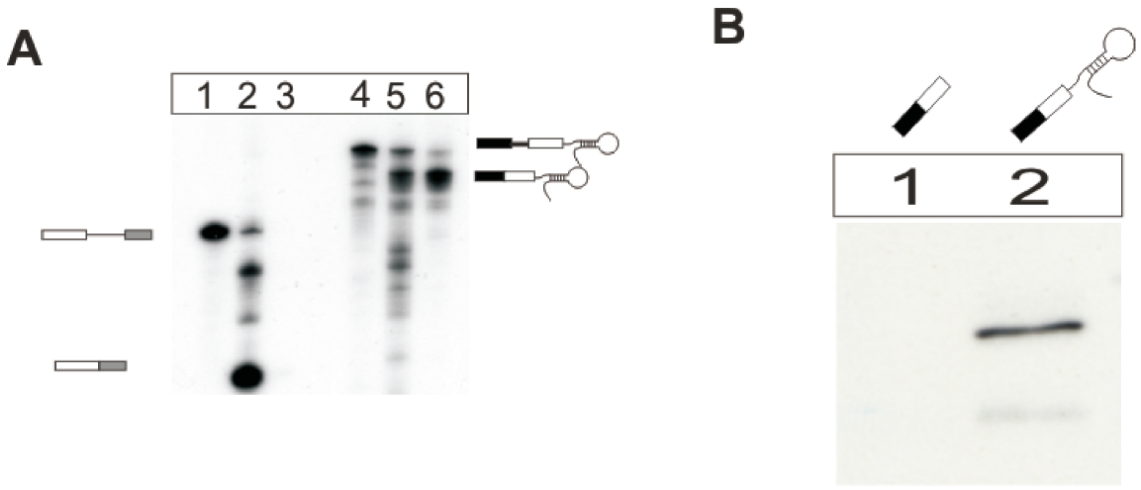


Figure 21. GRNA Chromatography affinity purifies eIF4A3-containing mRNPs spliced *in vitro*. (A) Affinity selection of RNA products. Lanes 1-3 correspond to control pull-down performed with untagged MINX pre-mRNA in the presence of λN-GST. Lanes 4-6 correspond to the tagged MINX. Lanes 1 and 4 show transcript inputs, lanes 2 and 5 splicing inputs (of the sample that is consecutively loaded onto the affinity purification resin) and lanes 3 and 6 show each an aliquot of the affinity purified sample that is subsequently used for the immunoblot shown in the (B) section of this figure. **(B)** Immunoblot analysis of affinity purified mRNPs using the eIF4A3 antibody (figure 20). The control pull-down (lane 1) and the tagged MINX pull-down (lane 2) are shown. Protein eIF4A3 was specifically immunodetected within affinity purified mRNPs from the tagged MINX (lane 2).

Protein eIF4A3 was immunodetected within affinity purified mRNPs from *in vitro* splicing reactions (Figure 21 B). The result is specific, since protein eIF4A3 was only detected at the affinity purified sample from the transcript with the affinity tag (Figure 21 B, lane 2) and not the control (Figure 21, lane 1). This outcome showed that the core EJC component eIF4A3 is present in the affinity purified mRNPs.

In addition to eIF4A3, the presence of UAP56 was investigated. UAP56 was identified within the affinity-selected tagged MINX sample (Figure 22, lane 2). UAP56 is an essential splicing factor needed for the U2 snRNP-branchpoint interaction (Fleckner et al., 1997). The stable binding of U2 snRNP at the pre-mRNA branchpoint is the first ATP-dependent pre-mRNA splicing process, which requires the hydrolysis of ATP from UAP56 (Fleckner et al., 1997).

To sum up the findings, eIF4A3 and UAP56 were detected within the affinity purified samples using immunoblot analysis. On the other hand, the proteins RNPS1, MAGOH, Y14 and UPF3b could not be detected.

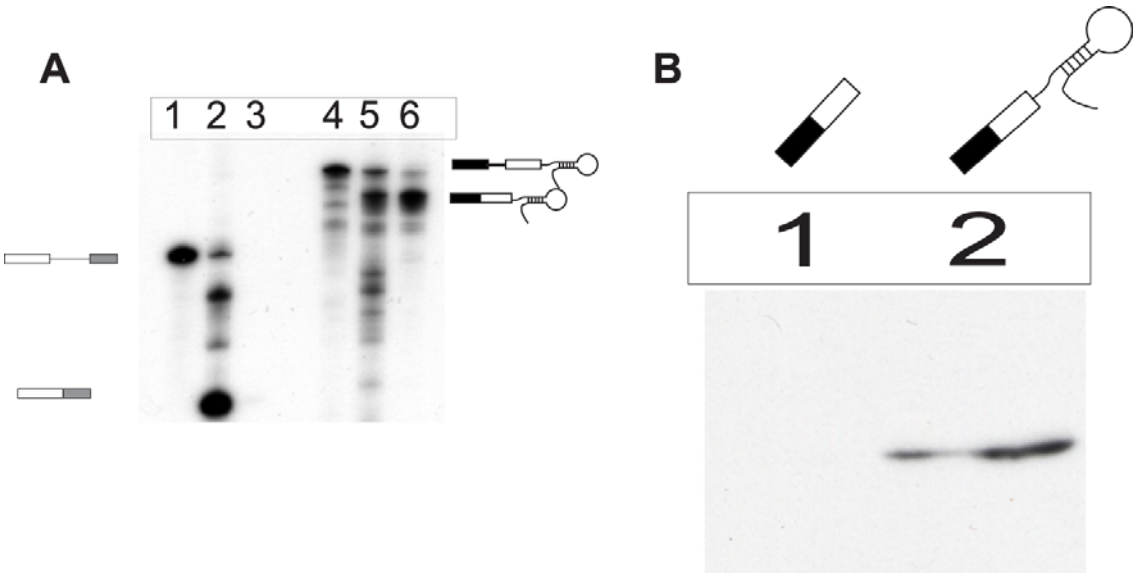


Figure 22. UAP56 is present within the affinity purified mRNPs. **(A)** Affinity selection of RNA products. Lanes 1-3 correspond to control pull-down and lanes 4-6 correspond to the tagged MINX transcript. Transcript inputs (lane 1-control and lane 4-tagged MINX) and splicing inputs (lane 2-control and lane 5-tagged MINX) are depicted. Pull-down inputs (lane 3-control and lane 6-tagged MINX), namely aliquots of the affinity purified sample that was subsequently used for the immunoblot shown in the (B) section of this figure are depicted. **(B)** Immunoblot analysis of affinity purified mRNPs with a commercially available UAP56 antibody. Control pull-down (lane 1) and tagged MINX pull-down (lane 2) are depicted.

Despite optimized splicing and GRNA chromatography conditions, no EJC proteins were identified by mass spectrometry and the immunoblot analysis could detect only one EJC component, eIF4A3, and a TREX component, UAP56, but not other critical components of the EJC. Therefore, the experimental system was not a suitable method to address the question whether substrate specific differences in the composition of EJC exist. At this point, this approach was stopped and the attention was directed towards a new set of biological questions, starting with the hierarchy of the EJC assembly.

3.2 The *in vitro* splicing system faithfully recapitulates the splicing-dependent deposition of EJC proteins.

EJC has been assembled *in vitro* from recombinant proteins on a RNA substrate in the presence of ATP (Ballut et al., 2005). Therefore, its assembly can occur in the absence of the physiological deposition machinery, the spliceosome. However, under these experimental conditions, EJC's physiological assembly hierarchy was not determined. One of the goals of this thesis was to study EJC deposition during splicing. The stepwise assembly of EJC during splicing, in parallel to the spliceosome assembly was investigated. More specifically, we tried to investigate which EJC components bind before exon ligation and if splicing is essential for the deposition of all the EJC core components (eIF4A3, MAGOH-Y14 and Barentsz) and an EJC core interacting factor, Upf3b. The EJC is assembled in the nucleus and it is removed from the spliced mRNA in the cytoplasm. A very interesting question is what mediates the EJC disassembly. Apart from the EJC, a number of protein complexes are assembled on the mRNA in the nucleus. Another protein complex that is assembled onto the mRNA in the nucleus is the TREX complex. Some of the TREX complex components have been suggested to be EJC components (Gatfield et al., 2001; Le Hir et al., 2001b; Le Hir et al., 2000). Therefore, I was interested to investigate the interconnection between EJC and TREX. One of the goals was to test whether TREX deposition is splicing and/ or EJC dependent. Although the splicing dependency has been addressed before (Masuda et al., 2005), I first wanted to recapitulate the existing published results using our *in vitro* system.

A robust experimental system was required to investigate the splicing-dependent deposition of exon junction complex (EJC) components onto an mRNA. The whole cell lysate system for *in vitro* splicing developed in the Dreyfuss' lab (Kataoka, N. and Dreyfuss, G., 2004) was employed and modified to address the questions described in sections 3.2-3.5. *In vitro* splicing reactions with the MINX pre-mRNA were performed in HeLa nuclear extract supplemented with splicing-competent whole cell extracts from HEK 293 cells expressing different FLAG-tagged EJC proteins.

During the splicing reaction, these tagged proteins are incorporated into the EJC and allow the immunoprecipitation of splicing-specific mRNPs. A schematic representation of the system is depicted in figure 23.

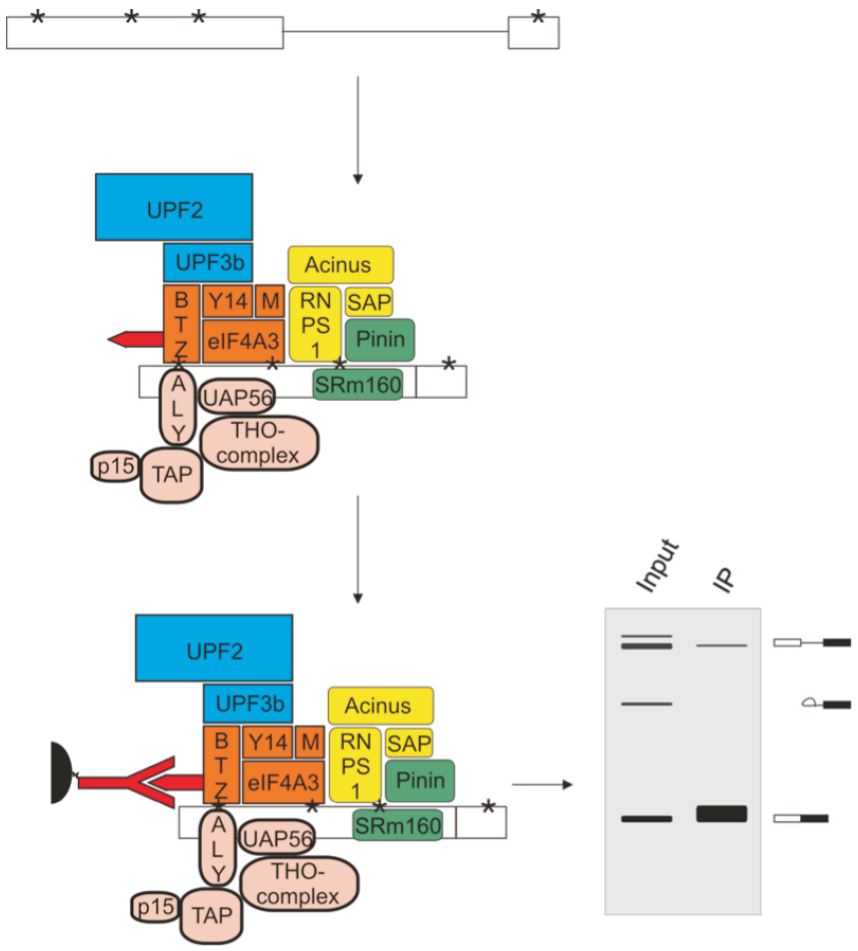


Figure 23. Schematic representation of the *in vitro* splicing system. The experimental strategy involves the splicing of a radiolabeled (α-^{32}P labeling is signified by the stars) transcript in a combination of HeLa nuclear extract and whole cell extracts prepared from HEK 293 cells previously transfected with FLAG-tagged proteins of interest. During splicing, the tagged EJC protein is assembled onto the RNA. Next using the FLAG tag (red arrow), the RNA products are immunoprecipitated. The precipitated products can be resolved on a denaturing urea gel, as shown schematically.

From these splicing reactions, the FLAG-tagged protein is immunoprecipitated in association with the spliced mRNAs and other RNA binding proteins. As expected for EJC proteins, the spliced mRNAs specifically co-immunoprecipitated with FLAG-eIF4A3, -Barentsz, -MAGOH, -Y14 and -Upf3b **(**Figure 24 A**)**. In order to test whether the immunoprecipitations are specific for spliced RNAs, a capped intronless MINX transcript (MINX Δi) was used. As expected, only traces of the intronless MINX transcript are pulled down under the same experimental conditions (Figure 24 C). As a positive control for the protein binding ability of the MINX Δi RNA, we have used the FLAG-CBP80. This protein associates with mRNAs in a splicing-independent but cap-dependent manner. The capped intronless transcript co-immunoprecipitated with the FLAG-CBP80 (Figure 24 D), as expected for a capped transcript. The expression of the FLAG-proteins used for the immunoprecipitations is compared in figure 24 B.

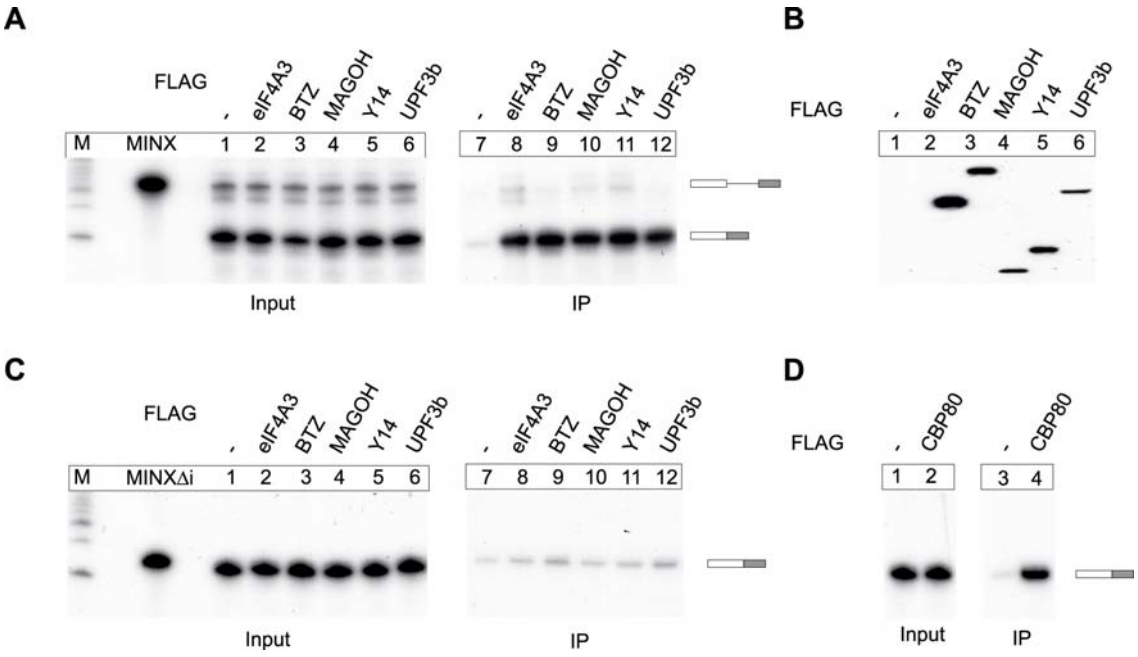


Figure 24. The *in vitro* splicing system recapitulates faithfully the splicing-dependent deposition of EJC proteins. (A) Splicing reactions using capped MINX substrate were supplemented with extracts expressing the indicated FLAG-tagged EJC proteins or unfused FLAG-tag as a negative control. The RNA products were precipitated through the FLAG tag using FLAG affinity agarose. The positions of the full length transcript and the spliced product are depicted schematically on the right side. **(B)** The expression of the FLAG-proteins used in

(A) is confirmed by immunoblot analysis with a FLAG antibody. **(C)** Splicing reactions and immunoprecipitations were performed as in (A) with a capped intronless MINX transcript (MINX Δi). **(D)** Splicing reactions and immunoprepcipitations with FLAG-CBP80 or unfused FLAG employing the intronless MINX transcript (MINX Δi) used in (C).

Concluding, this *in vitro* splicing system is a valuable tool that recapitulates faithfully the splicing-dependent deposition of EJC proteins. This *in vitro* splicing system has been used to address questions concerning the assembly and the disassembly of the EJC, as described in detail in sections 3.3, 3.4, and the deposition of TREX complex, as described in section 3.5. Initially the hypothesis that the assembly of EJC is an ordered procedure with intermediate steps was tested.

3.3 Ordered assembly of the Exon Junction Complex by the spliceosome

The EJC-proteins eIF4A3, Y14 and MAGOH were found in purified spliceosomes, purified spliceosomal B and C-complexes, whereas Barentsz has not been identified among the purified proteins (Bessonov et al., 2008; Makarov et al., 2002; Makarova et al., 2004). These findings lead to the questions of when and in which order individual EJC proteins are recruited to the RNA during splicing. Our initial hypothesis was that the EJC components assemble on the mRNA during the formation of distinct spliceosomal complexes. In order to test whether the EJC assembly is stepwise, a comprehensive set of eIF4A3, MAGOH, and Barentsz mutants in complete or C-complex-arrested splicing reactions was analyzed. The structure of mRNA was manipulated, so that the splicing process is arrested after the first step, at the stage of the C-complex. Arresting the spliceosome at its C form, allows the investigation of proteins associated before and after this step. The associated proteins can be compared with the proteins associated with substrates that have completed splicing.

3.3.1 eIF4A3 and MAGOH-Y14 form a pre-EJC before exon ligation

To address the question in which order during splicing are individual EJC proteins recruited to the RNA, FLAG-immunoprecipitations were performed from splicing reactions using a MINX pre-mRNA that contained an AG→GG mutation at the physiological 3' splice site. This alteration results in the blocking of the splicing reaction after 5' splice site cleavage and lariat formation and before exon ligation (Reed et al., 1985). The products at this stage include an intron lariat-3'exon and 5'exon, plus the pre-mRNA that did not initiate splicing. eIF4A3, MAGOH and Y14 specifically precipitated the splicing intermediates (intron lariat-3'exon, 5' exon) and the pre-mRNA (Figure 25, lanes 8, 10 and 11). While UPF3b precipitated the pre-mRNA only marginally above background and did not precipitate splicing intermediates (Figure 25, lane 12), BTZ precipitated only some pre-mRNA (Figure 25, lane 9). This finding demonstrates that both, BTZ and UPF3b do not associate with the RNA at an early step of splicing and stable binding thus appears to occur at later stages or after splicing of the mRNA has been completed. This result agrees with the current view that UPF3b binds only to the fully assembled EJC after release of the spliceosome (Kim et al., 2001a). The surprising finding is that BTZ does not associate with eIF4A3 during splicing and may even be actively excluded from the spliceosomal C-complex. The above finding is not expected if one takes into consideration the previously reported observation that BTZ can interact *in vitro* with eIF4A3 even in the absence of Y14 and MAGOH (Bono et al., 2006; Shibuya et al., 2006).

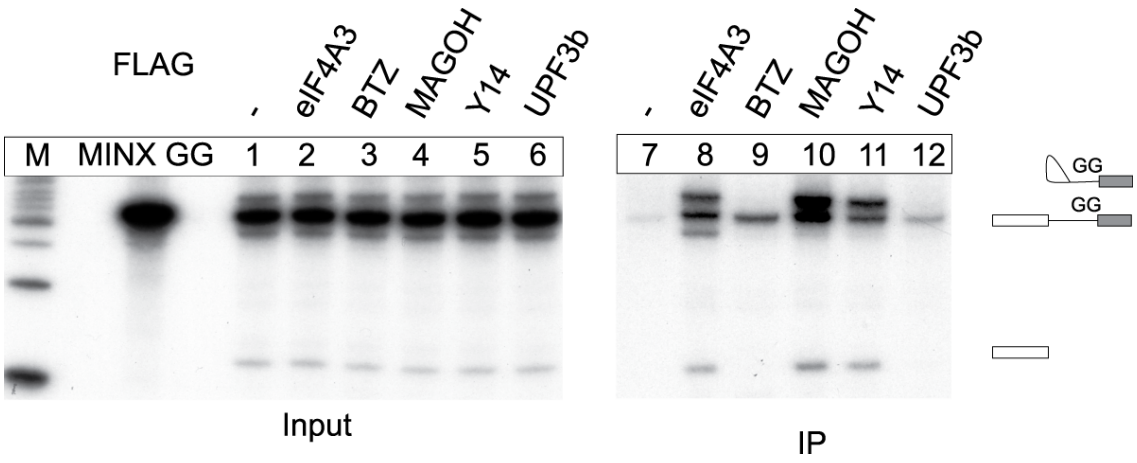


Figure 25. eIF4A3 and MAGOH-Y14 form a pre-EJC before exon ligation. Splicing reactions and immunoprecipitations were performed with a mutated MINX transcript (MINX GG) that does not undergo exon ligation. The unspliced transcript and the splicing intermediates (exon 1; lariat and exon 2) are depicted schematically. Proteins eIF4A3, MAGOH and Y14 precipitated both the unspliced transcript and the splicing intermediates. Proteins Barentsz and UPF3b precipitated only the unspliced transcript.

