

# CWC22 Supports Pre-mRNA Splicing and Exon Junction Complex Assembly

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Anna-Lena Steckelberg

aus Frankfurt am Main



Köln, 2014



Berichterstatte: PD Dr. Niels H. Gehring

Prof. Dr. Karin Schnetz

Tag der mündlichen Prüfung: 01.07.2014



**Parts of this thesis have been published as**

**Steckelberg AL, Boehm V, Gromadzka AM, Gehring NH** 'CWC22 Connects pre-mRNA Splicing and Exon Junction Complex Assembly'; Cell Rep. 2012 Sep 27; 2(3):454-61

**Steckelberg AL, Gehring NH** 'Studying the composition of mRNPs in vitro using splicing-competent cell extracts'; Methods. 2014 Feb;65(3):342-9



## Table of contents

<b>1</b>	<b>Abstract.....</b>	<b>1</b>
1.1	Deutsche Zusammenfassung .....	1
1.2	Summary .....	3
<b>2</b>	<b>Introduction .....</b>	<b>5</b>
2.1	Regulation of eukaryotic gene expression.....	5
2.1.1	5' end capping of mRNA.....	5
2.1.2	3' end processing of mRNA.....	6
2.1.3	Pre-mRNA splicing.....	7
2.2	mRNP assembly connects different steps of gene expression .....	9
2.3	The exon junction complex is part of the splicing-dependent mRNP.....	10
2.3.1	Structure of the EJC.....	11
2.3.2	Assembly and disassembly of EJCs .....	14
2.3.3	The EJC regulates different steps of the mRNA life cycle.....	18
2.3.4	EJC-independent functions of eIF4A3.....	23
2.4	DExD/H-box helicases are key regulators of mRNA metabolism .....	24
2.4.1	Structure of DExD/H-box helicases .....	24
2.4.2	DExD/H-box proteins function as components of large RNP complexes.....	26
2.4.3	The MIF4G-domain has evolved as a helicase interaction domain .....	27
<b>3</b>	<b>Aims and Significance of the project .....</b>	<b>31</b>
<b>4</b>	<b>Results .....</b>	<b>33</b>
4.1	CWC22 is a direct interaction partner of eIF4A3.....	33
4.1.1	An <i>in silico</i> search identifies CWC22 as a putative eIF4A3 binding protein .....	33
4.1.2	CWC22 interacts with eIF4A3 in cell extracts .....	35
4.1.3	eIF4A3 binds the MIF4G domain of CWC22 .....	36
4.1.4	Pulldown experiments confirm a direct interaction of eIF4A3 and CWC22 .....	36
4.1.5	CWC22 does not interact with the other EJC proteins .....	38
4.1.6	CWC22 is no component of the post-splicing mRNP.....	40
4.1.7	CWC22, eIF4A3 and Y14 interact with the C-complex spliceosome.....	41

4.1.8	CWC22 uses multiple domains to interact with the spliceosome .....	42
4.1.9	Initial spliceosome interaction of CWC22 and eIF4A3 is independent of the EJC deposition site on spliced mRNA.....	44
<b>4.2</b>	<b>CWC22 recruits eIF4A3 to the spliceosome and mediates EJC assembly .....</b>	<b>45</b>
4.2.1	The CWC22/eIF4A3-interaction can be modeled with an existing eIF4G/eIF4A crystal structure .....	45
4.2.2	CWC22-binding deficient eIF4A3 mutants form an EJC <i>in vitro</i> .....	49
4.2.3	CWC22-binding deficient eIF4A3 mutants fail to form splicing-dependent EJCs .	50
4.2.4	Recombinant CWC22(110-409) disturbs EJC assembly <i>in vitro</i> .....	52
4.2.5	Recombinant CWC22(110-409) disturbs EJC assembly in cell extracts.....	53
4.2.6	CWC22 interaction with the spliceosome is independent of eIF4A3 .....	56
<b>4.3</b>	<b>CWC22 is required for pre-mRNA splicing in human cells.....</b>	<b>58</b>
4.3.1	siRNA knock down of CWC22 leads to an accumulation of pre-mRNA .....	58
4.3.2	MIF4G and MA3 domains of CWC22 are required for pre-mRNA splicing .....	60
4.3.3	The interaction between CWC22 and eIF4A3 is dispensable for splicing .....	62
4.3.4	Over-expression of FLAG-CWC22(NK-171/2-DE) is toxic for HEK293 cells .....	63
4.3.5	Over-expression of FLAG-CWC22(NK-171/2-DE) inhibits EJC assembly without impairing pre-mRNA splicing .....	64
<b>5</b>	<b>Discussion .....</b>	<b>67</b>
<b>5.1</b>	<b>CWC22 links pre-mRNA splicing and EJC assembly.....</b>	<b>67</b>
<b>5.2</b>	<b>The eIF4A3-CWC22 complex shows a conserved interaction pattern.....</b>	<b>67</b>
5.2.1	The CWC22/eIF4A3 structure confirms a eIF4G/eIF4A-like interaction .....	67
5.2.2	The MIF4G domain of CWC22 regulates the conformation of eIF4A3.....	69
5.2.3	The MIF4G-domain emerges as a common regulator of RNA helicases .....	71
<b>5.3</b>	<b>CWC22 mediates splicing-dependent EJC assembly .....</b>	<b>72</b>
5.3.1	CWC22 recruits eIF4A3 to the spliceosome.....	72
5.3.2	The core EJC proteins are recruited to the spliceosome via different routes .....	73
5.3.3	EJC assembly requires a rearrangement of the eIF4A3-CWC22 interaction .....	73
5.3.4	The ATPase activity of eIF4A3 is dispensable for spliceosomal recruitment .....	75
<b>5.4</b>	<b>CWC22 is required for pre-mRNA splicing in human cells.....</b>	<b>77</b>
5.4.1	MIF4G and MA3 domain are required for pre-mRNA splicing in humans .....	77
5.4.2	N- and C-terminal domains of CWC22 regulate spliceosomal recruitment .....	79
<b>5.5</b>	<b>The roles of CWC22 in pre-mRNA splicing and EJC assembly can be functionally uncoupled.....</b>	<b>80</b>

5.5.1	Pre-mRNA splicing and EJC assembly are closely linked processes .....	80
5.5.2	EIF4A3 does not influence the splicing function of CWC22 .....	80
5.5.3	The selective inhibition of EJC assembly is toxic in human cells.....	81
<b>5.6</b>	<b>Concluding remarks .....</b>	<b>82</b>
<b>6</b>	<b>Material and Methods .....</b>	<b>83</b>
<b>6.1</b>	<b>Plasmids .....</b>	<b>83</b>
6.1.1	Mammalian expression plasmids.....	83
6.1.2	Bacterial expression plasmids.....	83
6.1.3	Plasmids for in vitro transcription .....	83
<b>6.2</b>	<b>Mammalian cell culture .....</b>	<b>83</b>
6.2.1	Mammalian cell lines .....	83
6.2.2	Generation of stable cell lines.....	84
6.2.3	Transfection of plasmid DNA.....	85
6.2.4	Transfection of siRNA .....	86
6.2.5	Cell viability assay .....	86
6.2.6	Preparation of cell lysates .....	87
<b>6.3</b>	<b>Molecular biology .....</b>	<b>87</b>
6.3.1	Cloning.....	87
6.3.2	Site-directed mutagenesis.....	88
6.3.3	Plasmid preparation.....	88
6.3.4	RNA extraction from cell lysates .....	88
6.3.5	Removal of genomic DNA.....	89
6.3.6	cDNA synthesis.....	89
6.3.7	Quantitative real-time PCR .....	89
<b>6.4</b>	<b>Protein biochemistry.....</b>	<b>91</b>
6.4.1	Immunoprecipitation.....	91
6.4.2	SDS-polyacrylamide gel electrophoresis (SDS-PAGE) .....	92
6.4.3	Western blot.....	92
6.4.4	Coomassie brilliant blue staining .....	93
6.4.5	Bacterial expression of recombinant proteins.....	93
6.4.6	Affinity purification of recombinant proteins.....	94
6.4.7	<i>In vitro</i> pulldown experiments.....	95
6.4.8	Assembly of recombinant EJCs .....	96
6.4.9	Isothermal titration calorimetry (ITC) .....	96