These results reveal a stepwise assembly of the biochemically defined EJC core. The assembly reaction entails three early (eIF4A3, MAGOH, Y14) and two late components (BTZ and UPF3b). Interestingly, BTZ behaves in this respect like the EJC-binding protein UPF3b. It seems that BTZ binds to eIF4A3 after the trimeric EJC core is fully assembled, or that binding occurs when the spliceosome has dissociated from the mRNA. Trying to distinguish between these two alternatives, splicing reactions were performed in extracts with or without the initial addition of 293 extracts expressing FLAG-tagged EJC proteins (Figure 26). The reactions that were performed without FLAG-tagged proteins were supplemented with the same amount of FLAG-extracts after splicing was completed and further incubated on ice to allow for the exchange of EJC-bound factors with supplemented proteins. At the end, recombinant proteins and associated mRNAs were affinity purified by FLAG-immunoprecipitation. Specifically, eIF4A3, MAGOH and Y14 co-purified spliced mRNA when they were present before splicing was initiated (Figure 26, lanes 13 & 14, 17-20). This is the predicted result for factors that bind to the mRNA in a splicing dependent manner

without any exchange of components after the EJC has been assembled and splicing is finished. Interestingly, BTZ and UPF3b co-purified the spliced mRNA when they were provided before or after splicing (Figure 26, lanes 15-16 and 21-22). Thus, BTZ and UPF3b interact with the trimeric EJC core that is loaded on the RNA in a splicing-dependent manner, but do not require splicing for the interaction. As shown in figure 26, the addition of BTZ and UPF3b protein before and after splicing led in both cases to the incorporation of these two proteins into the EJC. This observation suggests that splicing itself is not required for the recruitment of neither BTZ nor UPF3b. On the other hand, their deposition requires the presence of other proteins such as eIF4A3, MAGOH and Y14. The recruitment of the latter proteins (eIF4A3, MAGOH and Y14) is splicing-dependent, since the addition of these proteins after splicing does not lead to the precipitation of spliced product.

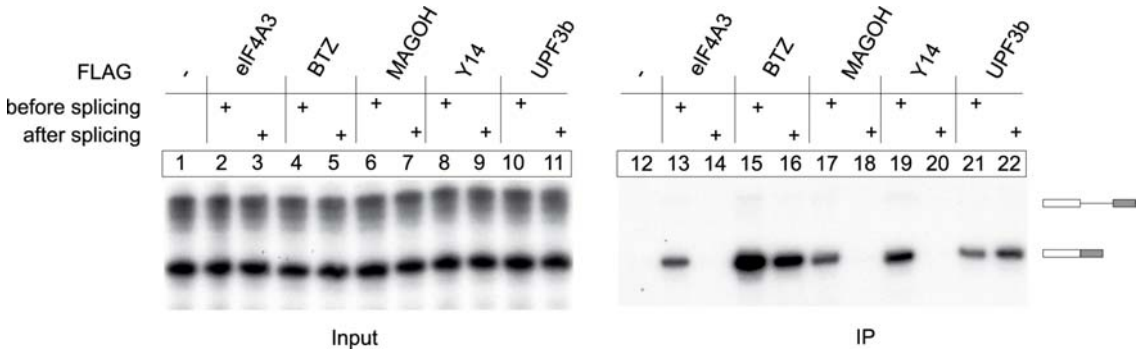


Figure 26. Deposition of eIF4A3, MAGOH and Y14 requires splicing, while Barentsz and UPF3b do not require splicing to bind at the EJC. In order to study the composition of EJC before and after the completion of splicing, splicing reactions were performed in the presence (lanes 2, 4, 6, 8, 10, 13, 15, 17, 19, 21) or absence (lanes 3, 5, 7, 9, 11, 14, 16, 18, 20, 22) of FLAG-protein extracts or with empty FLAG as a negative control (lanes 1 & 12). Reactions without added FLAG-proteins were supplemented after the completion of splicing with FLAG-protein extracts and incubated on ice (30 minutes).

To sum up the presented results, eIF4A3, MAGOH and Y14 assemble on the mRNA before exon ligation (as shown in figure 25) and they remain bound until after

the end of splicing (Figure 26). On the other hand, proteins BTZ and UPF3b bind only after exon ligation (Figure 25) and their binding is not splicing-dependent (Figure 26).

During the course of our analyses of the splicing-mediated deposition of EJCs, we noticed that mutants of MAGOH designed to lack the interaction with PYM (for more information about PYM, see beginning of next section, section 3.4) subtly but reproducibly co-immunoprecipitated more spliced MINX mRNA than wild type MAGOH. This observation suggests that these MAGOH mutants are either more efficiently incorporated into EJCs, or that EJCs containing such mutants are more stable than EJCs with wild type MAGOH. Both effects could be caused by the lack of PYM interaction. The results of the analysis of these alternatives are presented in the next section.

3.4 Disassembly of Exon Junction Complexes by PYM

PYM is a 29-kDa protein that binds the Y14-MAGOH heterodimer in the cytoplasm (Forler et al., 2003), but does not interact with other EJC components. It has been proposed that PYM functions to enhance the translation of spliced mRNAs by directly linking the EJC to the ribosomal 48S preinitiation complex in the cytoplasm. The N-terminus of PYM is responsible for the interaction between Y14-MAGOH and PYM (Bono et al., 2004), while the C-terminus is completely dispensable. The observation that PYM associates with the nuclear cap-binding protein CPB80 but not with the cytoplasmic cap-binding protein eIF4E, proposes that PYM has a role at the initial round of translation, rather than in steady-state translation (Chiu et al., 2004; Diem et al., 2007; Hosoda et al., 2005; Ishigaki et al., 2001; Lejeune et al., 2002). On the basis of its physical interaction with the EJC proteins and the finding that tethering of PYM to the 3'UTR of a reporter mRNA reduces its abundance, PYM was proposed to act as an NMD factor (Bono et al., 2004). Recently, PYM was shown to interact with ribosomes and to co-sediment with 40S ribosomal subunits in sucrose gradient centrifugations (Diem et al., 2007). Based on this association, PYM was proposed to

serve as a bridge between the EJC and the ribosome, and to mediate the stimulation of translation by EJCs bound within the open reading frame (Diem et al., 2007).

3.4.1 PYM reduces the amount of EJCs bound to spliced mRNAs

Based on the observation that MAGOH mutants lacking the interaction with PYM subtly but reproducibly co-immunoprecipitated more spliced MINX mRNA than wild type MAGOH, we hypothesized that EJCs containing such mutants are more stable than EJCs with wild type MAGOH. To directly test if PYM influences the EJC, we purified recombinant PYM (rPYM) from bacterial expression cultures and added increasing amounts of rPYM to splicing reactions containing FLAG-tagged EJC proteins. Splicing reactions containing one of the following FLAG-eIF4A3, -BTZ, -MAGOH, -Y14 or -UPF3b were performed and then increasing amounts of rPYM were added, followed by FLAG-immunoprecipitations. In all cases rPYM strongly reduced the amount of co-precipitated MINX mRNA in a concentration dependent manner (Figure 27). Therefore, the effect of rPYM is specific for EJCs and independent of the proteins used to precipitate the EJC. Importantly, the association of CBP80 with the mRNA was not inhibited by PYM under the same conditions (Figure 28). Thus, PYM specifically targets mRNA-bound EJCs rather than non-specifically disturbing mRNPs. This observation suggests that PYM is one of the factors able to disassemble the EJC.

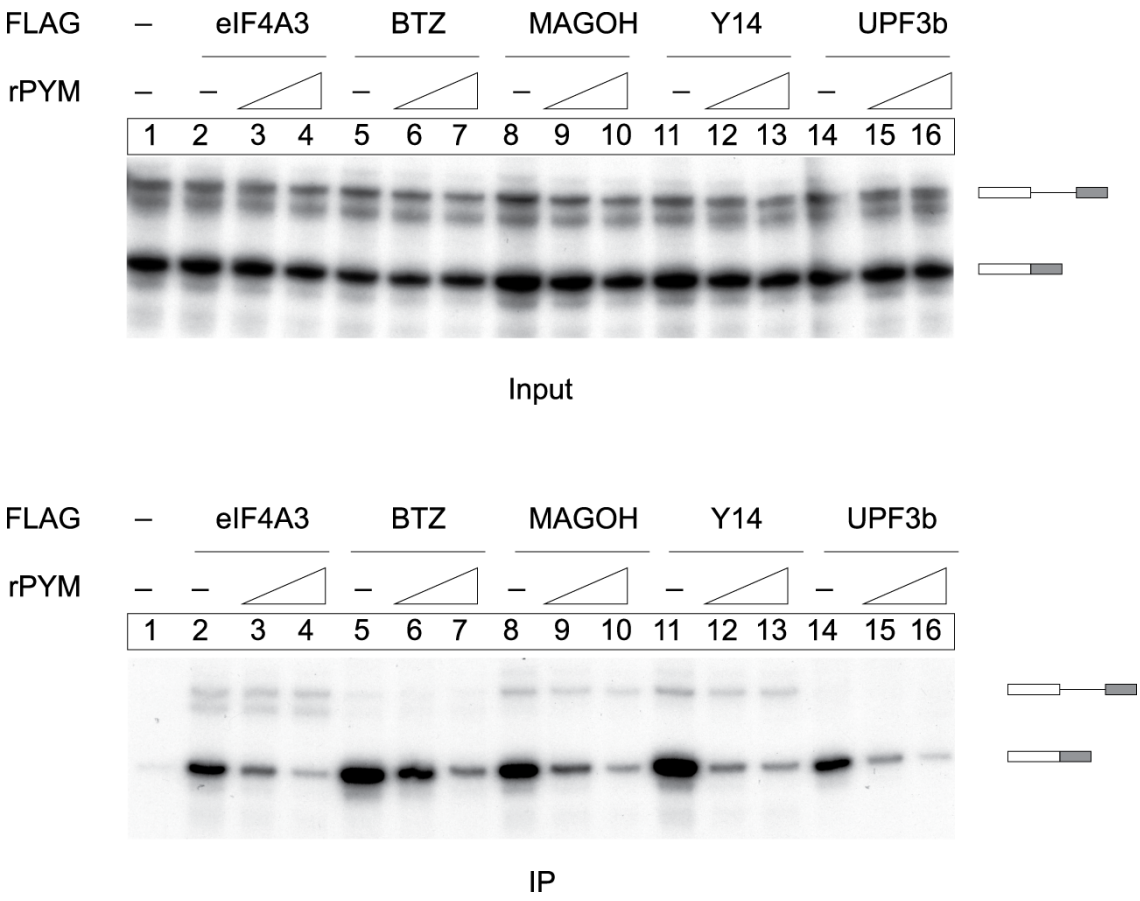


Figure 27. PYM reduces the amount of EJCs bound to spliced mRNAs. Splicing reactions in HeLa cell nuclear extract using MINX as a substrate were supplemented with 293 cell extracts expressing the indicated protein (eIF4A3, BTZ, MAGOH, Y14 and Upf3b) or unfused FLAG-tag as a negative control. Reactions were immunoprecipitated using FLAG affinity agarose. Increasing amounts of recombinant PYM (rPYM) were added to the reactions as shown.

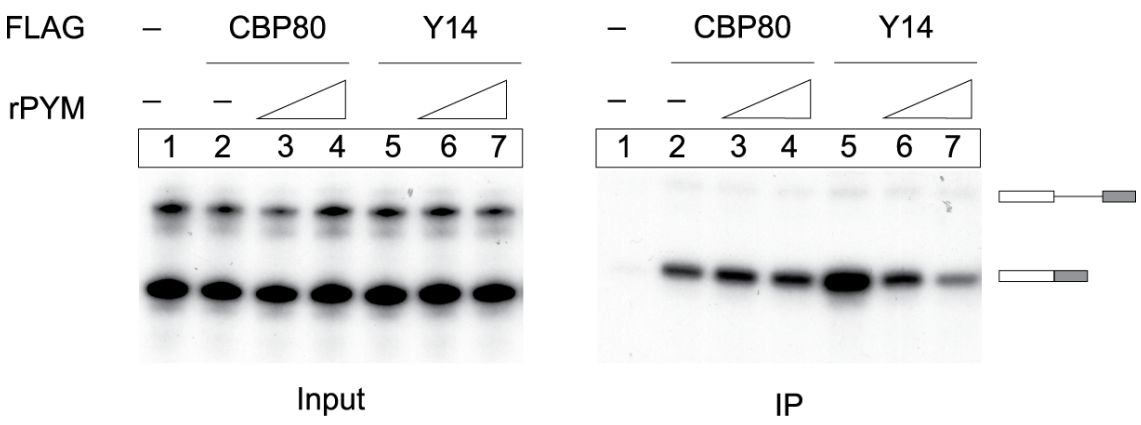


Figure 28. Addition of PYM has no effect on the binding of CBP80. Splicing reactions and immunoprecipitations with CBP80 and Y14 were performed as described in figure 27.

3.4.2 The N-terminus of PYM is required for EJC disassembly

It has been shown that PYM interacts with the MAGOH-Y14 heterodimer via its N-terminus (Bono et al., 2004). To further validate the specificity of the PYM effect on the EJC disassembly, a deletion mutant of PYM was generated. This mutant of PYM lacks the N-terminus and hence the interaction with MAGOH-Y14 should be abolished according to Bono et al. Splicing reactions were performed using either FLAG-eIF4A3 or FLAG-Y14, and increasing amounts of rPYM were added to the reactions. When equal amounts of rPYM and rPYM ΔN33 were tested in splicing reactions, rPYM ΔN33 did not affect the amount of precipitated MINX mRNA (Figure 29, lanes 5-6 and 10-11), while rPYM reproduced the usual interference (Figure 29, lanes 3-4 and 8-9). These results show that PYM requires its MAGOH-Y14 interacting N-terminus to interrupt the association of the EJC with spliced mRNA.

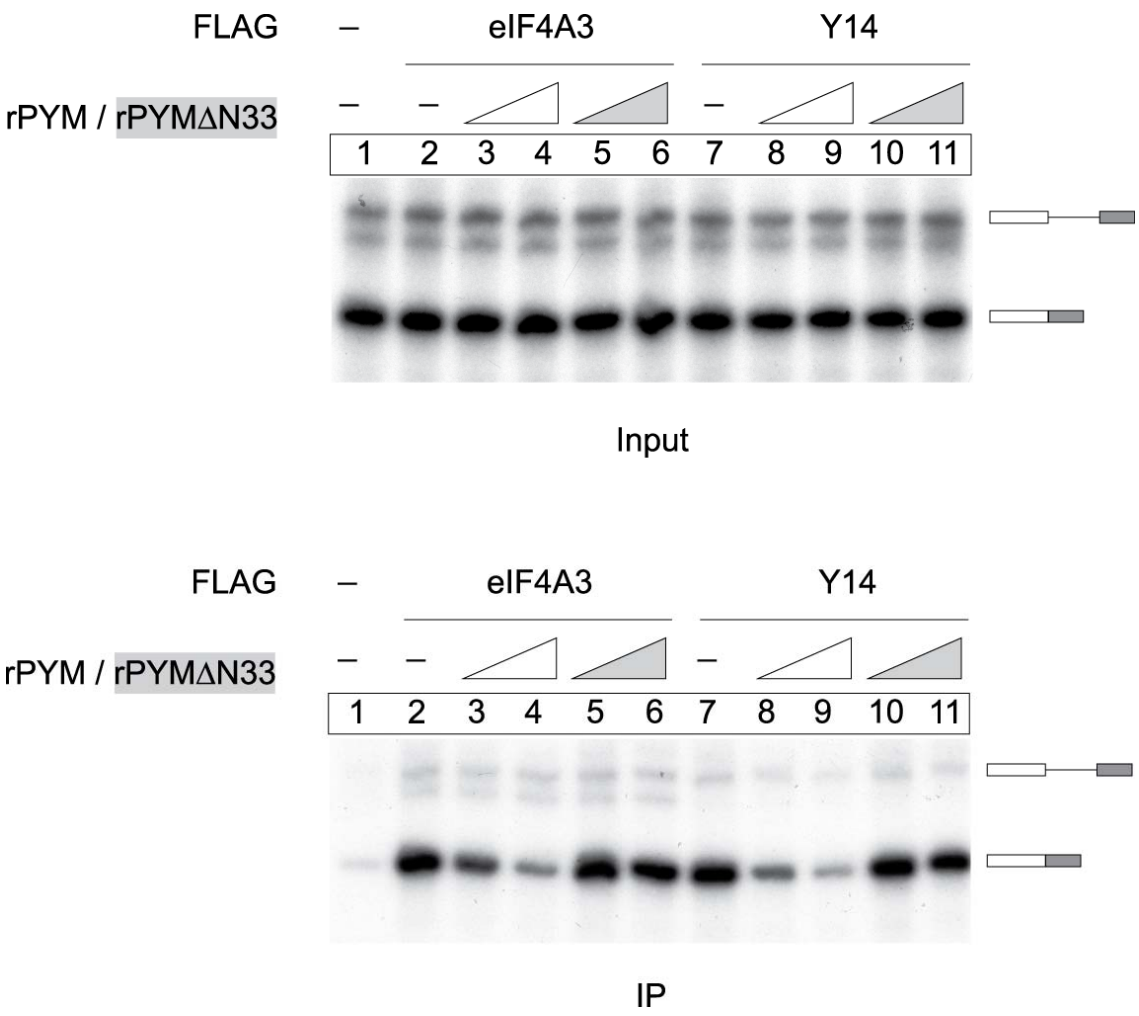


Figure 29. The N-terminus of PYM is required for EJC disassembly. Splicing reactions were supplemented with recombinant PYM (white triangles) or PYM ΔN33 (gray triangles). This PYM mutant did not disassemble the EJC. Immunoprecipitations were performed via eIF4A3 or Y14.

3.4.3 PYM dissociates assembled EJCs but does not inhibit EJC assembly

The question of whether PYM inhibits the assembly of stable EJCs or whether it disrupts assembled EJCs was addressed next. Thus, we used a splicing substrate that completes the first step of splicing but fails to undergo exon ligation (referred to as MINX GG, (Reed et al., 1985) and figure 25). The EJC proteins eIF4A3, MAGOH and Y14 already associate with the splicing intermediate (C complex) that is formed on MINX GG (Figure 25). Splicing reactions were performed using FLAG-eIF4A3, FLAG-MAGOH and FLAG-Y14. rPYM was added to the splicing reactions, after their completion. Addition of rPYM does not affect the association of any of these three proteins with the pre-mRNA or with the C complex (Figure 30 A). Comparable amounts of splicing intermediates were immunoprecipitated both, in the presence and in the absence of rPYM (Figure 30 A). This result indicates that PYM does not interfere with the EJC assembly intermediate that already includes its target, MAGOH-Y14. Rather, EJCs become sensitive to PYM at a later step, possibly when the spliceosome has dissociated from the mRNA.

If PYM is able to disassemble EJCs, it should be possible to dissociate EJCs from mRNAs after splicing is completed. This notion was tested by adding rPYM, or rPYM ΔN33 as a specificity control, to completed splicing reactions (120 minutes of splicing) and continued incubation for another 30 minutes. As shown in figure 30 B, the addition of rPYM to completed splicing reactions decreased the amount of EJCs associated with the spliced MINX mRNAs, whereas rPYM ΔN33 had no effect. These results demonstrate that PYM specifically disrupts mature EJCs, and that PYM is able to dissociate EJCs from spliced mRNAs.