<b>6.5</b>	<b><i>In vitro</i> splicing and RNP immunoprecipitation.....</b>	<b>96</b>
6.5.1	Generation of HEK293 whole cell extracts.....	96
6.5.2	<i>In vitro</i> transcription of radiolabeled mRNA.....	97
6.5.3	<i>In vitro</i> splicing reaction .....	97
6.5.4	RNP immunoprecipitation.....	98
6.5.5	RNA extraction.....	98
6.5.6	Denaturing TBE-gel electrophoresis.....	99
6.5.7	<i>In vitro</i> splicing and oligonucleotide-directed RNase H digestion.....	99
<b>6.6</b>	<b>Bioinformatics and statistics .....</b>	<b>100</b>
<b>7</b>	<b>Supplementary Figures.....</b>	<b>101</b>
<b>8</b>	<b>Bibliography.....</b>	<b>105</b>

# 1 Abstract

## 1.1 Deutsche Zusammenfassung

Pre-mRNA Spleißen in Säugerzellen beeinflusst die Proteinzusammensetzung von Ribonucleoproteinkomplexen (mRNP). Ein prominentes Beispiel hierfür ist der Exon-Verbindungs Komplex ('exon junction complex', EJC), der die gespleißte mRNA nahe der Exon-Exon Grenze bindet und so die ehemaligen Spleiß-Stellen markiert. Der EJC koordiniert viele wichtige Aspekte der Genexpression, so zum Beispiel den Export der mRNA ins Zytoplasma, ihre zelluläre Lokalisation, die Translations-Effizienz und die Qualitätskontrolle durch den Mechanismus des Nonsense-vermittelte mRNA-Abbaus (NMD). Doch obgleich die Zusammensetzung des EJC intensiv untersucht wurde, blieb bislang unklar, wie das Spleißosom den Zusammenbau des Komplexes reguliert.

In dieser Studie beschreibe ich nun die Identifikation von CWC22 als einen essentiellen Spleiß-Faktor, der den Zusammenbau des EJC in humanen Zellen vermittelt. Die mittlere Domäne (MIF4G) von CWC22 bindet direkt an das zentralen EJC Protein eIF4A3. Diese Bindung wiederum vermittelt die Interaktion zwischen eIF4A3 und dem Spleißosom. Bemerkenswert ist, dass CWC22 auch für das Spleißen von pre-mRNA benötigt wird. Diese Spleiß-Funktion des Proteins wird von einem zentralen Segment, bestehend aus der MIF4G Domäne und der benachbarten MA3 Domäne, vermittelt. Indem CWC22 sowohl das Spleißen von pre-mRNA, als auch den Aufbau von EJCs reguliert, nimmt das Protein eine duale Rolle in der Genregulation von Säugerzellen ein. Interessanterweise lassen sich beide Funktionen experimentell trennen, so dass die Überexpression einer dominant negativen Muntante von CWC22 zwar den Aufbau von EJCs inhibiert, gleichzeitig aber keinen Effekt auf das Spleißen von pre-mRNA hat. Die Überexpression dieser Mutante hat einen toxischen Effekt auf kultivierte Zellen, was die Bedeutung des EJC für die Kontrolle der Genexpression unterstreicht.

Zusammenfassend lässt sich sagen, dass das Protein CWC22 die erste direkte Verbindung zwischen dem EJC und dem Spleißosom darstellt, während die duale Funktion des Proteins die Komplexität der Genregulation in Säugerzellen illustriert.



## 1.2 Summary

The exon junction complex (EJC) is a multi-protein complex that assembles on mRNA during splicing. It binds upstream of exon-exon junctions and preserves the positional information of the former splice sites within the mRNP architecture. In metazoan cells, the EJC coordinates several key aspects of gene expression, such as mRNA export, cytoplasmic localization, translation efficiency and quality control by nonsense-mediated mRNA decay (NMD). Although the composition of EJCs is well understood, the mechanism mediating splicing-dependent EJC assembly and the factor(s) recruiting the EJC to the spliceosome remain elusive.

Here, I describe the identification of CWC22 as an essential splicing factor that is required for EJC assembly in human cells. The central domain (MIF4G) of CWC22 directly interacts with the EJC core protein eIF4A3, and this interaction is reminiscent of the eIF4A/eIF4G translation initiation complex. The interaction with CWC22 recruits eIF4A3 to the early spliceosomal complex and initiates EJC assembly. Mutations in eIF4A3 that disrupt the interaction with CWC22 also prevent splicing-dependent EJC assembly, but do not affect the formation of splicing-independent EJCs from recombinant components *in vitro*. Interestingly, CWC22 is also required for pre-mRNA splicing, and unspliced pre-mRNAs accumulate in CWC22-depleted cells. Complementation assays revealed that a central segment of CWC22, comprising of the MIF4G and an adjacent MA3 domain, supports pre-mRNA splicing in CWC22-depleted cells. By regulating both pre-mRNA splicing and EJC assembly, CWC22 plays a dual role in mammalian gene expression. The two functions of CWC22 can be experimentally separated, since over-expression of a CWC22-mutant that is unable to bind eIF4A3 inhibits EJC assembly without affecting pre-mRNA splicing. Over-expression of this mutant is highly toxic to cells, demonstrating the importance of EJCs for gene regulation in human cells.

Taken together, CWC22 represents the first direct link between the splicing machinery and the EJC, and the discovery of its dual function during pre-mRNA splicing highlights the complex interconnection of individual processing steps during mammalian gene expression.



## 2 Introduction

### 2.1 Regulation of eukaryotic gene expression

Eukaryotic gene expression is a complex, multi-step process (Moore, 2005; Reed, 2003). Transcription of protein-coding genes is carried out in the nucleus by the RNA polymerase II. But before a nascent RNA polymerase II transcript is transported to the cytoplasm, it usually undergoes 5' end capping, 3' end processing and pre-mRNA splicing. Eukaryotic cells have evolved an elaborate regulatory system to facilitate the proper coordination of individual mRNA processing steps and to ensure that only fully processed, mature mRNA molecules are exported. Most of the processing steps occur co-transcriptionally and RNA polymerase II itself plays a key role in orchestrating the sequential recruitment of processing factors to the nascent mRNA molecule, a function that is carried out by the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (Armache et al., 2005). The CTD is a long, unstructured region, consisting of multiple heptad repeats (52 repeats (Tyr-Ser-Pro-Thr-Ser-Pro-Ser) in mammals). Each repeat contains three serine and one tyrosine residues that undergo sequential phosphorylation and dephosphorylation cycles during transcription (Egloff and Murphy, 2008). Phosphorylation of serine 5 (Ser5P) is a hallmark of early transcription, whereas serine 2 phosphorylation (Ser2P) marks later stages of transcription elongation (Komarnitsky et al., 2000). Many mRNA processing factors differentially associate with the phosphorylated CTD of RNA polymerase II, thus facilitating the integration of mRNA transcription and processing. The following three paragraphs will characterize the individual processing steps in more detail.

#### 2.1.1 5' end capping of mRNA

The earliest mRNA processing event is 5' end capping. As soon as the nascent mRNA molecule reaches a length of 22-25 nucleotides (nt), it emerges from the RNA polymerase II exit channel and encounters several processing enzymes that catalyze the methyl-guanylation of its 5' end (Shatkin, 1976; Shuman, 2001). Capping requires three consecutive enzymatic reactions that are carried out by two proteins in

mammals: A triphosphatase guanylyltransferase removes a phosphate from the mRNA 5' end and catalyzes the atypical 5'-5' addition of a guanosine-monophosphate (GMP) (Yamada-Okabe et al., 1998). Subsequently, a methyl transferase modifies the guanosine by adding a methyl group (Tsukamoto et al., 1998). The triphosphatase guanylyltransferase is recruited to the transcription machinery upon Serine-5 phosphorylation of the RNA polymerase II CTD, thus coupling mRNA capping to early transcription elongation (Ho and Shuman, 1999). Marking the 5' end of an RNA polymerase II transcript with a 7-methylguanosine (m<sup>7</sup>G) cap helps to distinguish protein-coding mRNA from other types of RNA synthesized by RNA polymerase I or III.

The newly synthesized m<sup>7</sup>G-cap is immediately bound by the cap-binding complex (CBC), consisting of the two major cap-binding proteins (CBPs) CBP20 and CBP80. Hence, the CBC is the first protein complex to associate with nascent mRNA. The complex interacts with many other mRNA binding proteins and thereby couples mRNA capping to subsequent events of mRNA metabolism, such as splicing, export, translation and both nuclear and cytoplasmic mRNA surveillance. During the first round of translation, the nuclear CBC is replaced by a different cap-binding protein complex, called eIF4F. EIF4F consists of the cap-binding protein eIF4E, the RNA helicase eIF4A and the large scaffolding protein eIF4G. The protein complex is an important regulator of bulk protein synthesis by recruiting the ribosomal pre-initiation complex to mRNA and initiating ribosomal scanning (Sonenberg and Hinnebusch, 2009). Moreover, eIF4G interacts with PABPC, thus forming a 'closed loop' of mRNA that is supposed to facilitate ribosomal recycling (Kahvejian et al., 2001).

### **2.1.2 3' end processing of mRNA**

Similar to capping of the 5' end, the 3' end of protein-coding transcripts is modified in a very characteristic way: 3' end processing includes cleavage of the mRNA and subsequent addition of up to 200 non-templated adenosine nucleotides (the so-called poly(A) tail) to the 3' hydroxyl end. These processing steps are carried out by cleavage and polyadenylation (CPA) factors that are recruited to transcribing RNA polymerase II through direct interaction of CTD-interacting domains (CIDs) with Ser2P CTD, as well as RNA-binding domains that specifically recognize the poly(A) signal (AAUAAA) on the nascent pre-mRNA (Bentley, 2005; Proudfoot, 2004). Polyadenylation is a dynamic

process and the use of alternative polyadenylation sites can give rise to mRNA transcripts with different 3' UTRs (Elkon et al., 2013). Tight coupling of mRNA polyadenylation to export ensures that only correctly terminated transcripts are transported to the cytoplasm (Proudfoot, 2011). Moreover, polyadenylation generates the binding site for an important class of gene regulatory proteins, the poly(A) binding proteins (PABPs) (Kuhn and Wahle, 2004). PABPs impinge on both nuclear and cytoplasmic mRNA metabolism. Nuclear PABP (PABPN1) is involved in poly(A) tail synthesis itself by stimulating the processivity of poly(A) polymerases, whereas cytoplasmic PABP (PABPC) plays a crucial role during translation initiation and mRNA surveillance by interacting with the m<sup>7</sup>G-cap and forming a 'closed loop' between the 5' end and the 3' end of the transcript (Kahvejian et al., 2001).

Hence, by marking both ends of an mRNA molecule with characteristic modifications, a cell can ensure that only fully processed RNA polymerase II transcripts are translated into protein.

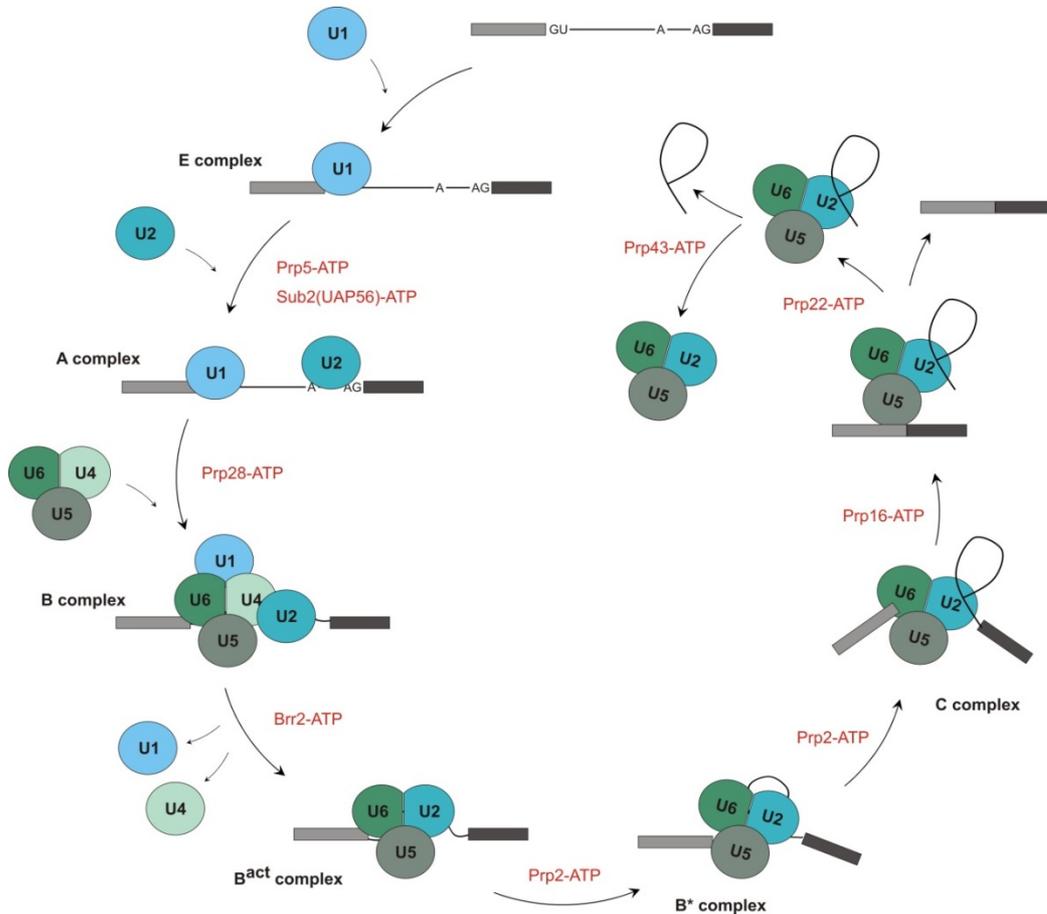
### **2.1.3 Pre-mRNA splicing**

Splicing is undoubtedly the most extensive form of pre-mRNA processing. The spliceosome, a large multi-unit ribonucleoprotein (RNP) complex, catalyzes the excision of large non-coding intronic regions from the pre-RNA. In contrast to the rather invariant 5' and 3' modifications described before, splicing is a highly dynamic process and the use of alternative splice sites greatly enhances the complexity of the proteome in higher eukaryotes (Blencowe, 2006).

Chemically, pre-mRNA splicing consists of two consecutive transesterification reactions. During the first step, a 2'-hydroxyl group of a conserved adenosine in the branch point sequence (BPS) of the intron attacks the phosphodiester bond at the 5' splice site (GU), generating a 5' exon with a free 3' hydroxyl group and an intron-lariat that is covalently linked to the 3' exon. Subsequently, the free hydroxyl group of the 5' exon attacks the phosphodiester bond of the 3' splice site (AG), simultaneously joining 5' exon to 3' exon and liberating the intron-lariat.

The major structural components of the spliceosome are uridine-rich small nuclear RNAs (snRNAs) that associate with a variable number of proteins to form the

spliceosomal building blocks, called snRNPs (Rappsilber et al., 2002; Zhou et al., 2002). Interestingly, the spliceosome is not a rigid structure, but instead assembles stepwise on the mRNA (Wahl et al., 2009). In order to distinguish the individual snRNP complexes that form during splicing, the splicing intermediates have been designated as E, A, B, B<sup>act</sup>, B\* and C complex (Figure 1).