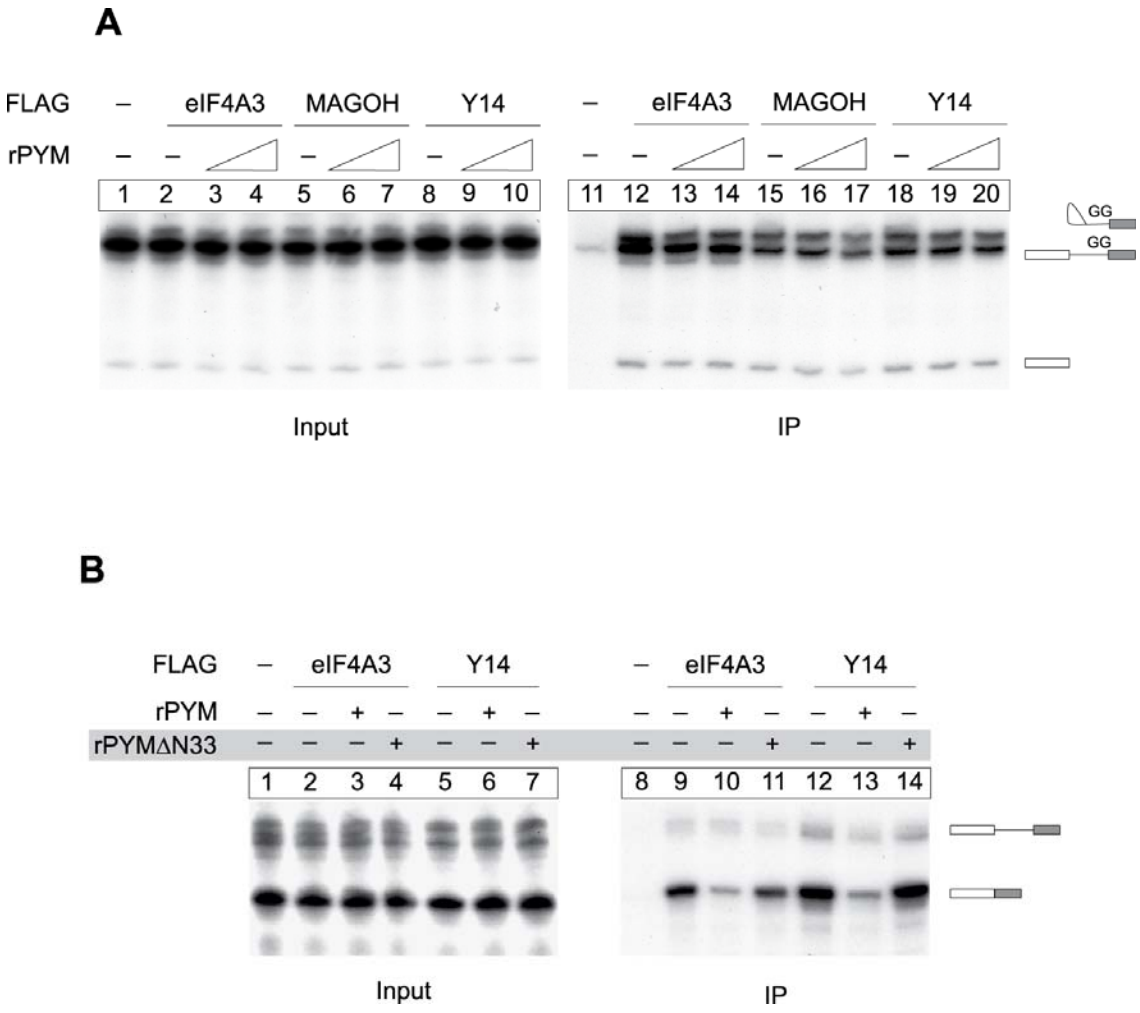


Figure 30. **PYM targets mature EJCs, but not spliceosome-bound EJC precursors**. **(A)** Splicing reactions were performed using MINX GG as a substrate precursor RNA that does not undergo the second step of splicing. Immunoprecipitations were performed via eIF4A3, Y14 and MAGOH. Reactions were supplemented with increasing amounts of recombinant PYM (rPYM) as indicated. **(B)** Splicing reactions were performed using MINX as a splicing substrate for 120 minutes in extracts supplemented with the indicated FLAG-proteins. After splicing was completed, recombinant PYM or PYM ΔN33 (gray shaded) were added and the reactions were incubated for another 30 minutes. Immunoprecipitations were done via eIF4A3 or Y14.

3.5 Deposition of the mammalian TREX complex

The two previous sections (3.3 and 3.4) describe the use of the *in vitro* splicing system to address the assembly and disassembly of EJC. Now we used this system to investigate the deposition of other related protein complexes. The deposition of human TREX complex has been addressed in a number of studies. In previous work, the TREX members UAP56 and ALY/REF were identified as components of the exon junction complex (EJC) (Gatfield et al., 2001; Le Hir et al., 2001b; Le Hir et al., 2000). The EJC was also reported to contain the mRNA export receptor TAP/p15 (Le Hir et al., 2001b; Tange et al., 2004). The next step was to use the *in vitro* splicing system, described in previous sections (section 3.2), and the freshly discovered disassembly effect of PYM to EJCs (section 3.4), to study the deposition of the TREX complex and the EJC contribution to this deposition.

The TREX (TREX TRanscription EXport) protein complex is found in all eukaryotes including yeast, *Drosophila melanogaster* and humans (Gatfield et al., 2001; Jensen et al., 2001; Luo et al., 2001; Strasser et al., 2000, 2001; Stutz et al., 2000; Zhou et al., 2000). This conserved protein complex integrates steps in mRNA biogenesis with nuclear export. Despite its conservation, it is associated with the transcription apparatus in yeast and the splicing machinery in humans. The yeast TREX complex contains the mRNA export factors Sub2, Yra 1 and the THO complex, which functions in transcription elongation (Fischer et al., 2002; Jimeno et al., 2002) and transcription dependent recombination (Huertas et al., 2003). In yeast, this complex plays a role in coupling transcription to mRNA export (Abruzzi et al., 2004; Fischer et al., 2002; Huertas et al., 2003; Jimeno et al., 2002; Rondon et al., 2003). The human TREX complex is comprised of the following proteins: ALY/REF, UAP56, hTex1 and the hTHO complex (hTho2, hHrp1, fSAP79, fSAP35, fSAP24) (Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002). A detailed table of the TREX members in yeast, *Drosophila melanogaster* and human (table 1) is present in the introduction section. In human, UAP56 and ALY/REF colocalize with splicing factors in nuclear speckle domains and confocal microscopy studies in mammalian cells indicate that

UAP56 and ALY/REF are recruited to sites of active transcription in a splicing-dependent manner (Custodio et al., 2004). The recruitment of the human TREX complex to spliced mRNA has reported to occur by a splicing-coupled mechanism rather than by the direct transcription-coupled mechanism that occurs in yeast (Masuda et al., 2005).

In the literature several possible models exist concerning the deposition of the TREX complex. Initial studies suggested a close link between splicing and mRNA export in higher eukaryotes (Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002). It has been proposed that ALY/REF, UAP56 and the mRNA export receptor TAP/p15 are recruited through the EJC (Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002). Recent studies show that hTREX is recruited in a splicing- and cap- dependent manner only to the 5'end of the mRNA (Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002). This recruitment requires the cap-binding subunit CBP80, which interacts directly with the ALY/REF subunit of the hTREX (Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002). The contribution of the EJC to the recruitment of TREX factors awaits an answer.

3.5.1 Protein DDX39 is highly similar to protein UAP56

Looking through the literature for proteins involved in the mRNA export pathway, a protein called DDX39 was revealed. DDX39 or URH49 is a mammalian protein that is highly identical to the DExH/D box protein UAP56 (Pryor et al., 2004). *Saccharomyces cerevisiae* and *Drosophila melanogaster* express only a single protein corresponding to UAP56, while human and mouse cells express both mRNAs encoding DDX39 and UAP56 (Pryor et al., 2004). Possibly the proteins have similar functions, since both proteins interact with the ALY/REF and are able to rescue the loss of Sub2p (the yeast homolog of UAP56) (Pryor et al., 2004). Aligning the amino acid sequences of the two proteins (Figure 31), the proteins are 90% identical (Figure 31 and (Pryor et al., 2004)). There are references concerning DDX39 in the literature, but its role has not been

explored before in similar assays. Therefore, it is interesting to compare the two highly similar proteins in terms of their function.

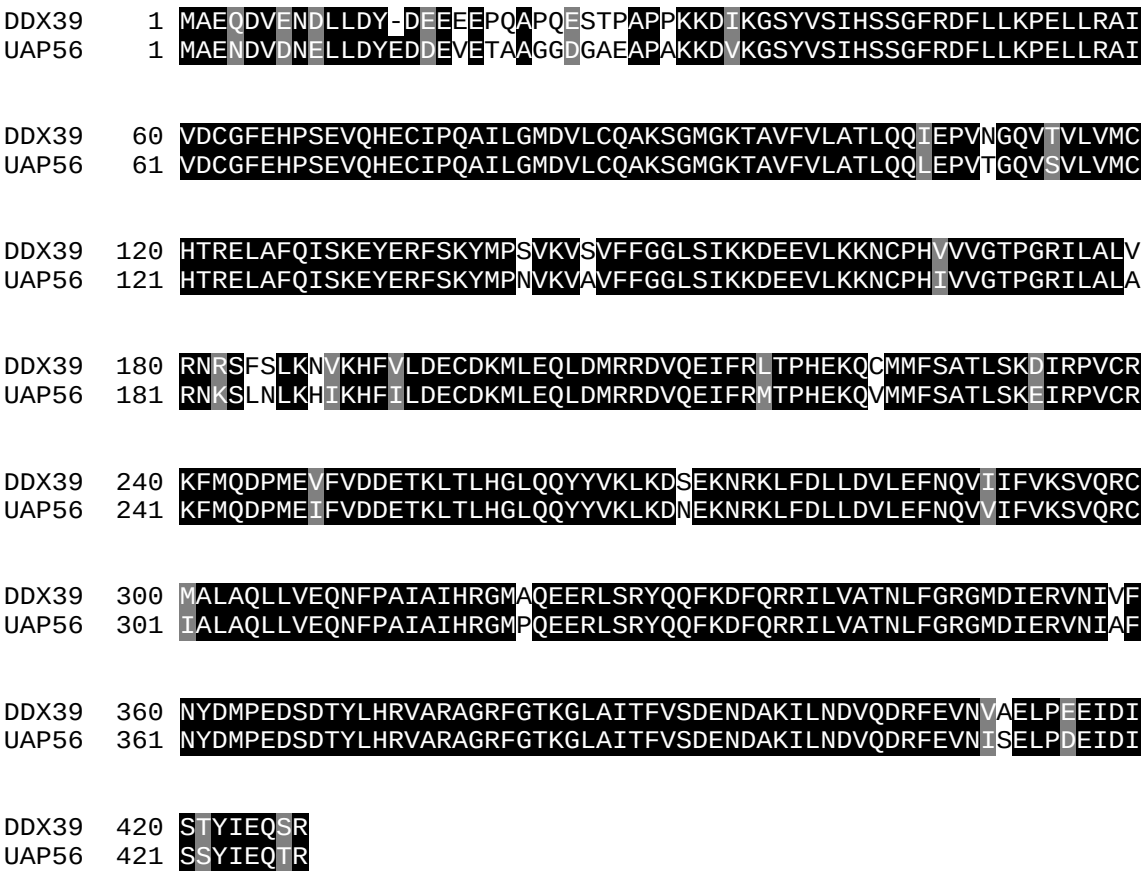


Figure 31. The proteins DDX39 and UAP56 are 90% identical. Alignment of the amino acid sequence of these two proteins showed that they have 384/428 identical (89.72≈ 90 %) residues. The alignment was performed using ClustalW2.

Having in hands a functional *in vitro* splicing system as well as a tool to specifically disassemble exon junction complexes without interfering with splicing or altering the structure of the mRNA, the deposition of the mammalian TREX complex was investigated. In particular, the following questions were addressed: is the

deposition of UAP56, ALY/REF and DDX39 proteins splicing-, EJC- or cap-dependent?

3.5.2 TREX complex deposition can be studied using an *in vitro* splicing system

The first step was to test whether the *in vitro* splicing system described in section 4.2 could be used to study TREX complex deposition. *In vitro* splicing reactions using the MINX pre-mRNA were performed in HeLa nuclear extract supplemented with splicing-competent whole cell extracts from HEK293 cells expressing different FLAG-tagged TREX proteins and the EJC protein Y14 (FLAG-Y14) as a control. During the splicing reaction, these tagged proteins are deposited onto the RNA and allow the immunoprecipitation of splicing-specific mRNPs. From these splicing reactions, FLAG-tagged proteins are immunoprecipitated in association with the spliced mRNAs. A number of factors associated with mRNA export (THOC1-7, UAP56, DDX39, NXF1 and TAP/p15) have been tested for their ability to co-immunoprecipitate spliced mRNA (some of these factors are shown in figure 32).

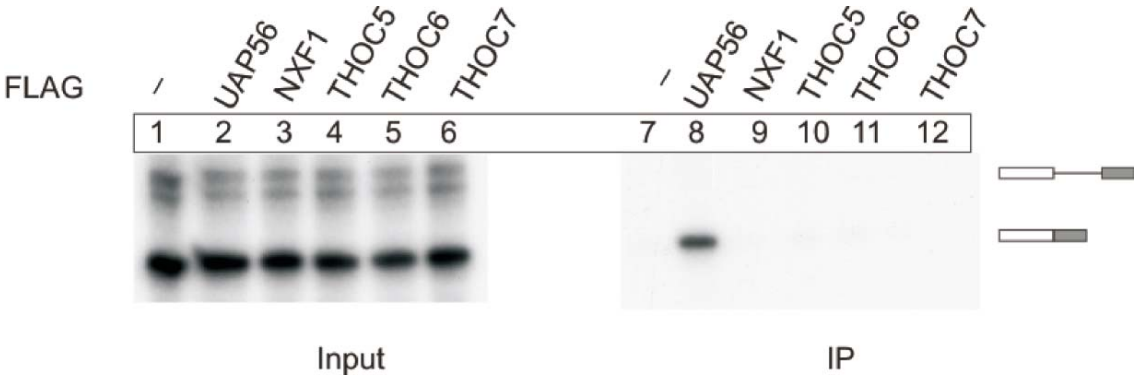


Figure 32. UAP56 specifically precipitates spliced mRNAs, while THOC5-7 associate neither with the pre-mRNA nor with the spliced product in a stable way. The *in vitro* splicing system (described in section 4.2) was used to test the deposition of TREX components. Splicing reactions in HeLa cell nuclear extract using MINX as a substrate were supplemented with 293 cell extracts expressing the indicated tagged protein (UAP56, NXF1 and THOC5-7)

or unfused FLAG-tag as a negative control. Reactions were immunoprecipitated using FLAG affinity agarose.

The components of the mammalian THO complex (THOC5-7 or else called fSAP79, fSAP35 and fSAP24 respectively shown in figure 32 and THOC1-3 referred to in humans as hHrp1, hTho2 and hTex1 respectively) associated neither with the pre-mRNA nor with the spliced product in our experimental system. This lack of association of these proteins hints that either these proteins are recruited in a splicing-independent fashion [possibly earlier, since they play a role in transcription elongation (Fischer et al., 2002; Jimeno et al., 2002)] or that the protein region responsible for the interaction with the RNA is not accessible, in order to immunoprecipitate the stable complex under the conditions used.

The THO complex (THOC1-3: hHrp1, hTho2 and hTex1 respectively and THOC5-7: fSAP79, fSAP35 and fSAP24) does not precipitate any RNA in our *in vitro* splicing experimental system, while ALY/REF, UAP56 and DDX39 were deposited during splicing. This shows that UAP56, DDX39 and ALY/REF assemble onto the RNA during splicing (Figures 32 and 33), while the THO complex possibly assembles at an earlier point or we cannot address its deposition under the experimental conditions used. The next steps address the role of splicing to the recruitment of ALY/REF, UAP56 and DDX39 proteins. For the rest of the thesis, these proteins UAP56, DDX39 and ALY/REF will be referred to collectively as TREX complex (describing a simplified TREX complex).

3.5.3 TREX deposition is splicing-independent

It has been reported that the deposition of the TREX complex is splicing-dependent (Masuda et al., 2005). In order to study the splicing-dependent deposition of TREX, a mutant MINX transcript lacking the intron was generated. This mutant transcript has been previously used in section 3.2, figure 24 C.

Radiolabeled transcripts, intron-containing and intron-lacking, were transcribed *in vitro*. Then they were subjected to *in vitro* splicing in HeLa nuclear extract and HEK 293 whole cell extracts with FLAG-tagged TREX complex components. The EJC protein Y14 was used as a positive control for the assay and splicing-dependent deposition.

During the splicing reaction, these tagged proteins bind to specific complexes and allow the immunoprecipitation of splicing-specific mRNPs. If a protein binds to the intron-lacking substrate, binding is interpreted as being splicing-independent.

As a first step the intron-containing transcript was analyzed. From these splicing reactions, the FLAG-tagged proteins (UAP56, DDX39, ALY/REF and Y14) were immunoprecipitated in association with the spliced mRNAs (Figure 33 A).

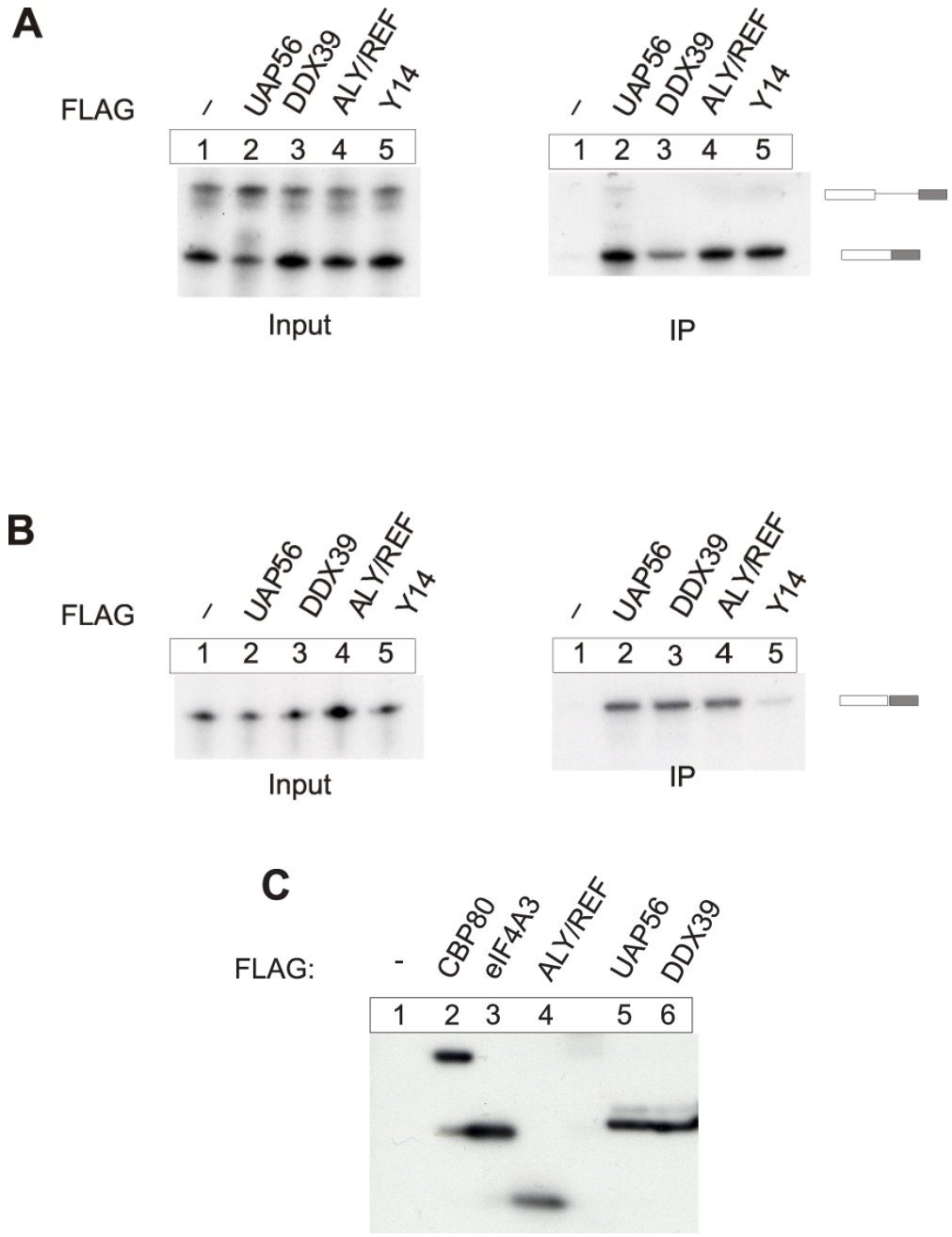


Figure 33. Deposition of the TREX complex components, UAP56, DDX39 and ALY/REF is splicing-independent. Splicing reactions in HeLa cell nuclear extract using MINX as a substrate were supplemented with HEK 293 cell extracts expressing the indicated protein (ALY/REF, UAP56, DDX39, Y14) or unfused FLAG-tag as a negative control. Reactions were immunoprecipitated using FLAG affinity agarose. **(A)** For the splicing reactions a MINX intron-containing substrate was used. **(B)** For the splicing reactions, a MINX intronless (MINX Δi) substrate was used. **(C)** Expression of the FLAG-proteins used in (A) and (B) is determined by immunoblot analysis with a FLAG antibody.