**Figure 1: Pre-mRNA splicing is a multi-step process.** Stepwise assembly of the spliceosome from individual snRNPs. The ATP-dependent remodeling steps and the respective DExD/H-box helicases are depicted in red.

The first spliceosomal component that encounters the substrate mRNA is the U1 snRNP. U1 binds the 5' splice site of the intron through base pairing. At the same time, the SF1/BBP protein and the U2 auxiliary factor (U2AF) interact cooperatively with the branch point sequence (BPS), the polypyrimidine tract and the AG-dinucleotide of the 3' splice site. Thus, this initial spliceosomal complex (E complex, commitment complex) is involved in the definition of both 5' and 3' splice site of the intron. Next, the U2 snRNP binds the BPS through base pairing, thereby displacing SF1/BBP and yielding the spliceosomal A-complex (pre-spliceosome). The base pairing is stabilized by two

protein complexes of the U2 snRNP, SF3a and SF3b. Subsequently, the pre-assembled U4/U6.U5 tri-snRNP is recruited to the pre-mRNA, forming the B complex (pre-catalytic spliceosome). Release of U1 and U4 yields the activated spliceosome ( $B^{\text{act}}$  complex) that is further rearranged by the helicase Prp2 to form the catalytically active  $B^*$  complex (Bessonov et al., 2010; Bessonov et al., 2008). The first catalytic step of splicing transforms  $B^*$  to C complex, while further rearrangements and the second catalytic step of splicing generate the post-spliceosomal complex and mediate the release of spliced mRNA, U2, U5 and U6 snRNP (Wahl et al., 2009).

Although the transesterification reactions are energetically neutral, most of the spliceosomal rearrangements are catalyzed through ATP hydrolysis. Eight evolutionary conserved helicases function at different steps during splicing to mediate the ATP-dependent remodeling of RNA-RNA interactions within the spliceosome. They all belong to the family of DExD/H-box RNA-dependent ATPases/helicases and will be further discussed in chapter 2.4.

Apart from modifying the mRNA molecule itself, all mRNA processing steps described here facilitate the assembly of protein complexes on the mRNA. These RNA-binding proteins (RBPs) regulate and integrate many different steps of the mRNA life cycle, from synthesis in the nucleus until degradation in the cytoplasm. A few of these RBPs (CBP20, CBP80, PABP) have already been briefly introduced; a more detailed description of the assembly and function of mRNPs will follow in the next chapter.

## **2.2 mRNP assembly connects different steps of gene expression**

Analogous to DNA-packaging into chromatin, cellular mRNAs associate with many RBPs to form ribonucleoprotein complexes (mRNPs). The first proteins engage the mRNA during early transcription elongation and subsequent remodeling steps in the nucleus and cytoplasm generate a diverse pool of different mRNPs. The composition of individual mRNPs reflect the character of the bound mRNA molecule as well as the history of earlier processing events. The roles of different RBPs are as diverse as the composition of the individual mRNPs. Some RBPs mainly induce the condensation of the mRNA molecule (Konig et al., 2010). Others are involved in regulating and integrating the different steps of gene expression (Moore and Proudfoot, 2009;

Muller-McNicoll and Neugebauer, 2013). And yet another function of RBPs is the prevention of R-loop formation (Aguilera and Klein, 1990; Huertas and Aguilera, 2003; Li and Manley, 2005). These DNA:RNA hybrids form during transcription when the newly synthesized single-stranded mRNA invades the DNA duplex behind the elongating polymerase, thus leading to elongation defects and genomic instability (Mischo et al., 2011).

Many studies aimed to characterize the mRNA interactome. Initial experiments focused on the analysis of individual mRNPs that were generated through *in vitro* splicing of mRNA substrates (Merz et al., 2007). Mass spectrometric analyses of these model mRNPs identified ~45 RNA-interacting proteins (Merz et al., 2007). Using a more global approach, two recent studies applied *in vivo* UV crosslinking and immunoprecipitation to compile a comprehensive list of proteins bound to endogenous mRNAs in human embryonic kidney (HEK) 293 and HeLa cells, respectively (Baltz et al., 2012; Castello et al., 2012). Both studies identified around 800 RNA-binding proteins, many of which were not previously known to interact with mRNA (Baltz et al., 2012; Castello et al., 2012). The large number of identified RBPs illustrates the high complexity of the mRNA interactome. Apart from canonical RBPs that invariantly interact with each mRNA transcript, differentially regulated RNA-protein complexes are likely to integrate many cellular signals and to adjust gene expression to the metabolic state of a cell. In line with this hypothesis is the finding that several metabolic enzymes interact with mRNA *in vivo* (Castello et al., 2012).

RBPs use diverse mechanisms to interact with mRNA. While some RBPs bind to mRNA through structural elements (e.g. sequence motifs, secondary structures or chemical modification such as the m<sup>7</sup>G-cap), the recruitment of other RBPs is linked to certain mRNA processing steps. A prominent example for the latter is the exon-junction complex (EJC), a multi-protein complex that assembles on mRNA during splicing.

### **2.3 The exon junction complex is part of the splicing-dependent mRNP**

Apart from enhancing the complexity of the proteome, pre-mRNA splicing in higher eukaryotes plays a crucial role in shaping the protein composition of mRNPs. A central component of post-splicing mRNPs is the exon junction complex (EJC). This multi-

protein complex binds mRNA in close proximity to the exon-exon junction, thus conserving the knowledge of the completed splicing reaction within the mRNP architecture (Le Hir et al., 2000a). The EJC is exported to the cytoplasm as an integral component of the post-splicing mRNP where it acts as a key regulator of gene expression. The details of EJC structure (2.3.1), assembly and disassembly (2.3.2) and function (2.3.3) will be discussed in the following chapters.

### 2.3.1 Structure of the EJC

The core of the EJC is a stable heterotetramer consisting of the proteins eIF4A3 (DDX48), MAGOH, Y14 (RBM8A) and Barentsz (BTZ, also known as MLN51 or CASC3) (Palacios et al., 2004; Shibuya et al., 2004). In living cells, the EJC core forms the binding platform for the dynamic assembly of many peripheral EJC components that mechanistically link the complex to downstream events in mRNA metabolism (Le Hir et al., 2001; Tange et al., 2005). *In vitro*, the EJC core can be reconstituted from recombinant proteins, allowing the analysis of its stable, yet sequence-independent interaction with RNA (Andersen et al., 2006; Ballut et al., 2005; Bono et al., 2006).