In order to test whether the immunoprecipitations are specific for spliced RNAs, we have used an intronless MINX transcript (MINX Δi). As expected, only traces of the intronless MINX transcript are pulled down under the same experimental conditions by the EJC protein Y14 (Figure 33 B). This intron-lacking transcript co-immunoprecipitated with FLAG- ALY/REF, -UAP56 and -DDX39 (Figure 33 B). Figure 33 C shows an immunoblot analysis of FLAG-proteins expressed in the extracts used for the *in vitro* splicing and immunoprecipitation. The expression levels of the FLAG-tagged proteins are similar.

In conclusion, splicing of MINX is not required for the deposition of UAP56, DDX39 and ALY/REF. This observation does not exclude a potential contribution of the EJC to the TREX complex deposition. The semi-quantitative nature of the assay does not allow the estimation of the deposition's stoichiometry. A certain amount of the complex components could interact with the mRNA without the splicing-dependent EJC deposition, but the rest could require the presence of EJC. In order to specifically address the role of the EJC to the deposition of the TREX complex, two approaches described in the next section were followed.

3.5.4 Recruitment of UAP56 and DDX39 is EJC-independent, while ALY/REF interacts with EJC

It has been reported that in metazoans the members of the TREX complex are recruited to the mRNA by the EJC, or that the members of TREX constitute part of the EJC (Gatfield et al., 2001; Le Hir et al., 2001b; Le Hir et al., 2000). The role of the EJC to the deposition of TREX was addressed by two different approaches.

The first approach makes use of a transcript with a short first exon (of only 15nts) that cannot support the assembly of the exon junction complex (Le Hir et al., 2001b; Merz et al., 2007). The EJC is assembled ~20-24 nucleotides upstream of an exon-exon junction (Le Hir et al., 2000). The spliced products of both, the intron-containing (Figure 34 A) and intron-lacking (Figure 34 B) MINX transcripts, co-immunoprecipitated with FLAG- ALY/REF, -UAP56 and -DDX39. As expected,

neither the spliced product nor the intron-lacking transcript co-immunoprecipitated with FLAG-Y14, confirming that the EJC is not assembled on the construct with the short 5'exon (Figure 34 A and B, lane 5). This observation already suggests that EJC is not the only determinant for the recruitment of the TREX factors.

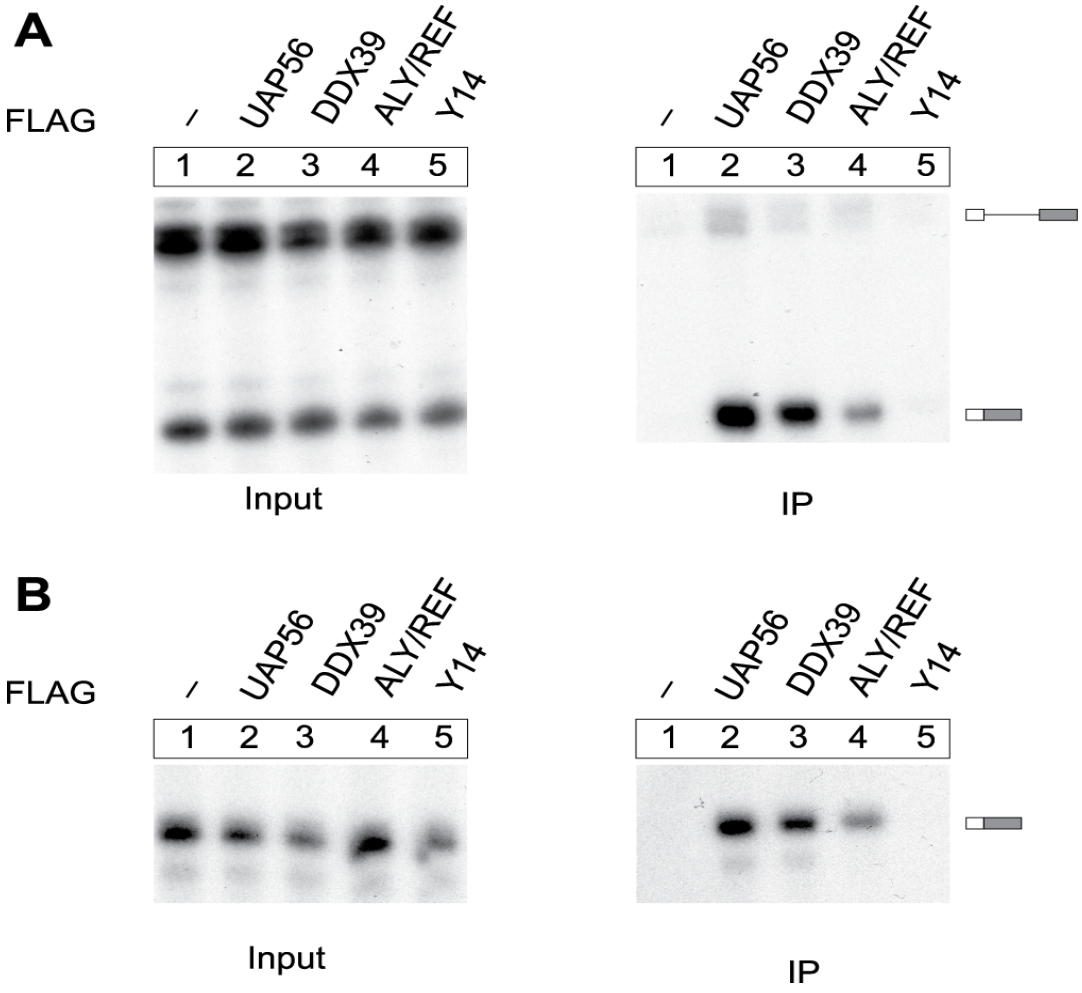


Figure 34. The EJC is not crucial for the recruitment of the TREX factors. Splicing reactions were performed using a substrate with a short first exon that fails to assemble an EJC. **(A)** For the splicing reactions a MINX intron-containing substrate was used. **(B)** For the splicing reactions a MINX intronless substrate was used.

As described in section 3.4, the addition of recombinant PYM to the splicing reactions reproducibly and specifically dissociated assembled EJCs. This function of rPYM was now employed to analyze the interdependence of the TREX and the exon junction complexes. If the TREX components were members of the EJC or if their deposition was depending on the EJC, both complexes would be expected to dissociate once rPYM was added. This dissociation will decrease the co-immunoprecipitation of MINX mRNA.

rPYM was thus added to the splicing reactions containing either FLAG-ALY/REF, -UAP56, -DDX39 or -Y14 and FLAG-immunoprecipitations were performed. As shown before, Y14 co-immunoprecipitation was limited after the addition of recombinant PYM in a concentration-dependent manner (Figure 35 A, lanes 11-13). No change was observed to the amount of spliced product that was immunoprecipitated by UAP56 and DDX39. Therefore, the addition of rPYM has no effect on the deposition of these TREX components (Figure 35 A, lanes 2-10). In the case of ALY/REF, the association of the protein with the spliced mRNA was reduced significantly after the addition of rPYM (Figure 35 A, lanes 8-10). The reduction has been quantified, which is shown in figure 35 B.

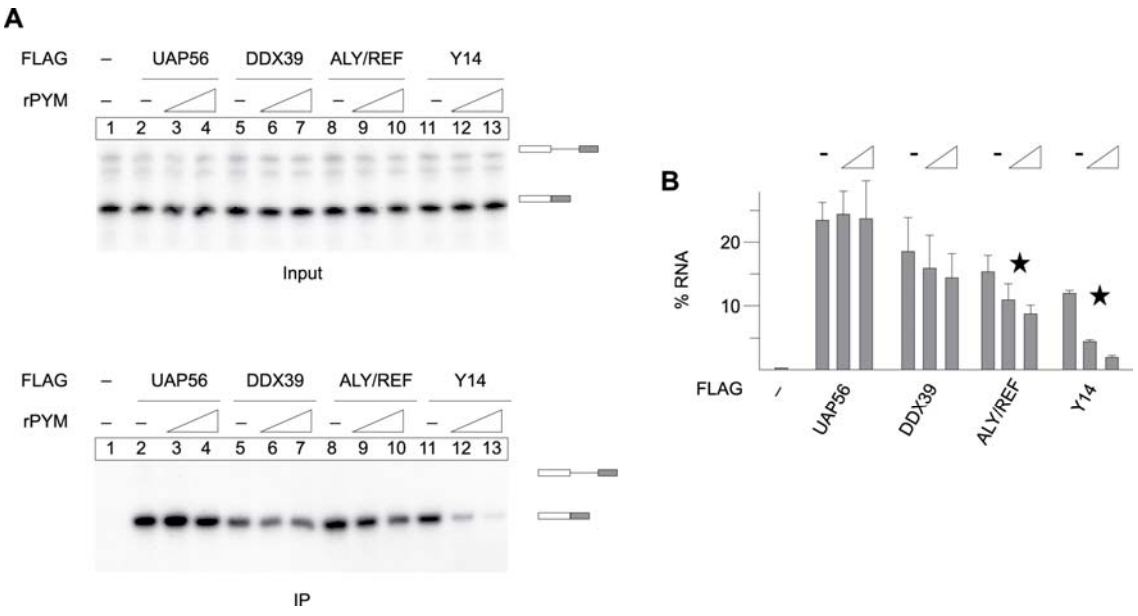


Figure 35. rPYM dissociates ALY/REF significantly, but does not interfere with UAP56 and DDX39 binding. (A) Splicing reactions were performed as described before. Reactions were immunoprecipitated using FLAG affinity agarose. Increasing amounts of recombinant PYM (rPYM) were added to the reactions as shown. **(B)** The graph is showing the % of RNA precipitated through the stated protein after the addition of increasing amounts of rPYM. The black asterisk indicates that the changes (for proteins ALY/REF and Y14) depicted in the graph are significant. A t-test was performed for the significance of the results: the $p< 0.01$ for ALY/REF and $p< 0.001$ for Y14. This experiment was repeated 3 times.

The approach with the short first exon showed EJC-independent TREX deposition, which can be concluded from the finding that UAP56, DDX39 and ALY/REF precipitated both the intron-containing and the intron-lacking mRNAs. The absence of EJC was confirmed by the absence of its component Y14. When a more subtle method was used, namely the addition of rPYM, UAP56 and DDX39 did not show any significant change in the amount of mRNA they precipitated. ALY/REF showed a significant change of the amount of mRNA precipitated, which was dose-dependent. When more rPYM was added, less mRNA was precipitated. Taken together, these data hint that the depositon of UAP56 and DDX39 is not directly affected by the EJC, while the deposition of ALY/REF seems to be.

3.5.5 TREX is deposited both to 5' and 3'ends of the mRNA

Cheng et al. suggested that the human export machinery is recruited to the 5'end of the mRNA. Its deposition was reported to be cap-dependent, binding in a region upstream of the EJC and only to the first exon of a transcript (Cheng et al., 2006).

The next step was to investigate the role of the 5'end of the mRNA to the deposition of the TREX complex and the potential contribution of the EJC. In order to address these questions, an RNase H digestion approach was used. The experimental strategy was the following: after the completion of the splicing reaction (after 120 minutes), a cDNA oligonucleotide complementary to position -36 of the first exon was added to the reaction. The addition of the complementary oligonucleotide activated the RNase H present in the extract which specifically cleaved the RNA substrate at position -36 of the first exon. This digestion generated 2 distinct fragments, the 5'end fragment or cap fragment and the 3'end fragment or so called EJC fragment (the position that the oligonucleotide binds and the cleavage products are depicted schematically in figure 36).

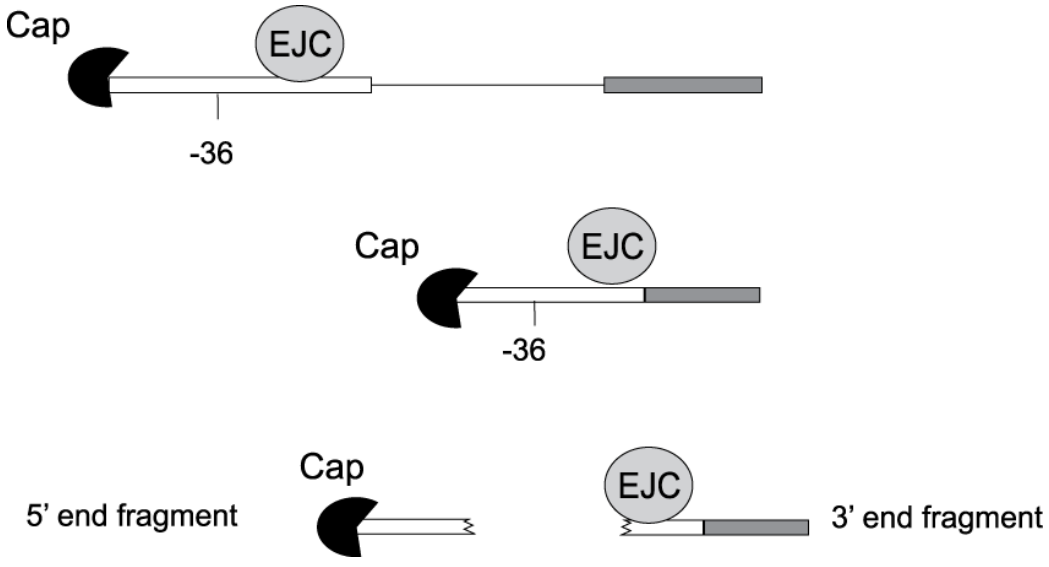


Figure 36. Schematic representation of the RNase H digestion assay. The cDNA oligonucleotide binds to position -36 (the oligonucleotide extends from position -42 to -31) of the 1st exon (position -36 refers to 36 nucleotides starting to count from the end of the 1st

exon). When the cDNA oligonucleotide is added to the splicing reaction, the endogenous RNase H activity digests the pre-mRNA. This digestion gives rise to two fragments: the 5'end that carries the cap and the 3'end fragment that carries the EJC.

In figure 37 A, splicing reactions and FLAG immunoprecipitations without RNase H digestion are shown. The bottom panel (B) shows the outcome of the FLAG immunoprecipitations after RNase H digestion. All 3 splicing products (the fully spliced transcript, the cap or 5'fragment and the EJC or 3' fragment) were immunoprecipitated by ALY/REF, UAP56 and DDX39. UAP56 and DDX39 exhibited a similar behavior, binding the cap fragment and the EJC fragment strongly (Figure 37 B, lanes 7 and 8 respectively). ALY/REF showed a weaker binding to both fragments (Figure 37 B, lane 9). While Y14 co-immunoprecipitated the fully spliced product and the EJC fragment, but not the cap fragment (Figure 37 B, lane 10).

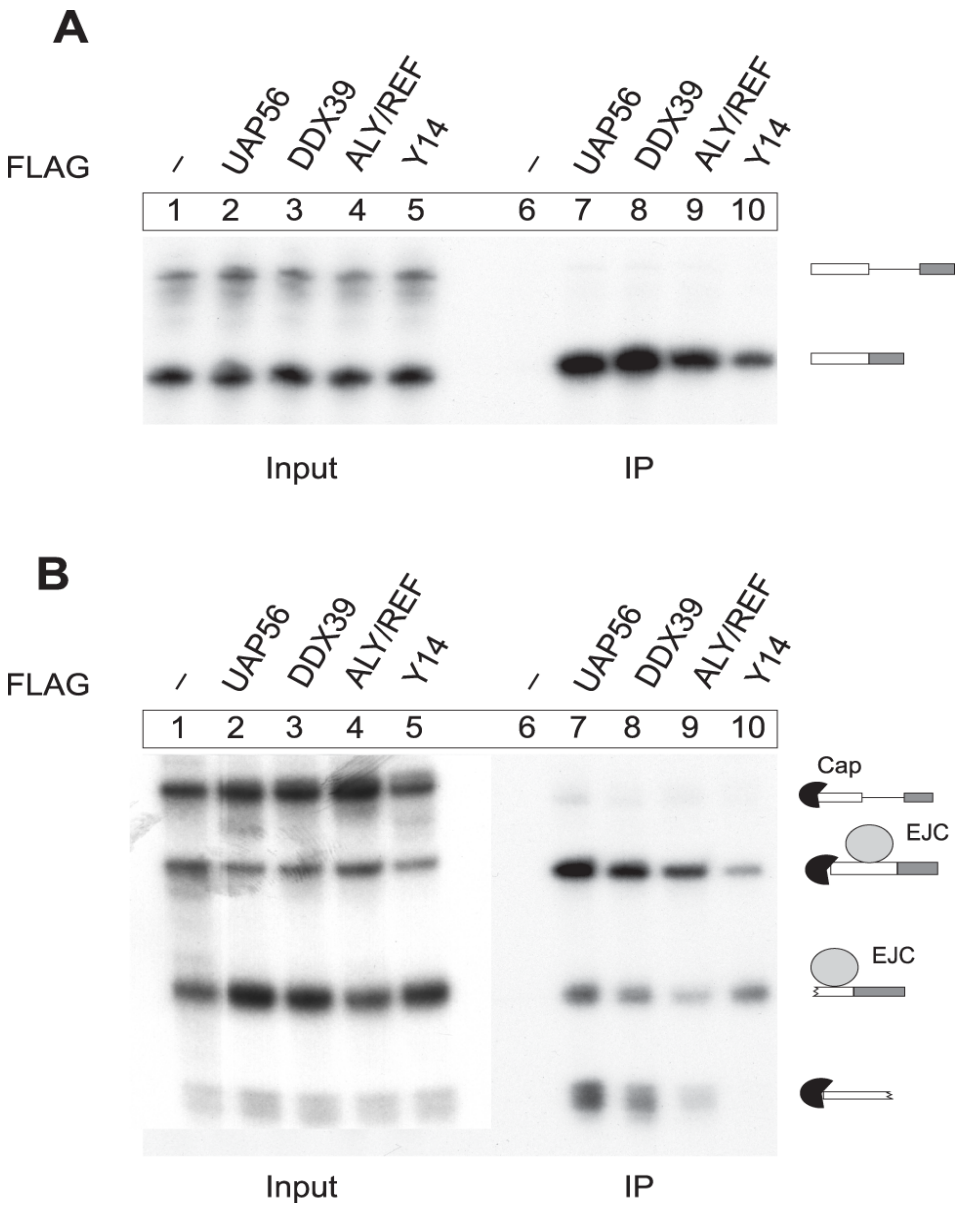


Figure 37. TREX complex proteins are deposited both to the cap-containing 5'end of the mRNA and the EJC-containing 3'end. (A) Splicing reactions and immunoprecipitations were performed as described before. UAP56, DDX39, ALY/REF and Y14 specifically precipitate the spliced product. **(B)** After splicing, RNase H digestion was performed. The resulting fragments, shown schematically on the side, were differentially precipitated by the tested proteins.

The same experiment, as the one described above (Figure 37), was performed with a MINX transcript lacking an intron. Instead of the EJC protein Y14 (which does not bind to an intronless transcript), CBP80 was used as a control. FLAG-UAP56 and FLAG-DDX39 precipitated the intronless transcript strongly (Figure 38 A) (as shown before). FLAG-ALY/REF and FLAG-CBP80 precipitated the intronless transcript as well (Figure 38 A). Splicing reactions of the MINX Δi and FLAG immunoprecipitation with RNase H digestion are shown at the bottom panel (B) of figure 38. FLAG-UAP56 precipitated all 3 splicing products strongly (the fully spliced transcript, the cap fragment and the EJC fragment). DDX39 precipitated all fragments as well, but to a smaller extent. ALY/REF showed a weaker binding to both fragments and the fully spliced transcript. CBP80 co-immunoprecipitated the fully spliced product and the cap fragment, but not the EJC fragment.

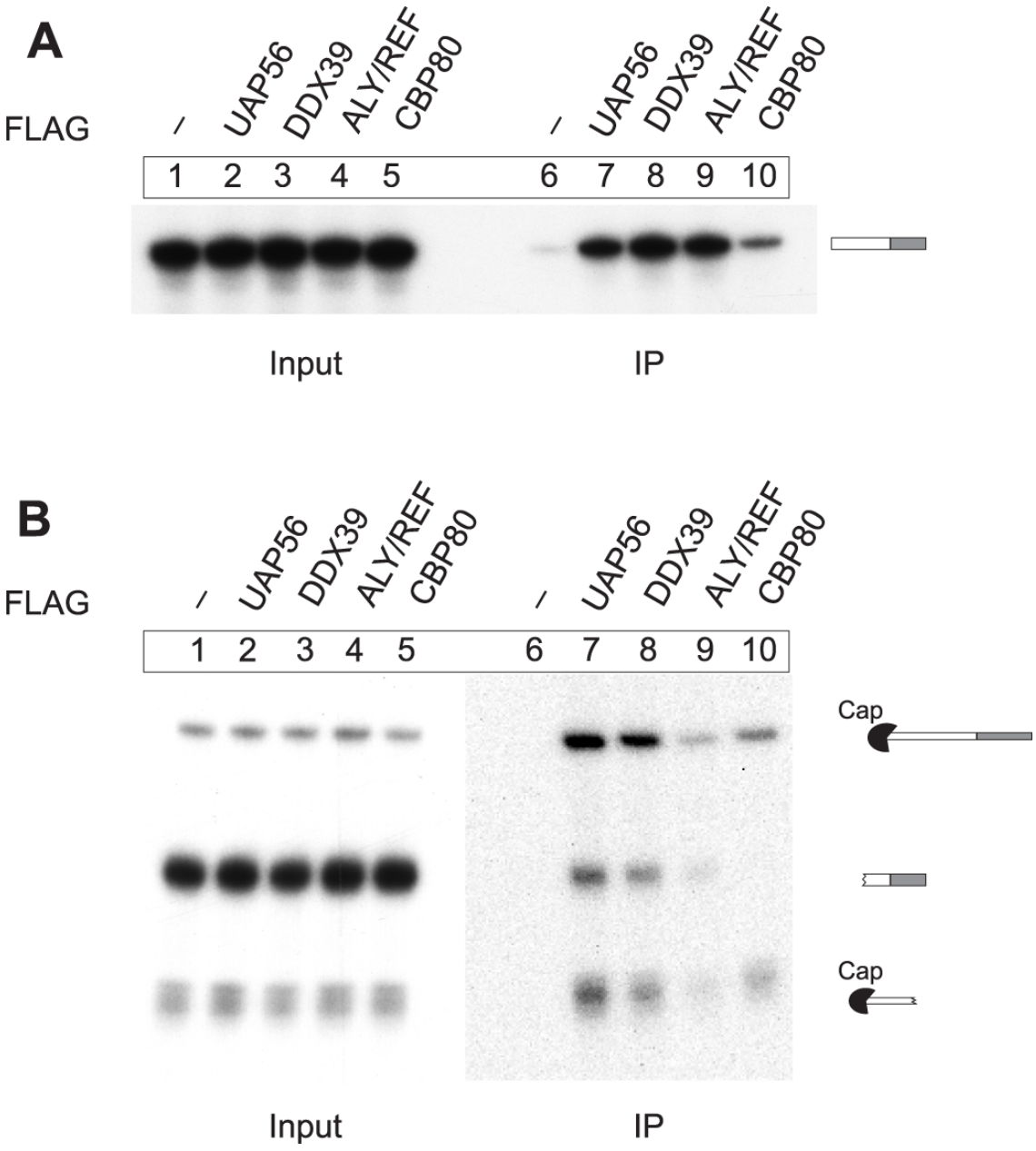


Figure 38. TREX proteins are deposited both to the 5' and 3'ends of an intronless MINX transcript. (A) Splicing reactions and immunoprecipitations were performed using MINX Δi. UAP56, DDX39, ALY/REF and CBP80 specifically precipitated the intronless transcript. **(B)** After splicing, RNase H digestion was performed. The resulting fragments, shown schematically on the side, were differentially precipitated by the tested proteins.

In conclusion, the TREX complex proteins UAP56, DDX39 and ALY/REF are deposited both to 5'and 3'ends of an mRNA. This observation applies to both intron-containing and intronless transcripts. UAP56 and DDX39 exhibited a similar behavior, showing a strong precipitation of the cap fragment both in the intron-containing and the intronless transcript. The protein ALY/REF binds strongly to the spliced mRNA, but very weakly to both 5'and 3'ends. The conclusion that derives from these observations is that for the recruitment of ALY/REF both the cap and the EJC are required or factors that are present both in 5'and 3'ends. In the case of the intronless transcript, proteins UAP56 and DDX39 assemble on both the 5'end or cap fragment and the 3' fragment, while ALY/REF precipitated both fragments very weakly. It should be noted that the binding of ALY/REF was weaker in the case of the intronless transcript than the binding of the intron-containing transcript.

3.5.6 Functional characterization of ALY/REF and UAP56

Trying to elucidate the role of ALY/REF and UAP56 in the deposition of the TREX complex, several mutants were generated. The goal was to identify mutants that precipitate specifically either only the cap fragment or the EJC fragment. The C-terminus of ALY/REF is required for the interaction with UAP56 (Luo et al., 2001). The mutants of ALY/REF generated were the following: ΔC16 is a mutant form lacking the C-terminal 16 amino acids, ΔC53 is a mutant form that lacks 53 C-terminal amino acids and mutant ΔN80 lacks 80 N-terminal amino acids. The C-terminal mutants were generated trying to abolish the interaction between ALY/REF and UAP56. The N-terminal mutant would be useful to investigate the contribution of the N-terminus to the protein interactions.

Three mutants of UAP56 were generated. UAP56K95N is a lysine to asparagine mutant at position 95 that disrupts the ATP binding site (Kota et al., 2008). The ATP binding site stretches from amino acid 92 to 96. The other two mutants were generated to lack one part of the helicase domain. Mutant ΔC171 (a fragment of the 1-256 protein amino acids) lacked 171 C-terminal amino acids, hence the helicase domain. A third

mutant lacking 251 amino acids from the N-terminus, was termed ΔN 251 (containing amino acids 252-427). This mutant lacked the first protein domain- DEAD box helicase domain. Figure 39 shows a schematic representation of the protein domains of the wild type proteins and the mutant forms generated.

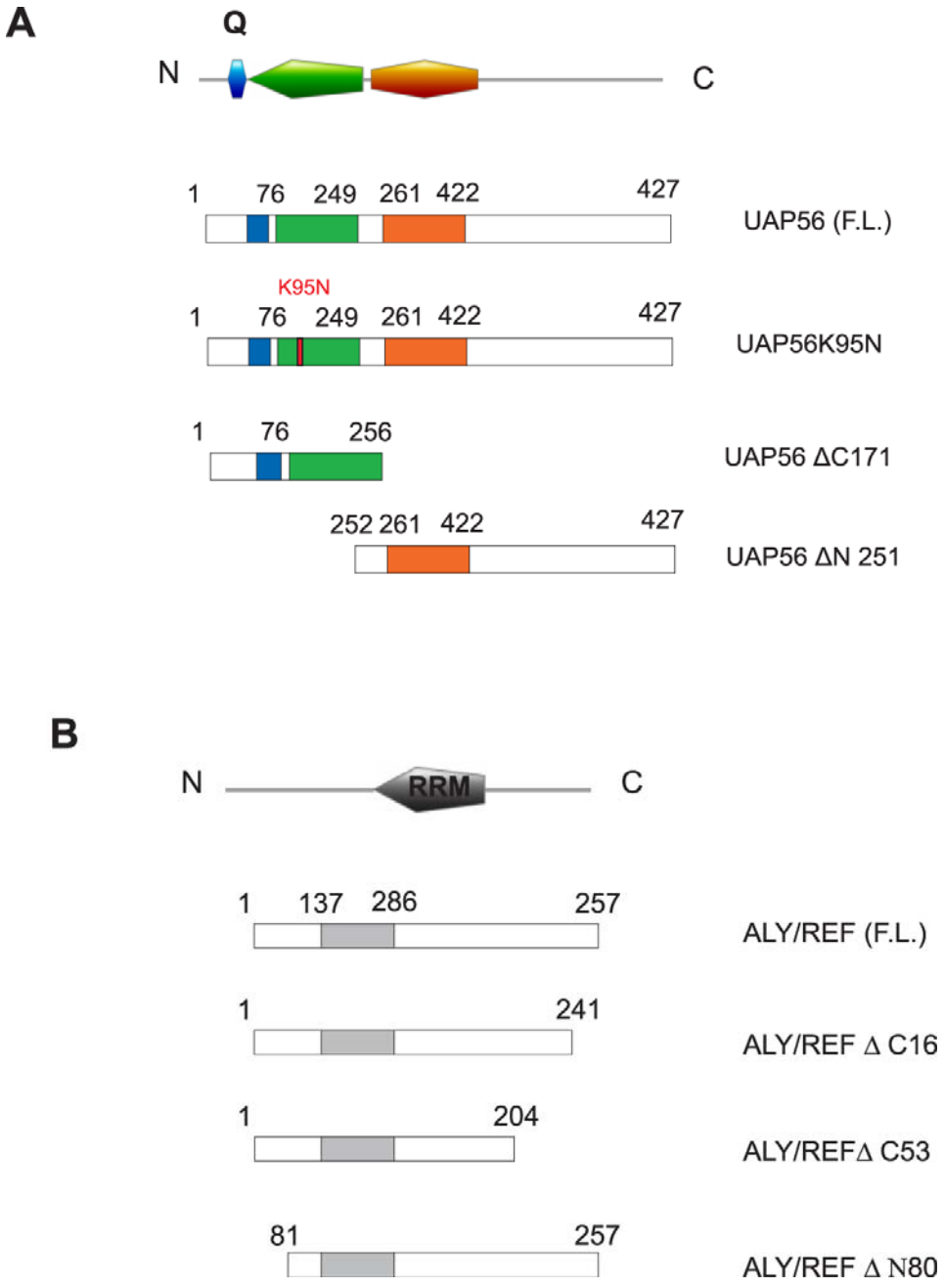
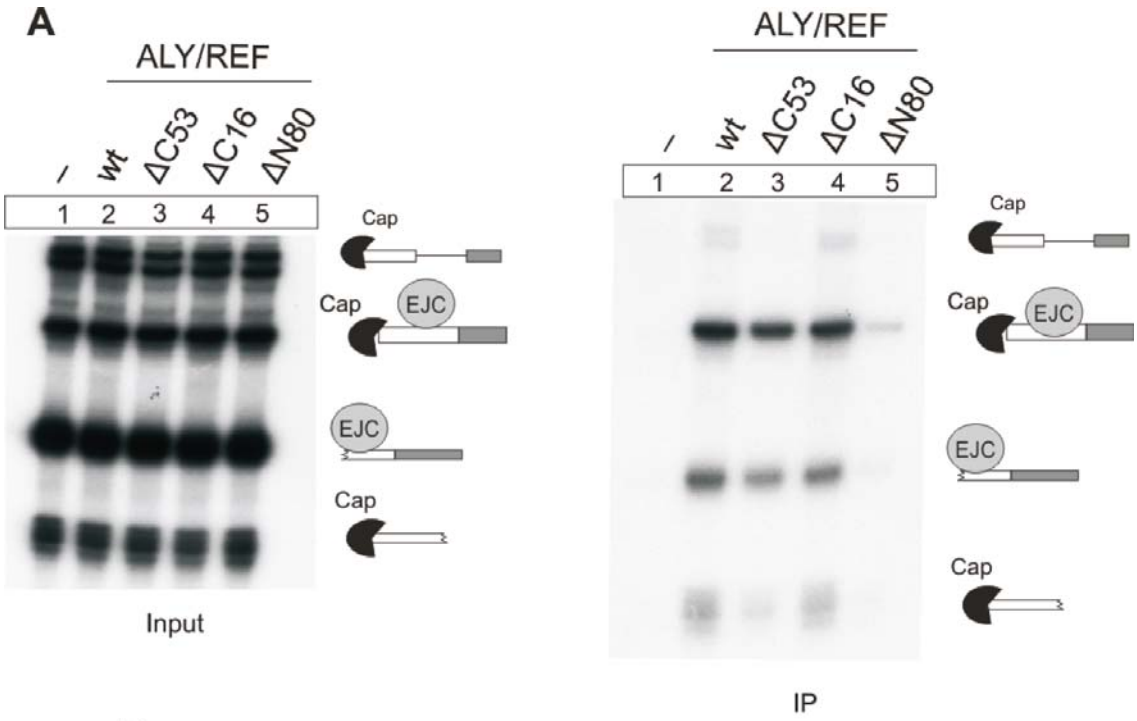


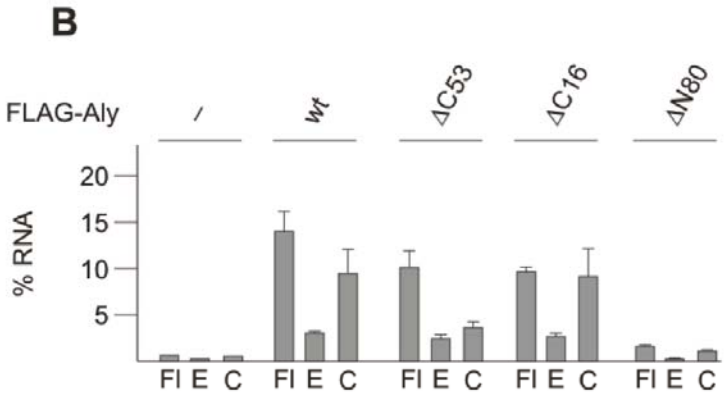
Figure 39. Schematic representation of the (A) UAP56 and (B) ALY/REF wild type proteins and mutants studied. On the top panel of both, a graphical representation of the protein domains made using the MyDomains-Image Creator of ExPASy-Prosite site. The rest of the lanes show the different truncation mutants used for the *in vitro* studies. (**A)** Schematic representation of the UAP56 deletion mutants. The different domains are displayed as coloured geometrical shapes: the blue stands for a predicted Q motif, the Helicase ATP binding domain is coloured green and the Helicase C terminal domain is shown in orange. The sizes of the respective N- or C-deletions are indicated. (**B)** Schematic representation of the ALY/REF deletion mutants. The one RRM domain is depicted in gray. The sizes of the respective N- or C-deletions are indicated. The abbreviation F.L. stands for full length. **Q motif** is a conserved element and it is involved in adenine recognition and in ATPase activity of DEAD-box proteins (Cordin et al., 2004).

These analyses aimed at identifying mutants that show reproducible changes in binding to the 5'-cap or the 3'-EJC fragments. Such mutants could also represent valuable tools for the future analysis of the mechanisms of mRNA export.

3.5.6.1 The ALY/REF residues 204-241 are required for binding to the cap fragment, but not to the EJC fragment.

The ALY/REF mutants were tested for their ability to interact with the cap or the EJC fragments. Splicing reactions, as described previously, were performed and RNase H digestion followed. The splicing reactions were carried out in a combination of HeLa nuclear extract and HEK 293 whole cell extracts expressing FLAG-tagged ALY/REF forms, wild type and mutants (Figure 40). As described before, wild type ALY/REF precipitated not only the full length but also both, the 5'and 3'end fragments. The mutant ΔC16 showed similar precipitation pattern to the wild type (Figure 40, lane 4 and 2 respectively) and mutant ΔN80 did not precipitate any of the resulting products (Figure 40, lane 5). The mutant ΔC53 displayed reduced binding of the 5'fragment, while the binding to the 3'fragment was comparable to that of the wild type protein (Figure 40, lane 3).


A
ALY/REF
/
wt
ΔC53
ΔC16
ΔN80
1 2 3 4 5
Cap
Cap
EJC
EJC
Cap
Input
ALY/REF
/
wt
ΔC53
ΔC16
ΔN80
1 2 3 4 5
Cap
Cap
EJC
EJC
Cap
IP


B
FLAG-Aly
/
wt
ΔC53
ΔC16
ΔN80
% RNA
20
15
10
5
Fl E C Fl E C Fl E C Fl E C Fl E C

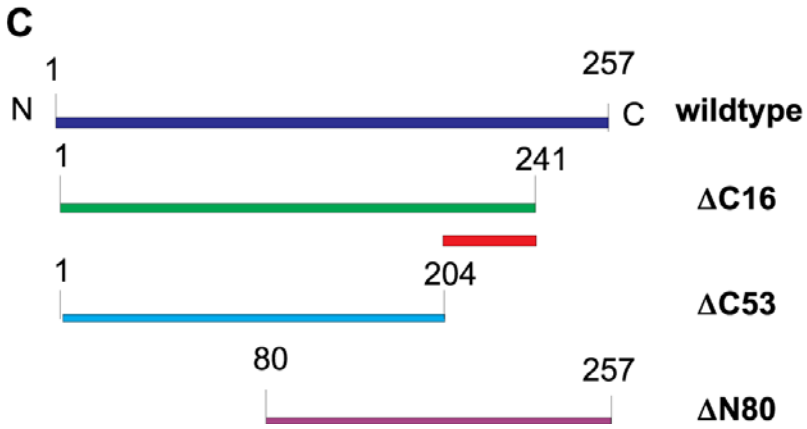

C
1
257
N
C
wildtype
1
241
ΔC16
1
204
ΔC53
80
257
ΔN80

Figure 40. The ALY/REF residues 204-241 are required for binding to the cap fragment, but not to the EJC fragment. (A) Splicing reactions, immunoprecipitations and RNase H digestion were performed as described before. The resulting fragments, shown schematically on the side, were differentially precipitated by the tested mutants. **(B)** The graph is showing the % of RNA fragment precipitated through the mutant of ALY/REF protein. This experiment was repeated 3 times. **(C)** The diagram depicts the ALY/REF wild type and mutants in a simple linear form. The residues 204-241 are depicted in orange. **Fl**: Full Length, **E**: EJC fragment, **C**: cap fragment

The generation of ALY/REF mutants led to the identification of mutant ΔC53, which showed a preferential binding for the EJC fragment rather than the cap fragment. This indicated that residues 1-204 are sufficient for the interaction with the EJC fragment, but not with the cap fragment. This observation shows that residues 204-257 are important for the interaction of ALY/REF with the cap fragment. Mutant ΔC16 showed a binding of the cap fragment comparable to the wild type. This binding is lost if more residues are removed from the C'-terminus. It seems that residues 204-241 are necessary for the interaction of ALY/REF with the cap. Mutant ΔN80 did not precipitate any of the fragments, indicating that residues 1-80 are crucial for these interactions. Further mutational analysis will help to map the protein regions involved at the interactions more finely.

3.5.6.2 UAP56 mutants do not precipitate unspliced pre-mRNA or spliced mRNA fragments.

Precipitations, as described at the previous section 3.5.6.1, were performed from splicing reactions with extracts expressing FLAG-tagged forms of UAP56 (Figure 41). In this case, only the wild type form of UAP56 precipitated all the products of the splicing and the RNase H digestion reactions. The mutants ΔC171 and ΔN175 exhibited a similar behavior, not precipitating any of the fragments (Figure 41, lanes 7 and 8 respectively). The data suggest that large deletions of either C'- or N'- terminus

of UAP56 destabilize the interaction with the mRNA. The point mutant UAP56K95N precipitated a reduced amount of the products, indicating that ATP binding may be required for the binding of this protein to the RNA. UAP56 has both RNA-stimulated ATP binding/hydrolysis activity and ATP-dependent RNA unwinding (Fleckner et al., 1997; Shen et al., 2008; Shen et al., 2007), therefore a disruption at the ATP binding site could lead to the abrogation of its function. Disruption of the ATP domain abrogates the ATP-dependent binding of the UAP56 to RNA and/or changes the protein conformation.

As in the case of ALY/REF, further mutational analysis may yield mutants lacking interactions with either the 5' end or the 3' end.