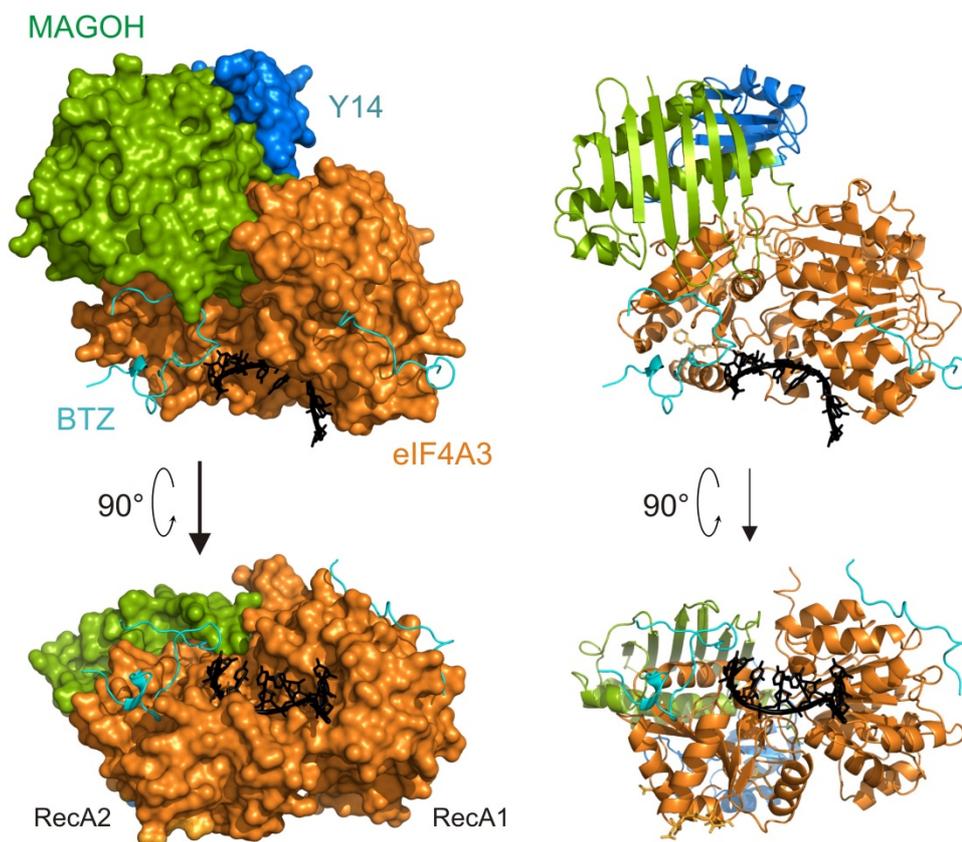
The major RNA-binding protein within the EJC is the DEAD-box protein eIF4A3 (for more details on DEAD-box helicases see chapter 2.4). EIF4A3 has a helicase core, consisting of two RecA-like domains (the N-terminal RecA1 domain and the C-terminal RecA2 domain) that are connected by a flexible linker. It is a close homolog of the eukaryotic translation initiation factor 4A (encoded by eIF4A1 and eIF4A2) and has an RNA-dependent ATPase and ATP-dependent RNA helicase activity *in vitro* (Li et al., 1999). But while eIF4A unwinds secondary structures in the mRNA 5'UTR during translation initiation, eIF4A3 anchors the EJC to spliced mRNA (Ferraiuolo et al., 2004; Kataoka et al., 2001; Palacios et al., 2004; Shibuya et al., 2004). The non-processive RNA binding of eIF4A3 is achieved through inhibition of the ATPase activity within the complex (Ballut et al., 2005). The molecular mechanism of this stable RNA interaction was revealed by two independent crystal structures of the EJC core at 2.2-2.3Å resolution (Andersen et al., 2006; Bono et al., 2006). EIF4A3 was crystallized in complex with the SELOR-domain of BTZ (residues 137–286), MAGOH, a piece of Y14 (66–154), polyuracil RNA and the non-hydrolyzable ATP analogue ADPNP (Andersen et al., 2006; Bono et al., 2006). The structures revealed that EJC-bound eIF4A3 adopts a

'closed' conformation with the two RecA-like domains in close proximity to each other (Figure 2). This conformation generates a deep interdomain cleft that holds the ADPNP. The RNA is bound in a shallow cleft across the two RecA-like domains at the surface opposite to the ATP binding pocket. Six nucleotides are resolved in the crystal structure, showing that the RNA binds to eIF4A3 in a bend conformation with the 3' end interacting with the RecA1 domain and the 5' end binding to RecA2 (Figure 2, bottom panel). EIF4A3 mainly contacts the ribose-phosphate backbone of the RNA, which explains the sequence-independent RNA binding, while interactions with the 2' OH-groups of the ribose explain the specificity for RNA.

The interaction between MAGOH and Y14 within the complex is almost identical to the conformation of the isolated heterodimer (Fribourg et al., 2003; Lau et al., 2003; Shi and Xu, 2003), suggesting that the two proteins form a constitutive unit. The MAGOH/Y14 heterodimer interacts mostly with the RecA2 domain of eIF4A3 and the flexible interdomain linker (Figure 2). Interestingly, the heterodimer can only interact with the ATP- and RNA-bound conformation of eIF4A3. Thus, the interaction with MAGOH and Y14 locks eIF4A3 in its 'closed' conformation on the RNA. This explains why MAGOH and Y14 inhibit the ATPase activity of eIF4A3 *in vitro* (Ballut et al., 2005). The EJC crystal structure with non-hydrolysable ATP analogue ADPNP does not define whether binding of MAGOH/Y14 to eIF4A3 prevents ATP hydrolysis itself or the release of the products (Le Hir and Andersen, 2008). This question was solved by replacing ADPNP with ADP-AIF(3), a mimic of the transition state during ATP hydrolysis (Nielsen et al., 2009). The ATP binding pocket within the EJC-ADP-AIF(3) complex was essentially identical to EJC-ADPNP, demonstrating that ATP hydrolysis can take place within the complex and proving that that MAGOH/Y14 do not inhibit the ATPase activity of eIF4A3 but rather the release of ADP and inorganic phosphate (Nielsen et al., 2009). Of note, disruption of the canonical Walker A and Walker B motifs of eIF4A3 inhibit the RNA-dependent ATPase activity, but had no effect on EJC formation (Shibuya et al., 2006). Together with the structural data, this suggested that nucleotide binding, but not ATP hydrolysis is required for EJC assembly.

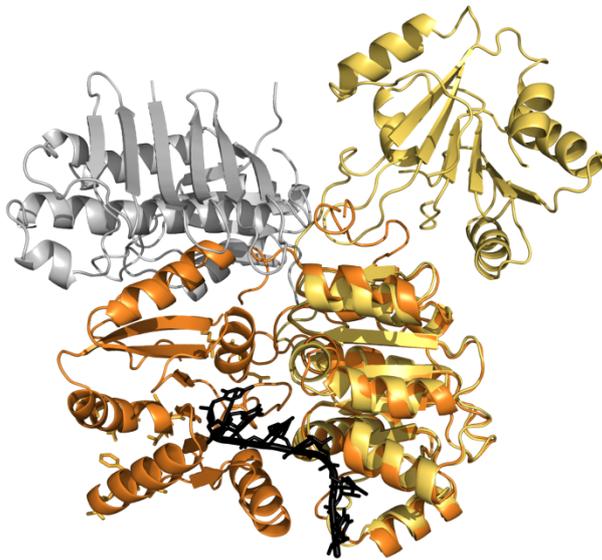
Neither MAGOH, nor Y14 directly interact with the RNA in the crystal (Andersen et al., 2006; Bono et al., 2006). Instead, the SELOR domain of BTZ contributes to the RNA

binding pocket through stacking interactions with the 5' base of the RNA (Figure 2, bottom panel). This observation suggests that BTZ increases the RNA-binding affinity of eIF4A3 and explains why BTZ stimulates the ATPase activity of eIF4A3 *in vitro* (Ballut et al., 2005). Most of the BTZ residues in the crystal structure are disordered except for two distinct EJC-interaction patches. Residues 214-248 interact with the RecA1 domain of eIF4A3, whereas residues 168-196 bind to RecA2, RNA and the MAGOH/Y14 heterodimer. Of note, the same C-terminal residues of BTZ mediate the interaction with apo-eIF4A3 in the eIF4A3-BTZ crystal, whereas the interactions with the RecA2 domain of eIF4A3 are specific for the EJC.



**Figure 2: Crystal structure of the EJC.** EIF4A3 in complex with the MAGOH/Y14(66-154) heterodimer, BTZ (137–286), poly(U) RNA and ADPNP (PDB file 2J0S). View from the side (top panel) and rotated 90° (bottom panel)

Comparison of eIF4A3 as part of the EJC and in complex with BTZ revealed large conformational changes upon complex formation (Figure 3). When bound to BTZ, eIF4A3 adopts an 'open' conformation. In this conformation, the two RecA-like domains are tilted at an angle of approximately 180°, which leads to a disruption of the ATP and RNA-binding sites (Figure 3). This suggests that eIF4A3 has to undergo drastic structural rearrangements to form the EJC. How these changes are initiated during pre-mRNA splicing is discussed in the following chapter.



**Figure 3: EIF4A3 undergoes large conformational rearrangements upon EJC assembly.** Overlay of the structure of eIF4A3 as part of the EJC (in orange, PDB file 2J0S) and apo-eIF4A3 (in yellow, PDB file 2J0U). The picture was generated by superimposing both RecA1 domains. Note that the RecA2 domain of apo-eIF4A3 is tilted approximately 180° compared to its position within the EJC. The MAGOH/Y14 heterodimer is depicted in grey.