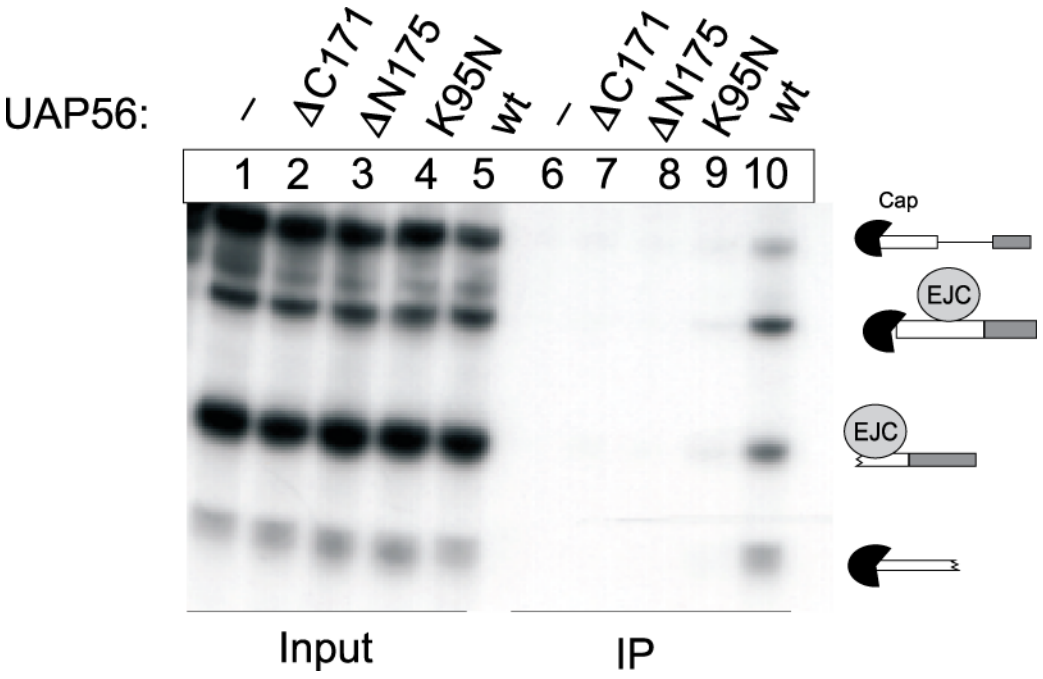


Figure 41. UAP56 mutants do not precipitate unspliced pre-mRNA or spliced mRNA fragments. Splicing reactions, immunoprecipitations and RNase H digestion were performed as described before. The resulting fragments, shown schematically on the right side, were differentially precipitated by the tested mutants. Only the wild type UAP56 precipitated the generated fragments. **Fl**: Full Length, **E**: EJC fragment, **C**: cap fragment

In this section (3.5), the deposition of proteins UAP56, DDX39 and ALY/REF onto mRNAs was studied using our *in vitro* splicing system. It is the first time that DDX39, a protein 90% identical to UAP56, is used in an in vitro study and its deposition is directly compared to UAP56. To sum up the results of this section, the deposition of TREX components, UAP56, DDX39 and ALY/REF is splicing-independent. In addition, the deposition of UAP56 and DDX39 is EJC-independent, while ALY/REF interacts with EJC. As a next observation, TREX is deposited both to 5'and 3'ends of the mRNA. Through mutation analyses, the ALY/REF region, consisting of residues 204-241, was identified as necessary for binding to the cap fragment, but not to the EJC fragment (the cap and EJC fragments refer to the fragments generated from the RNase H digestion performed on a MINX transcript).

4. DISCUSSION

4.1 Affinity selection of mRNPs spliced *in vitro* to investigate potential substrate specific differences in the composition of EJCs

Gaining insight into the different steps of mRNA metabolism can be achieved by studying the protein composition of the respective mRNPs. The elucidation of the protein components requires the affinity purification of RNAs together with the accompanying proteins. To this end, an RNA affinity purification system was developed to study the mRNPs' composition. The primary goal was to develop a system to affinity purify spliced mRNPs using GRNA chromatography. The ultimate goal was to use the established system to investigate the composition of EJCs assembled on different NMD substrates.

4.1.1 Affinity purification of mRNPs employing GRNA Chromatography

In this thesis, the main process under investigation is splicing. During this process, the protein complex of interest, the exon junction complex, is assembled onto the mRNA. The experimental strategy included the use of transcripts with and without RNA affinity tags. These transcripts are transcribed and spliced *in vitro*, mimicking the procedures in the cell that result in protein deposition. These spliced mRNPs were selected and their protein composition was investigated.

RNA affinity tags have been used in the literature for similar approaches. The method of affinity purification employed is GRNA chromatography, a module developed by Kevin Czaplinski in the Mattaj laboratory (Czaplinski et al., 2005). In the initial experimental plan, the protein composition was to be assessed by mass spectrometry. This technique could have led to the potential identification of novel associated factors.

We succeeded in purifying spliced and unspliced mRNAs from HeLa nuclear extract, but we were not able to determine their protein composition by mass spectrometry. The mass spectrometry analysis did not identify EJC components within the affinity purified mRNPs. The protein gel bands investigated were identified as abundant cellular proteins. Most likely the enrichment of proteins in the affinity purified sample was not enough for the mass spectrometry analysis to identify the proteins. In addition, the background is quite high, preventing the possible identification of proteins under investigation. Alternatively the sample could have been damaged from the mass spectrometry procedure. Certain steps of the procedure could damage the sample, such as the generation of not properly-digested peptides, ionization by laser and/or the not-correct peptide mass assignment to proteins.

The second method employed, immunoblot analysis is very sensitive and used to detect the presence of a previously-known protein. The necessary prerequisite is the existence of an antibody against the protein of interest. A specific antibody is less sensitive to background. The production of antibodies against proteins of interest was one of the taken steps. The production of antibodies against EJC components is a valuable tool to detect proteins in purified samples and to study protein interactions. Proteins eIF4A3 and UAP56 were identified within the affinity selected mRNPs. These results are in agreement with previously published studies (Merz et al., 2007). Moreover, the fact that we could not detect Barentsz is in agreement with figure 26 and (Gehring et al., 2009b), which show that Barentsz binds to the core EJC after splicing.

The λN/boxB purification system adopted is based on the same principle as the Maltose Binding Protein (MBP)-MS2 purification system, used by other laboratories for the isolation of spliceosomal complexes. Using the MBP-MS2 purification system and mass spectrometry analyses, Makarov and colleagues have provided a list of the protein composition of the A, B, B* and C spliceosomal complexes (Makarov et al., 2002). Furthermore, they have identified proteins that interact with pre-mRNA. The outcome of such analyses (proteins identified) can vary among different studies, depending on the stringency conditions and on the transcript used for the affinity selection. Nonetheless, our understanding of the protein composition of spliced mRNPs is far from complete. During the progress of this work, a study using a similar technical

approach, the MBP-MS2 system, was published by the Luehrmann laboratory (Merz et al., 2007). Merz and colleagues have identified 45 proteins studying two different mRNPs, one assembled on AdML mRNA and one on β-globin mRNA (Merz et al., 2007).

In conclusion, the GRNA chromatography approach to purify spliced mRNPs is a powerful tool to study the biogenesis of mRNPs. The big advantage is that the conditions used are close to the physiological ones. Plus, the elution with glutathione guaranties the smooth and total elution of the bound mRNPs. A number of experimental details could possibly be improved. First of all, one can notice that the transcription of the tagged transcript does not yield a uniform product, but rather one main product, the transcript at the expected size, and multiple bands of smaller size. This observation suggests that the transcription of this transcript is not complete. One explanation could be that the quality and efficiency of the RNA polymerase is compromised due to high GC content of the boxB regions. In order to improve the quality of the transcript used for *in vitro* splicing, one could use gel extraction of the expected product. Such a solution was not applicable in this case, since a combination of labeled and unlabeled transcripts was used. Therefore, the quality of both species could not be assessed on a single gel. The quality of the labeled transcript was checked on a urea gel and it did not look uniform. The transcription procedure in the presence of radiolabeled GTP led to the production of multiple RNA fragments. The quality of the unlabeled transcript was checked on an agarose gel. Although the resolution of agarose gel is limited, the unlabeled transcript looked uniform having one product at the expected product size.

The production of shorter and incomplete products during transcription can be visible also later at the *in vitro* splicing products. One could notice that the splicing products of the tagged transcript include multiple bands of varying size. These incomplete products were not affinity purified, when GRNA chromatography was employed. This observation supports the previously stated incomplete transcription, which stops before the transcription of the 3'end where the boxBs are situated. Another issue is that the addition of the RNA hairpins seems to diminish the splicing efficiency of the transcript. This could be attributed to the fact that the size of the three boxBs is equal to the size of the rest of the MINX transcript itself. Another possible explanation

is that the boxB sequence could be inhibitory. Trying to improve the splicing efficiency, one could potentially overcome these issues by increasing the size of the exons used or by adding spacer sequences among the different boxBs. These spacer sequences could reduce the density of CGs in the boxB sequence.

The absence of spliced mRNPs' specific proteins could be attributed to the rather stringent purification conditions. Our initial plan was to avoid heparin or similar competitors and keep the experimental conditions as close to physiological as possible. The use of competitors was considered necessary to avoid unspecific binding. Trying to avoid the background, the salt concentration was increased to 500 mM, since lower concentrations resulted to higher backgrounds. The possibility that the use of high salt concentration (500 mM) and heparin (12 μg/ml) has led to the dissociation of protein components could not be excluded. Merz and colleagues have used a much lower salt concentration (150 mM) and 8 μg/ml of heparin.

As becomes evident, the above described purification system can be used to address a number of biologically relevant questions, not only related to the EJC. Improving certain technical aspects could potentially lead to a powerful tool that can be used in the future to purify mRNPs for studying a number of interesting biological questions.

4.1.2 Differentially assembled EJCs

Our initial hypothesis was that different introns lead to EJCs of differential composition. This hypothesis was formulated based on the numerous observations that NMD can be activated from distinct routes. The term «distinct routes» refers to the protein factors that are sufficient and necessary to activate NMD. The goal was to splice these different substrates *in vitro* and using the GRNA Chromatography (section 3.1) to assess and analyze their protein composition focusing on the EJC (s).

The transcripts MINX and β-globin were efficiently spliced *in vitro.* The affinity selection of mRNPs seems to function properly, but the identification of protein components was not possible neither with mass spectrometry (as discussed above) nor

with immunoblot analysis. The immunoblot analysis could identify two EJC components, but not important others.

It still remains undefined whether the heterogeneity of NMD-activating complexes is reflected into EJC assemblies or if EJC is a uniform entity.

4.2 A system to study splicing-related deposition of proteins *in vitro*

To study splicing-related deposition of proteins, we employed the simple whole cell lysate system for *in vitro* splicing. Splicing reactions were performed in a combination of splicing competent HeLa nuclear extracts and the HEK 293 whole cell extracts. Such a system has been introduced by the Dreyfuss laboratory to address the assembly of the EJC (Kataoka et al., 2004). We have combined our optimal *in vitro* splicing conditions with this whole cell lysate system. This modified *in vitro* splicing system is a very valuable tool to address a number of biological questions. It is a simple and easy to use system, making use of whole cell extracts prepared from cells transfected with tagged-proteins of interest. Using such a system, one overcomes the need for separate production of recombinant proteins or the need for large amounts of protein-specific antibodies. A limitation of recombinant proteins is that they do not undergo post-translational modifications as many proteins physiologically do. Moreover, this strategy eliminates differences that can result from different efficiencies using various antibodies against the EJC components; thus allowing a direct comparison among the EJC proteins. Certain features of this system can be modified, for instance the structure of mRNA in order to block splicing at specific steps or by tagging different proteins the direct effect of each protein can be assessed. As shown in section 3.3, a mutation of the splice site leads to the blocking of the splicing process after the first step. Such a modification allows the dissection of the splicing process, facilitating the study of protein composition between the two steps. Another possible manipulation, not shown in this thesis, is to perform splicing reactions in extracts, prepared from cells where certain factors have been depleted and determine the changes in the composition of protein complexes.

The system is used to answer a number of questions, as shown in the present thesis. Section 3.2 of the results and figure 24 shows the splicing-dependent EJC deposition *in vitro.* The RNA products can be precipitated using the FLAG tag of a protein. Using this purification system, the EJC formations and interactions can be studied and the assembly of the EJC can be assessed. This system has been used for the rest of the results discussed in this thesis. Sections 4.3 and 4.4 discuss results obtained from experiments using this system, plus additional experiments included in two publications. Section 4.5 reveals the ability to study the deposition of additional protein complexes with this *in vitro* splicing system.

4.3 Ordered assembly of the Exon Junction Complex by the spliceosome

The core EJC consisting of the recombinant proteins eIF4A3, BTZ, MAGOH and Y14 self-assembles in the presence of ATP on an RNA substrate *in vitro* (Ballut et al., 2005). Ballut et al. in their study do not explain how these core EJC components bind to RNA during the spliceosome assembly. The crystal structure of the recombinant EJC core (Andersen et al., 2006; Bono et al., 2006) has provided a basis for our experimental strategy to analyze the assembly of the EJC by the spliceosome using an array of EJC protein mutants.

4.3.1 Definition and assembly of a minimal trimeric pre-EJC

We report that the exon junction complex is assembled in an ordered and hierarchical way during splicing. The proposed model of the EJC assembly pathway by the spliceosome is depicted schematically in figure 42. The first step includes the binding of eIF4A3 and MAGOH-Y14 to the spliceosome before exon ligation takes place (step 1). Taking into consideration the data presented here [for more details see (Gehring et al., 2009a)] and the crystal structure of the EJC core (Ballut et al., 2005; Bono et al., 2006), we propose that the spliceosome introduces eIF4A3 to the RNA with an

associated change of the conformation of this protein after ATP binding (step 2). This conformational change facilitates the binding of MAGOH-Y14 (step 3), completing the assembly of the first stable EJC intermediate that we refer to as minimal pre-EJC or trimeric pre-EJC. The suggested assembly pathway is in agreement with a previous analysis of EJC assembly (Reichert et al., 2002) and the analysis of purified spliceosomes and spliceosomal sub-complexes, where eIF4A3, MAGOH and Y14 were identified in the splicing complex B, the activated spliceosome (B*) and splicing complex C (Bessonov et al., 2008; Makarov et al., 2002; Makarova et al., 2004).

Our data also suggest that the trimeric pre-EJC represents the platform that physiologically interacts with additional EJC components. Affinity-purified spliced mRNPs lacking BTZ have been reported, suggesting that stable (pre-) EJCs without BTZ can occur under physiological conditions (Merz et al., 2007). Surprisingly, recombinant EJC proteins from bacteria do not interact with the spliceosome. Moreover, these bacterially expressed proteins appear to assemble on the RNA during the second step of splicing (Zhang et al., 2007). The differences observed between the use of bacterial proteins and recombinant proteins expressed in HEK293 cells suggest that posttranslational modifications of EJC proteins in mammalian cells may be required for the association of EJC protein(s) with the spliceosome.

During its assembly and the different steps of splicing, the spliceosome undergoes a number of conformational and structural changes (Bessonov et al., 2008; Makarov et al., 2002). An open question that remains to be answered is how the suggested conformational change of eIF4A3 is achieved. The conformational change of eIF4A3 may be a consequence of the rearrangements that occur within the spliceosome when splicing proceeds. eIF4A3 is found in the activated spliceosome, so it may adopt the closed, RNA-bound conformation during the transition from the activated spliceosome to splicing complex C. The results of Gehring et al., 2009a suggest that a significant fraction of eIF4A3 adopts the MAGOH-Y14-bound state before exon ligation. It is thus likely that upon binding of ATP and MAGOH-Y14, eIF4A3 adopts a closed confirmation and is stably locked onto the RNA in the C-complex.

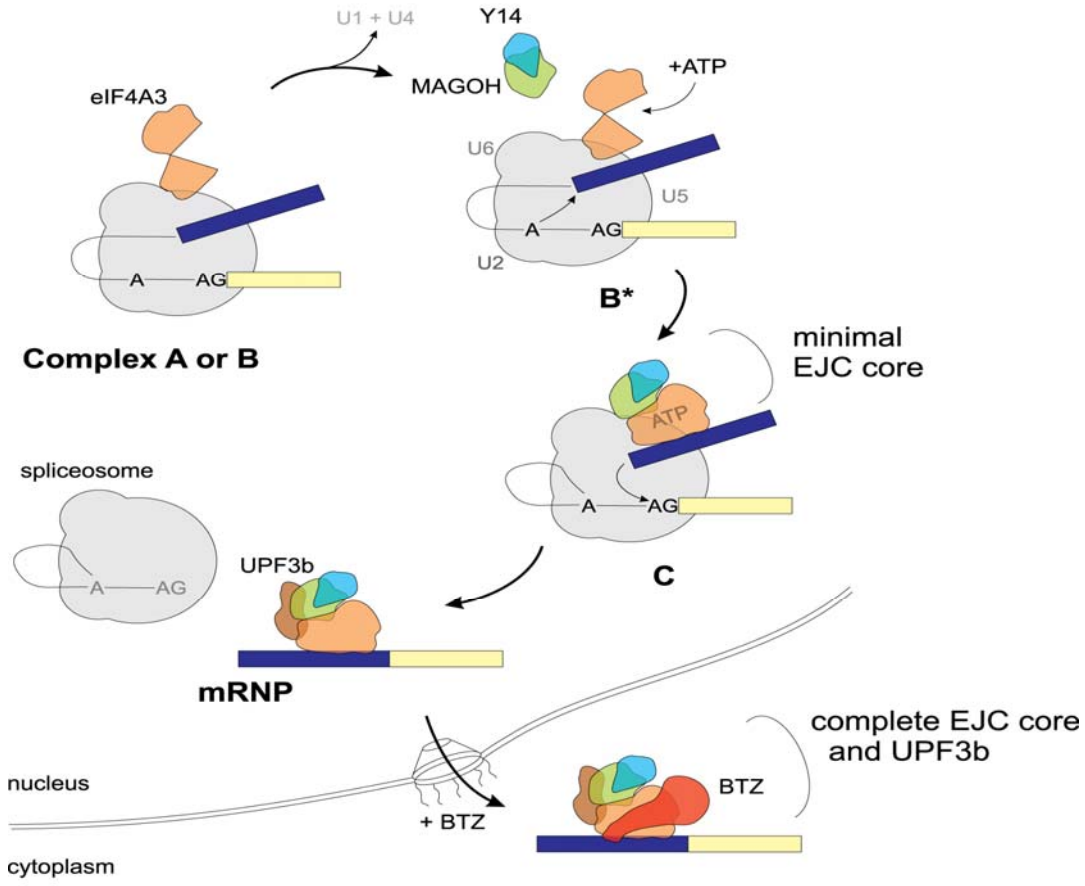


Figure 42. The proposed model of the EJC assembly pathway by the spliceosome.

4.3.2 From the trimeric pre-EJC to NMD-competent core EJC

One of the main interests in our laboratory is to elucidate the mechanism of NMD. The results presented in this part of the thesis were used as a basis for studying the functional implications of this assembly into the NMD activation (Gehring et al., 2009a). As mentioned already, the EJC plays a central role in the activation of NMD in mammalian cells. Apart from the core EJC members, our investigation was extended to more peripherally associated EJC members that play a role in NMD (Gehring et al., 2009a). Some of these EJC members, BTZ and the NMD factor UPF3b, bind to the

trimeric pre-EJC after the mRNA's release from the spliceosome. BTZ is predominantly localized in the cytoplasm (Degot et al., 2004; Degot et al., 2002). Taking into consideration its localization, as well as our data that BTZ joins the trimeric pre-EJC after splicing (Figure 26), this interaction may occur after nuclear export of the mRNP. Therefore BTZ is proposed to join the EJC at a later (possibly cytosolic) step of assembly. Gehring et al. 2009a have shown that the interaction between the trimeric pre-EJC and BTZ depends on exposed surface residues of eIF4A3 including aa 178/179. These residues also stabilize the association of eIF4A3 with spliceosomal complexes prior to the second step of splicing. This overlap indicates that BTZ and the spliceosome interact with the same region of eIF4A3. Such a finding can explain why nuclear BTZ fails to enter spliceosomal complexes and why BTZ has not been found in purified spliceosomal complexes (Bessonov et al., 2008).

4.4 Disassembly of Exon Junction Complexes by PYM

It has been proposed that the EJC is removed by ribosomes as they scan the mRNA during initiation and elongation (Dostie et al., 2002). PYM is identified as an active, ribosome-bound EJC disassembly factor in mammalian cell, using the *in vitro* splicing system. This work reveals that PYM is an EJC disassembly factor *in vitro* (Figure 27), antagonizing important EJC functions.

4.4.1 PYM-mediated disassembly of mature but not maturing EJCs

PYM has been crystallized as part of a heterotrimeric complex with the EJC proteins MAGOH-Y14 (Bono et al., 2004). Tethered PYM was reported to mediate NMD (Bono et al., 2004), suggesting that PYM associates with the EJC to support its functions. However, we noticed that the binding of EJCs to spliced mRNAs is augmented with mutants of MAGOH that cannot bind PYM. This observation suggested that PYM may

antagonize EJC assembly and/or stability, and initiated our further biochemical analysis.

Remarkably, PYM does not interfere with the assembly of MAGOH-Y14 into EJC complexes by the spliceosome, but specifically disrupts mature EJCs after release of the spliceosome from the mRNA (Figure 30). This activity is mediated by direct protein-protein interactions between MAGOH and the N-terminus of PYM (Gehring et al., 2009b). The proteins eIF4A3 and PYM bind the same residues of MAGOH-Y14. Therefore stable binding of PYM to MAGOH-Y14 would not be possible within the EJC (Bono et al., 2006), suggesting that stable PYM-EJC assemblies should not exist. This prediction was confirmed *in vitro* and *in vivo* (Gehring et al., 2009b). This observation is in agreement with the results of Chamieh et al., 2008; the MAGOH-Y14 heterodimer appears to form mutually exclusive, stable complexes with either PYM or within the EJC.

These findings raise the questions of why EJC assembly intermediates including MAGOH-Y14 are resistant to PYM, and how PYM initiates the disassembly of mature EJCs. The insensitivity of maturing EJC assembly intermediates to PYM may have different explanations. First, spliceosomal proteins may protect MAGOH-Y14 from PYM during EJC assembly. This would involve interactions of MAGOH-Y14 with the spliceosome during the deposition of the EJC on the RNA. Second, after the release of the spliceosome MAGOH-Y14 heterodimers may undergo conformational changes that enable their interaction with PYM, or the PYM interaction may be helped by an EJC protein that is still lacking from the minimal EJC core after the first step of splicing.

Based on available structural and biochemical data (Andersen et al., 2006; Bono et al., 2006; Bono et al., 2004; Chamieh et al., 2008; Gehring et al., 2005), we suggest that the interaction of PYM with solvent-accessible side chains of MAGOH-Y14 within the mature EJC leads to the formation of a transient intermediate that destabilizes MAGOH-Y14 from the EJC. As a consequence, the inhibition of the ATPase activity of eIF4A3 imposed by MAGOH-Y14 is removed, and ATP hydrolysis can occur since the ATPase active site of eIF4A3 is already positioned for catalysis (Andersen et al., 2006; Ballut et al., 2005; Bono et al., 2006). Once the γ-phosphate of ATP is released, it induces the dissociation of eIF4A3 from the mRNA (Andersen et al., 2006; Ballut et al.,

2005; Bono et al., 2006). Moreover, eIF4A3 adopts an open conformation after ATP hydrolysis, preventing the re-binding of MAGOH-Y14 (Andersen et al., 2006; Bono et al., 2006), which is additionally avoided by the formation of trimeric PYM-MAGOH-Y14 complexes. It will be interesting to investigate whether PYM alone is sufficient to disassemble EJCs that have been assembled from recombinant subunits, or whether PYM requires co-factors that are present in splicing extracts. Similarly, it will be informative to explore whether cytoplasmic proteins exist that can protect the EJC from disassembly by PYM.

4.4.2 Functional implications of EJC disassembly by PYM for EJC recycling and NMD

Translation and EJC removal have previously been proposed to be coupled (Dostie et al., 2002; Lejeune et al., 2002), but the mechanism of EJC removal has remained enigmatic and been attributed to the ribosomal act of translation elongation (Dostie et al., 2002). The results presented in section 3.4 and in Gehring et al, 2009b establish PYM as a ribosome-bound EJC disassembly factor. Perceiving a mammalian mRNA that bears premature termination codon requires the translation-dependent assessment of the position of the stop codon relative to the last EJC. If a messenger mRNA retains an EJC downstream from the stop codon, it is considered as “faulty” and it is degraded by NMD (Thermann et al., 1998; Zhang et al., 1998). Hence, coupling of EJC disassembly to mRNA translation seems logical, as it ensures (1) that EJCs are only removed from mRNAs that have been subjected to at least one round of translation, and (2) that EJCs are only removed from positions within the open reading frame but not from positions sufficiently far downstream of (premature) translation termination codons.

These two criteria would be met if the ribosome itself was the EJC disassembly machine, but not if EJC disassembly is catalyzed by a soluble cytoplasmic protein that could remove EJCs irrespective of mRNA translation and EJC position within the mRNA. We propose that the tight binding of PYM to the ribosome and the lack of “free” cytosolic PYM serve to couple EJC removal by PYM to translation, and hence to

ensure proper NMD function. Indeed, accumulation of “free” cytosolic PYM inhibits NMD of classical NMD substrates and in tethering assays (Gehring et al., 2009b).

The number of PYM molecules and ribosomes per cell were estimated. It was found that HeLa cells express far fewer molecules of PYM than the number of ribosomes per cell (Gehring et al., 2009b). It is interesting to discover how PYM associates preferentially with the ribosomes that perform (the first round of) translation, or how PYM-bound ribosomes are preferentially selected for translation. The way PYM binds to the ribosome is unknown. It may interact indirectly through the association with translation (initiation) factors. If this hypothesis holds true, PYM would be recruited only to ribosomes that are translationally active. Hence, explaining why overexpression of PYM leads to the accumulation of “free” cytosolic PYM despite the high abundance of ribosomes. Recently, PYM has been suggested to stimulate translation through EJCs (Diem et al., 2007); such a mechanism could help assure that the translation of EJC-containing mRNAs is performed by PYM-bound ribosomes.

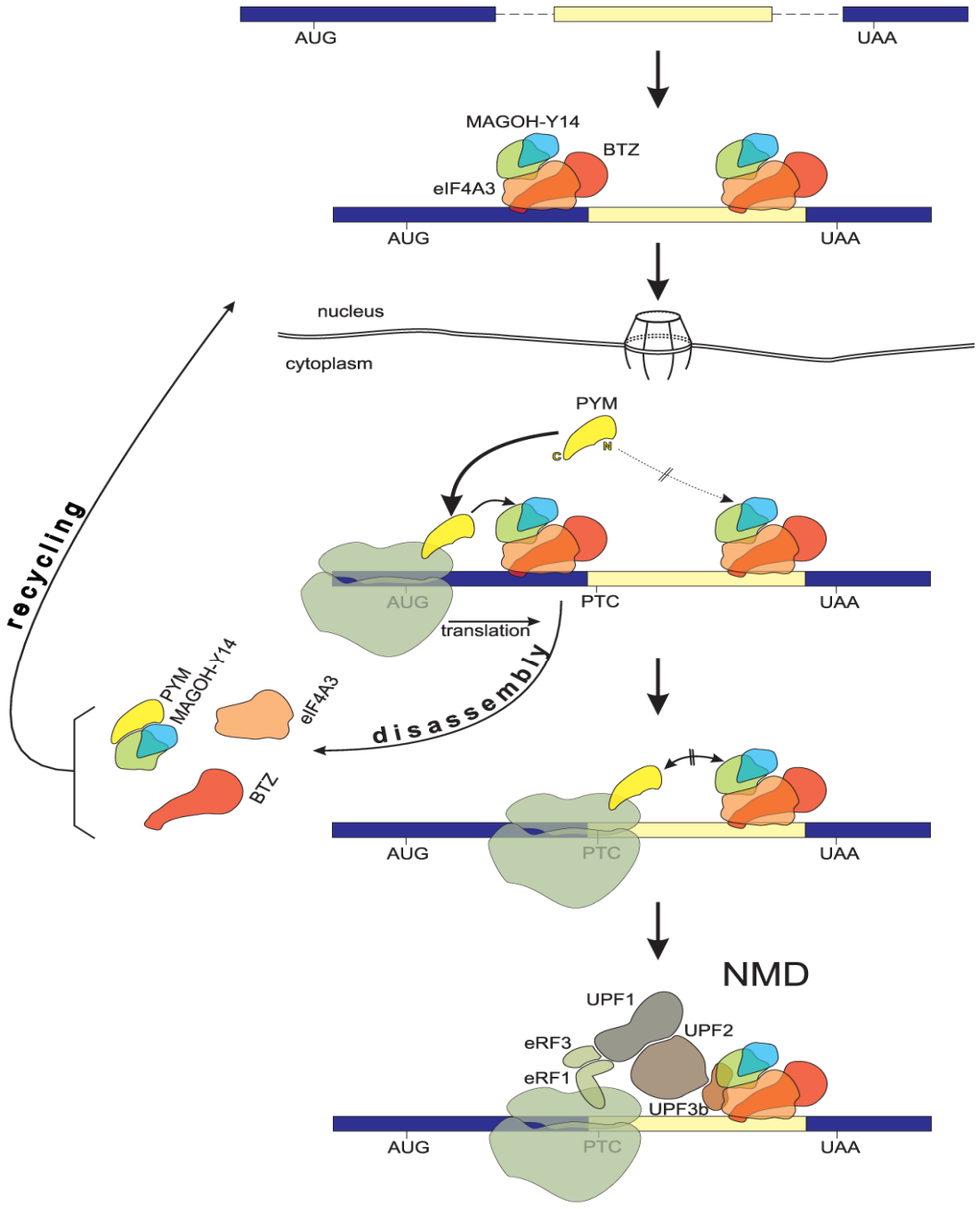


Figure 43. PYM is a ribosome-bound EJC disassembly factor. Exon junction complexes are assembled during splicing in the nucleus. The core components of EJC, MAGOH-Y14, eIF4A3, BTZ are depicted in coloured shapes. BTZ most probably joins the complex at a later step in the cytoplasm. PYM (shown in yellow) interacts via its C-terminus (indicated by C) with the small ribosomal subunit and removes EJC during ribosomal passage via interaction of its N-terminus (N) with MAGOH-Y14. Cytosolic «free» PYM is limited by its ribosomal association, because it may disassemble EJCs independently of translation and thus inhibit

more downstream processes, such as NMD. Disassembled EJC components are recycled and transported back to the nucleus, with the likely exception of BTZ that may remain in the cytoplasm. When a premature translation termination codon (PTC) is encountered and a downstream EJC is recognized, NMD is activated. During the activation of NMD, UPF-proteins bridge the EJC to the terminating ribosome. Adapted from (Gehring et al., 2009b).

An important question is whether the number of the EJC components is enough to cover all the exon-exon junctions in a cell. Trying to address this issue, the number of exon-exon junctions within the steady-state transcriptome of HeLa cells was calculated to approximately 400,000 (100,000 transcripts with a mean of 4 introns; (Deutsch et al., 1999; Jackson et al., 1998). This number far exceeds the abundance of EJC subunits per cell [10,000 for eIF4A3; 40,000 for Y14 and MAGOH, respectively (Gehring et al., 2009b)]. This suggests that newly exported mRNAs bearing EJCs must be preferentially translated, and after they are translated, their EJC components should be efficiently disassembled and recycled. PYM may be the key factor that serves both needs, helping to direct translation preferentially to EJC-bound mRNAs (Diem et al., 2007; Le Hir et al., 2003; Lu et al., 2003; Nott et al., 2004; Nott et al., 2003; Wiegand et al., 2003), and mediating efficient EJC disassembly and recycling. Future studies will address whether PYM facilitates ribosomal EJC disassembly or is required for this process. Future studies will also analyze the pathway(s) along which the EJC subunits recycle into the nucleus.

4.5 Deposition of the mammalian TREX complex

4.5.1 TREX deposition is splicing-independent

A number of models has been proposed for the deposition of the TREX complex. Some of these models suggest the involvement of EJC. Our recent discovery, that PYM specifically and reproducibly disassembles mature EJCs, could serve as a tool to assess the direct role of EJC into the mammalian TREX complex deposition. The initial goal was to recapitulate results in the literature which have been obtained by different experimental approaches. Therefore our *in vitro* splicing system, described earlier in 4.2, was adapted to investigate the deposition of the mammalian TREX complex. Splicing reactions were performed in a combination of HeLa nuclear extracts and 293 whole cell extracts prepared from FLAG-tagged TREX complex components. We initiated this study by investigating the role of splicing, EJC and cap to the mammalian TREX complex recruitment. This study focused on the TREX complex components UAP56 and ALY/REF. The investigation included protein DDX39, which is highly similar to UAP56 protein. The deposition of DDX39 has not been studied before in similar assays. The main goal was to specifically address the contribution of the EJC for the first time, using the addition of rPYM. These results will add valuable information and could lead to a more complete picture of the deposition of this complex.

From the proteins tested in the *in vitro* system, UAP56, DDX39 and ALY/REF precipitated the spliced mRNAs and could be used for further studies. One of the basic questions in the field is whether the deposition of these factors is splicing-dependent. To test this question, intron-containing and intronless transcripts were used. In our experimental system, UAP56, DDX39 and ALY/REF are recruited both to intron-containing and intronless transcripts. This is in agreement with a previous study, which has shown that proteins UAP56 and ALY/REF are recruited to an intronless transcript, but less efficiently than to spliced mRNA (Masuda et al., 2005).

In order to investigate the role of EJC to TREX deposition, two approaches were followed. The first approach employed the use of a transcript with a short first

exon (of only 15 nts) that does not support the assembly of EJC and the second approach employed the newly found tool, the addition of rPYM to disassemble EJCs. Both approaches showed that UAP56 and DDX39 precipitate both, the spliced and unspliced mRNAs (intron-containing and intronless transcript), while the precipitation observed in the case of ALY/REF was less efficient than the precipitation of UAP56 and DDX39. When rPYM was added, the amount of RNA precipitated was quantified. UAP56 and DDX39 did not show any significant change in the amount of mRNA they precipitated. On the other hand, ALY/REF showed a significant change of the amount of mRNA precipitated, which was dose-dependent. When more rPYM was added, less mRNA was precipitated.

Previous studies have shown that UAP56 localizes to nuclear speckles (Gatfield et al., 2001) and that it associates with spliced mRNAs (Cheng et al., 2006; Gatfield et al., 2001; Masuda et al., 2005). Concerning ALY/REF, it was found to be present in spliced mRNP complexes, but absent from an analogous mRNP complex containing unspliced mRNA (Zhou et al., 2000) and spliced mRNPs-not carrying the EJC (Le Hir et al., 2000). In addition to these observations, ALY/REF was found to be among the mRNA-specific proteins associated with the 8-nucleotide-protected RNA fragment due to the presence of EJC (Le Hir et al., 2000). Our results suggest that ALY/REF is not a member of the core EJC components, but its deposition is affected by the EJC. If ALY/REF was a core EJC component, it would exhibit a similar behavior to eIF4A3 and MAGOH-Y14, leading to its complete dissociation from the RNA once rPYM was added.

ALY/REF does not seem to require the EJC for its interaction with the mRNA. It is conceivable, however, that an EJC is required for the stable association of ALY/REF and/ or for the proper conformational changes necessary for the ALY/REF to function as an export adaptor in more downstream processes, such as mRNA export.

It has been suggested that the human TREX complex is recruited in a splicing- and cap-dependent manner to a region near the 5'end of the mRNA (Cheng et al., 2006). Particularly, it was proposed that ALY/REF binds to the mRNA closest to the cap and THO complex, while UAP56 binds downstream of ALY/REF but upstream of the EJC (Cheng et al., 2006; Masuda et al., 2005). In this model, the recruitment of the

TREX complex to mRNA occurs via an interaction between the TREX component ALY/REF and the cap-binding protein CBP80. While the TREX complex is bound only on the first exon, the EJC is recruited at every exon-exon junction independently of the 5' cap/CBP80 and does not contain the mRNA export machinery (Cheng et al., 2006; Masuda et al., 2005). Having this information as a basis, we wanted to address the role of the cap-containing-5'end. The experimental approach included *in vitro* splicing followed by RNase H digestion assays. UAP56 and DDX39 precipitate strongly all three products (spliced product and the two RNase H digestion products: 5'and 3'fragment), while ALY/REF precipitated the spliced product efficiently but the 5'and 3'fragments less efficiently. A similar experiment was performed with an intronless transcript. UAP56 and DDX39 precipitated all three products, while ALY/REF precipitated all three products much less efficiently. The observation that ALY/REF precipitated the spliced mRNA efficiently, but the 5'and 3'fragments less efficiently could suggest that for the efficient and stable association of ALY/REF with the mRNA both the cap and the EJC are required. The observation is in accordance with previously published results, that TREX is recruited in a cap-dependent manner (Cheng et al., 2006).

4.5.2 Functional characterization of ALY/REF and UAP56

It has been shown that ALY/REF interacts with the UAP56-C terminus as efficiently as with full-length UAP56, indicating that the C terminus is sufficient for ALY/REF binding (Masuda et al., 2005). The C terminus of ALY/REF is required not only for binding UAP56, but also for the hTHO complex (Masuda et al., 2005). Trying to acquire valuable tools for the future analysis of the mechanisms of mRNA export, several UAP56 and ALY/REF mutants were generated. These analyses were carried out aiming at the identification of mutants that show reproducible changes in binding of the 5'or the 3'fragments. From the analyzed mutants, the UAP56K95N mutant with a single amino acid change at the ATP binding site precipitates very weakly the RNase

H-generated fragments. UAP56 mutants, having a N- or C-terminal deletion, did not precipitate unspliced pre-mRNA or spliced mRNA fragments. This observation hints that large deletions inhibit the protein's function. This inhibition can derive from the fact that both the N- and C- termini are required in order for the protein to acquire its functional conformation.

A yeast two-hybrid study has shown that the most frequent interaction partner for human UAP56 (103 of the total 117 positive clones) was either itself or a close paralogue (DDX39), indicating that it exists as a homo- or heterodimer (Lehner et al., 2004). The crystal structure of human UAP56 has shown that it is composed of two domains connected by a flexible linker (Zhao et al., 2004). Moreover, the crystal structure of the N-terminal domain reveals a dimer interface that could potentially be important for its function (Zhao et al., 2004). From the proposed crystal structure, two possibilities derive. One possibility is that a dramatic conformational change is induced by the binding of ATP and/or substrate. This change brings the two domains closer together. An alternative explanation for the different domain organization is that UAP56, and similar DExD/H proteins, could use the different domains to fulfill distinct functions. The crystal structure observations support the results of our mutational analysis that both termini are necessary for the UAP56 protein to function properly. Further mutational, biochemical, and functional analyses are required to investigate the roles of UAP56 and DDX39 and assess the role of the dimer formation in their function in splicing and export. The ATPase and helicase activities of DDX39 have yet to be demonstrated, together with the helicase activity of UAP56.

Testing the ability of ALY/REF mutants to precipitate spliced mRNAs and RNase H-digested fragments, it was observed that the deletion of 16 C- terminal amino acids did not affect the binding to the spliced products. In our study, the ALY/REF mutant ΔC16 precipitated the RNase H-generated cap fragment as efficiently as the wild type protein. This observation is consistent with a previous study (Cheng et al., 2006). The ΔC16 ALY/REF mutant precipitated the cap fragment, therefore still binds to the cap and the cap-binding proteins like CBP80. Deletion of 80 N-terminal residues leads to no RNA precipitation at all. The ALY/REF mutant, lacking 53 C-terminal residues, shows a reduced binding of the 5'fragment, while the binding to the 3'end is

similar to the wild type. Taken the results from the two ALY/REF mutants (ΔC16 and ΔC53) together, it is evident that protein residues 204-241 play a role in the protein binding to the cap. Further mutational studies are required for the identification of the protein regions involved in the different protein interactions.

The loading of the TREX complex seems to be directed neither solely by the cap nor by the EJC, but the cap seems to have a major influence. Detailed studies addressing the role of the cap are required to shed light on its contribution to the deposition of this complex. The role of the cap can be addressed, either by cap competition experiments or by the use of transcripts having a non-functional cap.

As next steps, the validation of interactions between TREX members and the investigation of their export function is required. I would propose in order to clarify the export function to use *Xenopus laevis* oocytes system. Microinjection export assays in *Xenopus* oocyte, as described before (Hamm et al., 1990), can be performed using capped and uncapped transcripts, in previously depleted oocytes of certain export-related factor, such as ALY/REF or UAP56. Depletion factors involved will help assess their contribution to mRNA export.