### 2.3.2 Assembly and disassembly of EJCs

EJC assembly on nascent mRNAs is mediated by the spliceosome (Kataoka et al., 2000; Le Hir et al., 2000b). Independent of the mRNA sequence, the complex assembles 20-24 nt upstream of the splice site (Le Hir et al., 2000a; Mishler et al., 2008).

Purified spliceosomal B and C complexes contain the EJC core proteins eIF4A3, MAGOH and Y14, indicating that those proteins are recruited to the spliceosome prior to exon

ligation (Bessonov et al., 2008; Deckert et al., 2006). Interestingly, BTZ is absent from spliceosomal purifications. Accordingly, *in vitro* splicing experiments showed that eIF4A3, MAGOH and Y14, but not BTZ, interact with splicing intermediates that are stalled after the first catalytic step of splicing (Gehring et al., 2009a; Merz et al., 2007; Reichert et al., 2002). Moreover, disruption of the BTZ-interaction site on eIF4A3 did not prevent splicing-dependent EJC assembly (Gehring et al., 2009a). Together, these findings suggest that BTZ is dispensable for early EJC formation. This led to the proposal of a stable trimeric 'pre-EJC', consisting of eIF4A3, Y14 and MAGOH (Gehring et al., 2009a). In mammalian cells, BTZ is predominantly localized in the cytoplasm (Degot et al., 2004); it is therefore conceivable that it interacts with the preassembled trimeric pre-EJC after nuclear export of the mRNP (Gehring et al., 2009a). This idea is supported by the observation that BTZ can bind to mRNA after completion of the splicing reaction (Gehring et al., 2009a). In this regard, BTZ behaves more like a peripheral EJC component that requires the assembled pre-EJC, but not the splicing reaction itself, to interact with mRNA.

Even though eIF4A3, MAGOH and Y14 all associate with spliceosomal B and C complexes, *in vitro* splicing experiments revealed differences concerning their interaction with spliced mRNA (Gehring et al., 2009a). Whereas the MAGOH/Y14 heterodimer requires the EJC binding site (position -24) on the mRNA to interact with splicing intermediates, the early spliceosomal interaction of eIF4A3 is independent of this mRNA sequence. Furthermore, the MAGOH/Y14 heterodimer requires the interaction with eIF4A3 for its recruitment to the spliceosome, whereas eIF4A3 is able to interact with the splicing machinery independently of the heterodimer (Gehring et al., 2009a). Nevertheless, the binding of MAGOH/Y14 to eIF4A3 is required to stabilize the EJC on the mature mRNA, presumably through inhibition of the ATPase activity of eIF4A3 (Ballut et al., 2005; Gehring et al., 2009a). Together, these results led to the current model for the stepwise assembly of EJCs on spliced mRNA (Figure 4 (Gehring et al., 2009a)):

- (1) EJC assembly is initiated when eIF4A3 is recruited to the spliceosomal B or C complex in an RNA-independent manner.

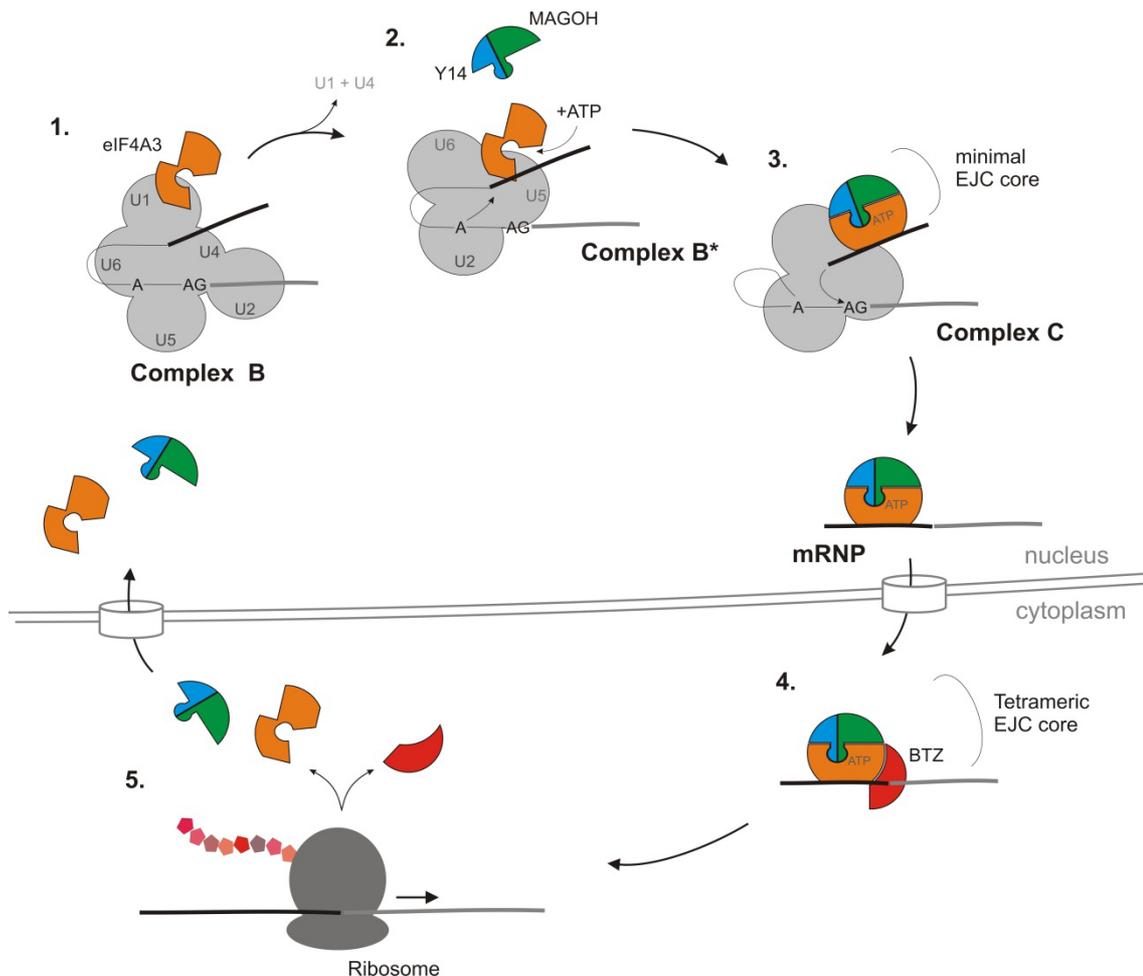
(2) Upon interaction with the mRNA 20-24 nt upstream of the splice site, eIF4A3 adopts its closed conformation, thereby forming a binding platform for the recruitment of the MAGOH/Y14 heterodimer.

(3) Binding of MAGOH/Y14 inhibits the ATPase activity of eIF4A3 and stabilizes the complex on the mRNA.

(4) Binding of BTZ to the pre-EJC (in the nucleus or cytoplasm) further stabilizes the complex.

How the interaction between EJC proteins and the splicing machinery is established is currently unknown. It was suggested however, that the intron-binding protein IBP160 might be involved in the recruitment of EJC proteins to spliced mRNA (Ideue et al., 2007).

While the assembly of EJCs is closely linked to pre-mRNA splicing, disassembly of the complex is connected to yet another important step of gene expression, the translation of mRNA in the cytoplasm (Dostie and Dreyfuss, 2002; Lejeune et al., 2002). It is believed that the processive helicase activity of the ribosome itself is sufficient to displace RBPs from the mRNA (Takyar et al., 2005). However, the ribosome-associated protein PYM has been shown to aid the translation-dependent disassembly of EJCs (Gehring et al., 2009b). The C-terminal part of PYM binds the 40S subunit of the ribosome (Diem et al., 2007), whereas the N-terminus can interact with the MAGOH/Y14 heterodimer (Bono et al., 2004). The binding of MAGOH/Y14 to either PYM or eIF4A3 is mutually exclusive, so that PYM most likely triggers EJC disassembly by displacing eIF4A3 (Bono et al., 2006; Gehring et al., 2009b). PYM-mediated disassembly of EJCs is important for the recycling of EJC proteins, as evidenced by an accumulation of cytosolic EJC components in PYM-depleted cells (Gehring et al., 2009b). Through interaction with the 40S ribosomal subunit, PYM-mediated EJC disassembly is restricted to mRNAs that undergo their initial round of translation. This close link to translation ensures that EJCs stay associated with the mRNA until they have fulfilled their regulatory role during mRNA metabolism (e.g. in NMD, discussed in chapter 2.3.3).