4.5.3 Future directions and next steps to be taken

The results presented in this thesis indicate that splicing is not necessary for the recruitment of ALY/REF, UAP56 and DDX39. Plus they do not provide evidence for the connection of UAP56 to EJC. Nonetheless, ALY/REF could be the factor that connects the cap and the EJC, since in our study it precipitates both the cap and EJC fragment.

Current models propose that UAP56 is loaded on the mRNA first and then it recruits ALY/REF (Luo et al., 2001). Consecutively, ALY/REF interacts with the cap binding protein-CBP80. It is also possible that ALY/REF is recruited by CBP80 (Masuda et al., 2005), but this interaction is not stable and it needs UAP56 to stabilize it. The ATP-dependent deposition of ALY/REF requires the hydrolysis of ATP.

Our model does not suggest that unspliced mRNAs are trapped into the nucleus, because they recruit the proteins necessary for export. It can be that export is not necessarily splicing-dependent, maybe splicing enhances export. Moreover, it is possible that the protein interaction with export receptors is splicing dependent. UAP56 is required for the recruitment of ALY/REF to mRNA, and it is displaced from ALY/REF by the mRNA export factor TAP (TAP: p15) as the message is getting ready to be exported (Kataoka et al., 2001). It has been shown that TAP:p15 interacts with Y14 and MAGOH (Kataoka et al., 2001). This observation suggests a possible role for the EJC in recruitment of TAP: p15.

Nojima and his colleagues suggest that ALY/REF can be deposited to an intronless mRNA through the cap (Nojima et al., 2007). In co-immunoprecipitation experiments in HEK cells, ALY/REF interacts with CBP20, and it is hypothesized that the recruitment of ALY/REF may play a stimulatory role to export the capped intronless mRNAs (Nojima et al., 2007).

When a transcript is transcribed, the cap structure and the poly (A) tail are added to it, to ensure the stability of the RNA molecule. In order for a transcript to be capped and polyadenylated, a number of proof-reading mechanisms occur ensuring that the transcript has already passed the first control check point. A crucial decision in the process of gene expression is whether a transcript should be exported from the nucleus to the cytoplasm to be translated or not. At this point the cell should certify that only completely and correctly processed molecules are available for more downstream processes. Therefore some of the export factors, in this case ALY/REF, interact with the cap and a recently suggested possible link with the 3'end processing-Pcf11 to check the polyadenylation status of the transcript.

It was recently shown that Yra 1 (the yeast homologue of ALY/REF) is recruited to mRNA co-transcriptionally by Pcf11 (Johnson et al., 2009). Johnson and colleagues show that Pcf11 and Sub2 have overlapping binding sites, suggesting that they bind in a mutually exclusive way (Johnson et al., 2009). The functional importance of Pcf11 is shown in yeast, it still remains to be determined if Pcf11 plays a role in mammals. These new data suggest that the recruitment of TREX components requires both the cap (5'end) and the 3'end. If this model holds true, factors necessary for export

are recruited to the RNA, only if it is correctly processed (capped and polyadenylated). In this way, the cell could verify the integrity of the two ends. Only if both ends are properly processed, the message is ready to be exported to the cytoplasm. For this process, namely the final decision to export or not a message, splicing is not of a primary importance. Hence, naturally occurring messages that have a cap and poly (A) tail can be physiologically exported to the cytoplasm. But splicing could play a role in the interaction of the export adaptor proteins with the export receptors. It is possible that splicing is required for the recruitment of the export adaptors, thus enhancing the export of the spliced products.

One of the most appealing challenges is to understand how mRNA processing steps (capping, splicing and polyadenylation) can influence the downstream processes, starting from mRNA export. Future studies are required to find which proteins are recruited to the mRNPs and dictate their fates.

5. REFERENCES

Abruzzi, K.C., Lacadie, S., and Rosbash, M. (2004). Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *The EMBO journal* ***23***, 2620-2631.

Alcazar-Roman, A.R., Tran, E.J., Guo, S., and Wente, S.R. (2006). Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. *Nat Cell Biol* ***8***, 711-716.

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 (New York, NY* ***313***, 1968-1972.

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.

Bell, M., Schreiner, S., Damianov, A., Reddy, R., and Bindereif, A. (2002). p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *The EMBO journal* ***21***, 2724-2735.

Bennett, M., Pinol-Roma, S., Staknis, D., Dreyfuss, G., and Reed, R. (1992). Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Molecular and cellular biology* ***12***, 3165-3175.

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 reports* ***5***, 304-310.

Braun, I.C., Herold, A., Rode, M., Conti, E., and Izaurralde, E. (2001). Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export. *J Biol Chem* ***276***, 20536-20543.

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, C.C., Dostie, J., Diem, M.D., Feng, W., Mann, M., Rappsilber, J., and Dreyfuss, G. (2004). eIF4A3 is a novel component of the exon junction complex. *RNA* ***10***, 200-209.

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.

Chiu, S.Y., Lejeune, F., Ranganathan, A.C., and Maquat, L.E. (2004). The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. *Genes & development* ***18***, 745-754.

Cole, C.N., and Scarcelli, J.J. (2006a). Transport of messenger RNA from the nucleus to the cytoplasm. *Curr Opin Cell Biol* ***18***, 299-306.

Cole, C.N., and Scarcelli, J.J. (2006b). Unravelling mRNA export. *Nat Cell Biol* ***8***, 645-647.

Conti, E., and Izaurralde, E. (2001). Nucleocytoplasmic transport enters the atomic age. *Current opinion in cell biology* ***13***, 310-319.

Cordin, O., Tanner, N.K., Doere, M., Linder, P., and Banroques, J. (2004). The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *The EMBO journal* ***23***, 2478-2487.

Custodio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B.J., and Carmo-Fonseca, M. (2004). In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *RNA* ***10***, 622-633.

Czaplinski, K., Kocher, T., Schelder, M., Segref, A., Wilm, M., and Mattaj, I.W. (2005). Identification of 40LoVe, a Xenopus hnRNP D family protein involved in localizing a TGF-beta-related mRNA during oogenesis. *Dev Cell* ***8***, 505-515.

Daneholt, B. (1997). A look at messenger RNP moving through the nuclear pore. *Cell* ***88***, 585-588.

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.

Deutsch, M., and Long, M. (1999). Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res* ***27***, 3219-3228.

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. *Nature structural & molecular biology* ***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.

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.

Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J., and Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *The EMBO journal* ***21***, 5843-5852.

Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes & development* ***11***, 1864-1872.

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.

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.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. *The EMBO journal* ***22***, 3960-3970.

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.

Guzik, B.W., Levesque, L., Prasad, S., Bor, Y.C., Black, B.E., Paschal, B.M., Rekosh, D., and Hammarskjold, M.L. (2001). NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of intron-containing RNA in mammalian cells. *Mol Cell Biol* ***21***, 2545-2554.

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.

Hamm, J., and Mattaj, I.W. (1990). Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* ***63***, 109-118.

Hentze, M.W., and Kulozik, A.E. (1999). A perfect message: RNA surveillance and nonsense-mediated decay. *Cell* ***96***, 307-310.

Hilleren, P., and Parker, R. (1999). Mechanisms of mRNA surveillance in eukaryotes. *Annual review of genetics* ***33***, 229-260.

Hosoda, N., Kim, Y.K., Lejeune, F., and Maquat, L.E. (2005). CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. *Nature structural & molecular biology* ***12***, 893-901.

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

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell* ***12***, 711-721.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* ***106***, 607-617.

Jackson, D.A., Iborra, F.J., Manders, E.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* ***9***, 1523-1536.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology* ***3***, 318-356.

Jacobson, A., and Peltz, S.W. (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annual review of biochemistry* ***65***, 693-739.

Jensen, T.H., Boulay, J., Rosbash, M., and Libri, D. (2001). The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr Biol* ***11***, 1711-1715.

Jimeno, S., Rondon, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *The EMBO journal* ***21***, 3526-3535.

Johnson, S.A., Cubberley, G., and Bentley, D.L. (2009). Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. *Molecular cell* ***33***, 215-226.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* ***12***, 5-14.

Katahira, J., Strasser, K., Podtelejnikov, A., Mann, M., Jung, J.U., and Hurt, E. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *The EMBO journal* ***18***, 2593-2609.

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. *The EMBO journal* ***20***, 6424-6433.

Kataoka, N., and Dreyfuss, G. (2004). A simple whole cell lysate system for in vitro splicing reveals a stepwise assembly of the exon-exon junction complex. *J Biol Chem* ***279***, 7009-7013.

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. *Molecular cell* ***6***, 673-682.

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

Kim, V.N., Yong, J., Kataoka, N., Abel, L., Diem, M.D., and Dreyfuss, G. (2001b). The Y14 protein communicates to the cytoplasm the position of exon-exon junctions. *The EMBO journal* ***20***, 2062-2068.

Kole, R., and Weissman, S.M. (1982). Accurate in vitro splicing of human beta-globin RNA. *Nucleic Acids Res* ***10***, 5429-5445.

Konarska, M.M., and Sharp, P.A. (1986). Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* ***46***, 845-855.

Kota, K.P., Wagner, S.R., Huerta, E., Underwood, J.M., and Nickerson, J.A. (2008). Binding of ATP to UAP56 is necessary for mRNA export. *J Cell Sci* ***121***, 1526-1537.

Krainer, A.R., Maniatis, T., Ruskin, B., and Green, M.R. (1984). Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* ***36***, 993-1005.

Le Hir, H., and Andersen, G.R. (2008a). Structural insights into the exon junction complex. *Curr Opin Struct Biol* ***18***, 112-119.

Le Hir, H., Gatfield, D., Braun, I.C., Forler, D., and Izaurralde, E. (2001a). The protein Mago provides a link between splicing and mRNA localization. *EMBO reports* ***2***, 1119-1124.

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

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

Le Hir, H., Nott, A., and Moore, M.J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* ***28***, 215-220.

Le Hir, H., and Seraphin, B. (2008b). EJCs at the heart of translational control. *Cell* ***133***, 213-216.

Lehner, B., Semple, J.I., Brown, S.E., Counsell, D., Campbell, R.D., and Sanderson, C.M. (2004). Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* ***83***, 153-167.

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.

Lejeune, F., Ranganathan, A.C., and Maquat, L.E. (2004). eIF4G is required for the pioneer round of translation in mammalian cells. *Nature structural & molecular biology* ***11***, 992-1000.

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

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.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science (New York, NY* ***298***, 2205-2208.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *Embo J* ***23***, 2381-2391.

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 & development* ***19***, 1512-1517.

Mattaj, I.W., and Englmeier, L. (1998). Nucleocytoplasmic transport: the soluble phase. *Annu Rev Biochem* ***67***, 265-306.

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

Mehlin, H., Daneholt, B., and Skoglund, U. (1992). Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. *Cell* ***69***, 605-613.

Merz, C., Urlaub, H., Will, C.L., and Luhrmann, R. (2007). Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. *RNA* ***13***, 116-128.

Mohr, S.E., Dillon, S.T., and Boswell, R.E. (2001). The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during Drosophila oogenesis. *Genes & development* ***15***, 2886-2899.

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

Neu-Yilik, G., Gehring, N.H., Thermann, R., Frede, U., Hentze, M.W., and Kulozik, A.E. (2001). Splicing and 3' end formation in the definition of nonsense-mediated decay-competent human beta-globin mRNPs. *The EMBO journal* ***20***, 532-540.

Nojima, T., Hirose, T., Kimura, H., and Hagiwara, M. (2007). The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *J Biol Chem* ***282***, 15645-15651.

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 (New York, NY* ***9***, 607-617.

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.

Pryor, A., Tung, L., Yang, Z., Kapadia, F., Chang, T.H., and Johnson, L.F. (2004). Growth-regulated expression and G0-specific turnover of the mRNA that encodes URH49, a mammalian DExH/D box protein that is highly related to the mRNA export protein UAP56. *Nucleic Acids Res* ***32***, 1857-1865.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol* ***17***, 269-273.

Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell* ***41***, 95-105.

Reichert, V.L., Le Hir, H., Jurica, M.S., and Moore, M.J. (2002). 5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes & development* ***16***, 2778-2791.

Rondon, A.G., Garcia-Rubio, M., Gonzalez-Barrera, S., and Aguilera, A. (2003). Molecular evidence for a positive role of Spt4 in transcription elongation. *The EMBO journal* ***22***, 612-620.

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

Shen, H., Zheng, X., Shen, J., Zhang, L., Zhao, R., and Green, M.R. (2008). Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes & development* ***22***, 1796-1803.

Shen, J., Zhang, L., and Zhao, R. (2007). Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. *J Biol Chem* ***282***, 22544-22550.

Shibuya, T., Tange, T.O., Sonenberg, N., and Moore, M.J. (2004). eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nature structural & molecular biology* ***11***, 346-351.

Shibuya, T., Tange, T.O., Stroupe, M.E., and Moore, M.J. (2006). Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. *RNA (New York, NY* ***12***, 360-374.

Smith, C.W., Patton, J.G., and Nadal-Ginard, B. (1989). Alternative splicing in the control of gene expression. *Annual review of genetics* ***23***, 527-577.

Sreedhar, A.S., Kalmar, E., Csermely, P., and Shen, Y.F. (2004). Hsp90 isoforms: functions, expression and clinical importance. *FEBS Lett* ***562***, 11-15.

Strasser, K., and Hurt, E. (2000). Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *The EMBO journal* ***19***, 410-420.

Strasser, K., and Hurt, E. (2001). Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* ***413***, 648-652.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., *et al.* (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* ***417***, 304-308.

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.

Tange, T.O., Nott, A., and Moore, M.J. (2004). The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol* ***16***, 279-284.

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.

Tarn, W.Y., and Steitz, J.A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes & development* ***8***, 2704-2717.

Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeier, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* ***17***, 3484-3494.

Visa, N., Alzhanova-Ericsson, A.T., Sun, X., Kiseleva, E., Bjorkroth, B., Wurtz, T., and Daneholt, B. (1996a). A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. *Cell* ***84***, 253-264.

Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B., and Mattaj, I.W. (1996b). A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *J Cell Biol* ***133***, 5-14.

Weirich, C.S., Erzberger, J.P., Flick, J.S., Berger, J.M., Thorner, J., and Weis, K. (2006). Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. *Nat Cell Biol* ***8***, 668-676.

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.

Wilusz, C.J., and Wilusz, J. (2004). Bringing the role of mRNA decay in the control of gene expression into focus. *Trends Genet* ***20***, 491-497.

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E. (1998). 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, Z., and Krainer, A.R. (2007). Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. *Proc Natl Acad Sci U S A* ***104***, 11574-11579.

Zhao, R., Shen, J., Green, M.R., MacMorris, M., and Blumenthal, T. (2004). Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export. *Structure* ***12***, 1373-1381.

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.

6. LIST OF ABBREVIATIONS

A adenosine, Ampere

A_{260} absorbance at 260 nm

aa amino acid

amp ampicillin

APS ammonium persulfate

ATP adenosine triphosphate

BPS branch point sequence

BrEt ethidium bromide

BSA bovine serum albumin

C. elegans Caenorhabditis elegans

cDNA complementary DNA

Ci Curie

cpm counts per minute

C-terminus carboxy-terminus

CTP cytidine triphosphate

dATP deoxyadenosine triphosphate

dCTP deoxycytidine triphosphate

dGTP deoxyguanosine triphosphate

DMEM Dulbecco's modified Eagle's medium

DMSO dimethylsulfoxide

DNA deoxyribonucleic acid

dNTP deoxynucleotide triphosphate

D. melanogaster Drosophila melanogaster

DSE downstream sequence element

dsRNA double-stranded RNA

dT deoxythymidine

DTT dithiothreitol

dTTP deoxythymidine triphosphate

E. coli Escherichia coli

EDTA ethylenediaminetetraacetic acid

eIF eukaryotic translation initiation factor

EJC exon junction complex

eRF eukaryotic translation release factor

ESI-MS electrospray ionization mass spectrometry

FCS fetal calf serum

g gram, gravitational force

G guanosine

GDP guanosine diphosphate

GMP guanosine monophosphate

GST glutathione-S-transferase

GT guanylytransferase

GTP guanosine triphosphate

HEPES *4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid*

hr hour

H_2O_{bidest} double deinonized water

HRP horseradish peroxidase

H. sapiens Homo sapiens

IP_6 inositol hexaphosphate

K lysine

kb kilobases

kDa kiloDalton

Km kanamycin

L liter

LB Luria Bertani

µ micro

λN N-peptide from phage λ

M molar, methionine

m^7G 7-methyl guanosine

mol moles

MOPS 3-(N-morpholino)propanesulfonic acid

mRNA messenger RNA

mRNP messenger ribonucleoprotein particle

MS mass spectrometry

MS/MALDI mass spectrometry/matrix-assisted laser desorption/ionization

MT methyltransferase

neo neomycin

NE nuclear extract

ng nanogram

nm nanoMolar

NMD nonsense-mediated mRNA decay

NP40 Nonidet® P40

NPC nuclear pore complex

nt(s) nucleotide(s)

N-terminus amino-terminus

OD optical density

ORF open reading frame

PABP poly(A)-binding protein

PAGE polyacrylamide gel electrophoresis

PAS polyadenylation signal

PBS phosphate-buffered saline

PCR polymerase chain reaction

pre-mRNA precursor messenger RNA

PTC premature termination codon

PVA polyvinyl alcohol

rcf relative centrifugal force

RNase ribonuclease

RNA ribonucleic acid

RNAi RNA interference

RNA Pol II RNA Polymerase II

rpm rotations per minute

s seconds

S. cerevisiae Saccharomyces cerevisiae

SDS sodium dodecyl sulphate

siRNA small interfering RNA

snRNPs small nuclear ribonucleoprotein particles

SS splice site

Taq Thermophilus aquaticus

TBE tris/borate/EDTA

TBS tris-buffered saline

TEMED N,N,N',N'-tetramethylethylenediamine

Ter termination

Tm melting temperature

TREX complex trancription export complex

Tris tris(hydroxymethyl)aminoethane

U units, Uridine

UPF up-frameshift protein

uORF upstream open reading frame

UsRNPs uridine-rich small ribonuclear particles

UTR untranslated region

U units

V Volts

v/v volume/volume

wt wild type

w/v weight/volume

X. laevis Xenopus laevis

% per cent

7. APPENDIX

Parts of the thesis were published in:

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.

Parts of the thesis were presented at the following conferences:

o Thirteenth Annual Meeting of the RNA Society, Berlin, Germany, July 28- August 3 2008.

o Translational Control Meeting, Cold Spring Harbor, USA, September 3-7 2008.

8. ACKNOWLEDGEMENTS

The work presented in this thesis was carried out at the European Molecular Biology Laboratory in the research group of Prof. Dr. Andreas Kulozik and Prof. Dr. Matthias Hentze. I would like to thank both supervisors for their support, guidance and scientific advice on the projects presented in this thesis. Moreover, I would like to thank them for reading the thesis and giving their valuable feedback.

I would like to thank especially my thesis advisory committee, which consisted of Dr. Stephen Cusack, Dr. Anne Ephrussi from the EMBL and Prof. Dr. Ed Hurt from the University of Heidelberg. Their advice on the progress of the work was very much appreciated.

I would like to thank my defense committee members Dr. Stephen Cusack, Dr. Anne Ephrussi from the EMBL, Prof. Dr. Ed Hurt and Prof. Dr. Walter Nickel from the University of Heidelberg.

This work would have been impossible without the supervision and advice of Dr. Niels Gehring. I am thankful to Niels for translating the summary in German. Moreover, I would like to thank Niels for giving me the permission to use figures from our papers, which depict experiments he has performed (Results sections 3.3 and 3.4, plus figures 24, 35 and 42).

During these last four years, I was lucky to interact and work with many people at EMBL, to all of which I am indebted for their help, advice and friendship. I would like to thank particularly the EMBL Protein Expression and Purification Facility for expressing the proteins used for the production of antibodies and the Proteomic Facility for the mass spectrometry analysis.

Many thanks to all past and present members of the Kulozik and Hentze groups for sharing their valuable reagents and advice with me. Special thanks to Dr. Gaby Neu-Yilik and Dr. Sven Danckwardt for their valuable help and advice. Moreover, I would like to thank Claudia Strein and Dr. Kent Duncan from the Hentze's group for sharing their valuable reagents and expertise in GRNA chromatography.

Last but not least, I would like to thank my parents, Dimosthenis and Theckla, and my sister Victoria for their support and patience all these years. A very special "thank you" goes to Dimitris for his support and understanding during these years.