**Figure 4: Assembly and disassembly of EJC\*s.** EJCs assemble on the mRNA during pre-mRNA splicing in the nucleus (1-3). After export as a component of the mRNP, EJCs regulate the cytoplasmic fate of the mRNA (4). EJCs are disassembled by the translating ribosome (5) and the individual components are reimported into the nucleus. For details see paragraph 2.3.2. \*The image was inspired by drawings from Niels Gehring.

This chapter described the sequence-independent assembly of canonical EJCs in close proximity to the splice site. At least in *Drosophila melanogaster* however, it has been shown that certain *cis*-acting sequence elements can regulate the differential EJC assembly on spliced mRNA (Sauliere et al., 2010). This observation suggested that EJC assembly might not be a constitutive consequence of splicing, but indeed a regulated process. The hypothesis was further supported by two recent studies that applied deep sequencing approaches to analyze EJC binding sites in human cells on a transcriptome-wide level (Sauliere et al., 2012; Singh et al., 2012). While both studies show that EJC proteins mainly associated with spliced mRNAs and around half of the sequence reads correspond to canonical EJC deposition sites (cEJC), they also

identified a large number of non-canonical EJC (ncEJC) binding sites. Interestingly, occurrence of ncEJCs correlated with a high propensity for purine-rich sequences, exonic splice enhancers (ESE) and binding sites for SR proteins. Moreover, a physical interaction between adjacent EJCs seems to lead to a higher density of ncEJCs in the vicinity of cEJC sites. This is in line with the observation that EJC proteins form higher order mRNP structures comprising of several EJCs and super-stoichiometric amounts of SR proteins (Singh et al., 2012).

The EJC is a key regulator of post-transcriptional gene expression. The recent discovery of non-canonical EJC binding sites and EJC-SR super-complexes has therefore added another layer of complexity to the EJC-dependent gene regulatory network. In order to understand the implications of these findings for gene expression in human cells, the different cellular functions of the EJC will be described in the following chapter.

### **2.3.3 The EJC regulates different steps of the mRNA life cycle**

As an important component of the post-splicing mRNP, the EJC regulates and integrates many steps of the mRNA life cycle. To this end, the core EJC differentially interacts with many accessory proteins often referred to as peripheral EJC (Le Hir et al., 2001; Tange et al., 2005). Among these peripheral EJC components are proteins involved in mRNA export, splicing and mRNA quality control.

#### ***(a) EJCs stimulate protein synthesis***

It has been known for a long time that the presence of introns in the coding sequence of genes enhances protein synthesis (Braddock et al., 1994; Callis et al., 1987; Matsumoto et al., 1998). The effect of splicing on translational yield can be directly attributed to the presence of EJCs on the mRNA (Nott et al., 2004; Wiegand et al., 2003). To date, three different mechanisms have been proposed to explain the stimulatory effect of EJCs on translation. The first mechanistic insight came from a study showing that the protein PYM, through interaction with MAGOH/Y14 and the 48S pre-initiation complex, recruits EJC-bound mRNAs to the translation machinery (Diem et al., 2007). A second study described a link between EJC and mTOR1 pathway through recruitment of activated S6 kinase 1 (S6K1) to newly transcribed mRNAs by an EJC-bound protein called SKAR (S6K1/AlyREF-like target) (Ma et al., 2008). The mTOR

pathway regulates cellular growth in response to external and internal stimuli and its key component S6K1 promotes translation initiation by phosphorylating stimulatory and inhibitory factors bound to the 5' end of the mRNA (Bhaskar and Hay, 2007). The third link between EJC and translation machinery was established by a recent study that attributed the EJC-dependent translation activation to a direct interaction between the EJC core protein BTZ and the translation initiation factor eIF3 (Chazal et al., 2013).

Importantly, all described mechanisms influence early translation initiation. Since EJCs are disassembled during the first round of translation (see chapter 2.3.2), EJC-dependent translational activation is restricted to newly synthesized transcripts. The preferential translation of new mRNAs reduces the delay time between transcription activation and protein synthesis, therefore allowing a faster response to external stimuli (Giorgi and Moore, 2007).

### *(b) EJCs stimulate mRNA export*

The increased translational yield of spliced mRNA might also be partly due to a stimulatory effect of EJCs on the nucleocytoplasmic export of mRNA (Le Hir et al., 2001; Luo and Reed, 1999; Luo et al., 2001; Zhou et al., 2000). The peripheral EJC contains several proteins with known functions in mRNA export. The best studied export factors of the peripheral EJC are the proteins ALYREF (THOC4) and UAP56 (Luo et al., 2001; Rodrigues et al., 2001; Stutz et al., 2000; Zhou et al., 2000). Both function as export adaptors that mediate the recruitment of the nuclear export receptor NXF1/TAP:p15 to the mRNA (Stutz et al., 2000). The subsequent interaction of NXF1/TAP:p15 with the nuclear pore complex promotes the export of mRNA (Bachi et al., 2000). ALYREF and UAP56 are also components of another RNA-binding protein complex, the TREX1 (transcription export) complex. Like the EJC, the TREX1 complex preferentially assembles on spliced mRNAs (Masuda et al., 2005). Nevertheless, a splicing-independent, cap-dependent TREX1 assembly has been described (Cheng et al., 2006; Nojima et al., 2007). This suggests that the export adaptors ALYREF and UAP56 might also function in the export of intronless mRNAs.

### (c) EJC regulate the subcellular localization of mRNA

A well-documented function of EJCs is the localization of *oskar* mRNA in the *Drosophila* oocyte. Together with the EJC proteins eIF4A3, BTZ, Mago Nashi and Tsunagi (the *Drosophila* orthologs of MAGOH and Y14), *oskar* mRNA localizes to the posterior pole of the oocyte (Hachet and Ephrussi, 2001; Mohr et al., 2001; van Eeden et al., 2001). The subsequent local translation of mRNA is required for the formation of germ line and abdomen in the developing embryo. So far, no similar mechanism of subcellular mRNA localization through EJC proteins has been described in other organisms. However, the localization of eIF4A3, MAGOH, Y14 and BTZ to neuronal dendrites suggested that EJC proteins might also function in other cell types that require a regulated local mRNA translation (Giorgi et al., 2007; Glanzer et al., 2005; Macchi et al., 2003; Monshausen et al., 2004).

### (d) EJCs regulate nonsense-mediated mRNA decay

The best-characterized function of EJCs is the identification of faulty transcripts during nonsense-mediated mRNA decay (NMD). NMD is a cellular quality control mechanism that detects and degrades transcripts with premature translation termination codons (PTCs) (Conti and Izaurralde, 2005). A termination codon is recognized as premature when it is situated at least 50 nt upstream of an exon-exon junction that is marked by an EJC (Le Hir et al., 2000a; Nagy and Maquat, 1998; Thermann et al., 1998). Alternatively, NMD can be triggered by other features of the mRNA (e.g. long 3' UTRs) that inhibit efficient translation termination (Amrani et al., 2004), but this alternative (and EJC-independent) NMD will not be further discussed in this chapter. NMD is a translation-dependent process that requires the crosstalk between the stalled ribosome and factors bound to the downstream EJC. Genetic screens in yeast and *caenorhabditis elegans* identified a number of evolutionary conserved NMD factors, among them the Upf proteins Upf1, Upf2 and Upf3 (Upf3a and Upf3b in humans) and the SMG proteins SMG1, SMG5, SMG6 and SMG7 (Cui et al., 1995; He et al., 1997; Hodgkin et al., 1989; Leeds et al., 1991). The current model for NMD activation in mammalian cells suggests that Upf1 and SMG1 are recruited to the stalled ribosome through interaction with the release factors eRF1 and eRF3, forming the so-called SURF (SMG1-Upf1-eRF1-eRF3) complex (Czapinski et al., 1998; Kashima et al., 2006). Binding of Upf1 to Upf2 and Upf3 at the downstream EJC activates the kinase SMG1,

which in turn phosphorylates serine-residues at the N- and C-terminal domain of Upf1 (Chamieh et al., 2008; Yamashita, 2013). Phosphorylated Upf1 serves as a binding platform for the recruitment of SMG6 and the SMG5/SMG7 heterodimer through 14-3-3 like domains in all three proteins (Ohnishi et al., 2003; Okada-Katsuhata et al., 2012). SMG5, SMG6 and SMG7 act in concert to promote the endo- and exonucleolytic degradation of the target mRNA. SMG6 has an endonuclease activity that directly cleaves the mRNA in the vicinity of the PTC (Eberle et al., 2009; Huntzinger et al., 2008), whereas the SMG5/SMG7 heterodimer recruits a deadenylase complex to the mRNA, thus initiating deadenylation, deadenylation-dependent decapping and 5' to 3' exonucleolytic decay of NMD targets (Loh et al., 2013). Moreover, Upf1 itself is involved in target mRNA degradation through recruitment of decapping factors (Loh et al., 2013; Lykke-Andersen, 2002). Apart from promoting target mRNA degradation, SMG5, SMG6 and SMG7 trigger the dephosphorylation of Upf1 by protein phosphatase 2A (PP2A), thereby terminating the reaction and initiating the disassembly of the surveillance complex (Anders et al., 2003).

Originally considered as a cellular surveillance mechanism that prevents the generation of faulty and possibly toxic gene products, it has become more and more evident that the NMD machinery also targets endogenous mRNA. Around 5-10% of naturally occurring transcripts are regulated by NMD (Mendell et al., 2004). Among the endogenous NMD targets are genes that require very fast and limited bursts in protein expression. A prominent example is *Arc*, an immediate early gene involved in neuronal plasticity and consolidation of long-term memory (Giorgi et al., 2007). The *Arc* mRNA contains two introns in its 3' UTR, making it susceptible to translation-dependent NMD. The mRNA is stable as long as it remains translationally quiescent, but is rapidly degraded upon translation activation. This ensures a very tight control of gene expression, probably generating only one protein per transcript (Giorgi et al., 2007).

A different example for NMD-controlled gene expression is an auto-regulatory feedback-loop implemented by several splicing regulators (Lareau et al., 2007; Ni et al., 2007). Members of the SR-family of splicing regulators (e.g. SRSF2 (SC35) and SRSF3), promote the alternative splicing of their own transcripts, leading the generation of NMD-sensitive splicing variants (Sureau et al., 2001). This 'unproductive splicing'

tightly couples the amount of newly translated protein to the abundance of the respective SR protein within the cell.

### (e) EJCs function in splicing regulation

Recently it was discovered that the EJC is not only an important regulator of post-splicing mRNA metabolism, but can also influence the splicing process itself (Ashton-Beaucage et al., 2010; Haremaki and Weinstein, 2012; Michelle et al., 2012; Roignant and Treisman, 2010). The first indication that EJCs are involved in pre-mRNA splicing came from two studies in *Drosophila* that identified EJC components as regulators of MAPK (mitogen-activated protein kinase) signaling (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). They demonstrated that eIF4A3, Mago and Tsunagi (the orthologs of human eIF4A3, MAGOH and Y14) are required for splicing of *mapk* transcripts in *Drosophila*. Interestingly, *mapk* transcripts contain particularly long introns and whole transcriptome analyses revealed that other long-intron containing genes are susceptible to EJC-depletion in *Drosophila*. This finding suggested a general role for EJC proteins during the excision of long introns in *Drosophila*, presumably through a mechanism known as exon definition (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). Furthermore, eIF4A3 is required for the splicing of the ryanodine receptor mRNA during early embryonic development in *Xenopus laevis* (Haremaki and Weinstein, 2012). And finally, EJC proteins have been linked to alternative splicing of a subset of human pre-mRNAs (Michelle et al., 2012). Depletion of both core (eIF4A3, Y14) and auxiliary (RNPS1, Acinus, SAP18) EJC proteins shifted the splicing pattern of the apoptotic regulator Bcl-x to the proapoptotic isoform Bcl-x<sub>s</sub> (Michelle et al., 2012). Interestingly, the function of EJC components during splicing regulation of Bcl-x depended on *cis*-acting elements, suggesting that the splicing-function of the EJC differs from its canonical role (Michelle et al., 2012). Whether EJC proteins regulate pre-mRNA splicing of other human transcripts is still unknown. In contrast to *Drosophila*, human transcripts contain many long introns and undergo numerous alternative splicing events (Kim et al., 2007; McGuire et al., 2008). Of note, several known splicing regulators (e.g. RNPS1, SAP18, SRm160, SR proteins) interact with the EJC (Mayeda et al., 1999; Schwerk et al., 2003; Singh et al., 2012; Singh et al.,

2010). It is therefore conceivable that the role of the EJC during pre-mRNA splicing in human cells goes far beyond the control of the apoptotic regulator Bcl-x.

#### **2.3.4 EJC-independent functions of eIF4A3**

EIF4A3 is conserved from yeast to humans, even though yeast cells lack additional EJC proteins and NMD in yeast is independent of pre-mRNA splicing (Gonzalez et al., 2001). This suggests that a different function of eIF4A3 preceded its role in the EJC. And indeed, Fal1p, the yeast homolog of eIF4A3, is involved in ribosomal RNA (rRNA) processing. Depletion of Fal1p leads to reduced 18S rRNA levels and a deficiency in 40S ribosomal subunits (Kressler et al., 1997). The function of Fal1p/eIF4A3 during rRNA biogenesis is conserved from yeast to humans. Both yeast Fal1p and human eIF4A3 require the interaction with an additional protein to function in rRNA processing. This protein is called Sgd1p in yeast and NOM1 in humans and interacts with Fal1p/eIF4A3 via a conserved MIF4G domain (Alexandrov et al., 2011).

A different EJC-independent function of eIF4A3 is the regulation of selenoprotein-synthesis. Selenium is incorporated into proteins as selenocysteine through a specific recoding of the UGA stop codon. The recoding mechanism requires stem loop structures in the 3'UTR of the selenoprotein mRNA (SECIS elements) as well as certain *trans* acting proteins that bind to the SECIS element (Seeher et al., 2012). During selenium deficiency, some selenocysteine-containing proteins are preferentially produced at the expense of others. EIF4A3 is involved in the differential synthesis of selenoproteins through directly binding to SECIS elements (Budiman et al., 2009). EIF4A3-binding prevents the recruitment of other *trans*-acting factors to the mRNA and thereby inhibits selenocysteine incorporation. Strikingly, eIF4A3 only binds to a subset of SECIS elements, which is in stark contrast to the sequence-independent RNA interaction observed for the EJC (Budiman et al., 2011). Of note, in the absence of UGA-recoding, the codon is recognized as a PTC, leading to the mRNA degradation through NMD (Moriarty et al., 1998). Thus, EJC-independent and EJC-dependent functions of eIF4A3 act in concert to regulate the expression of selenoproteins during selenium deficiency.

The diverse functions of eIF4A3 described in this chapter all rely on its activity as an ATP-dependent RNA-binding protein. This activity is carried out by the conserved helicase core of eIF4A3, which is shared with many other proteins of the DExD/H-box family. The characteristics of DExD/H-box proteins are further discussed in the following chapter.

## **2.4 DExD/H-box helicases are key regulators of mRNA metabolism**

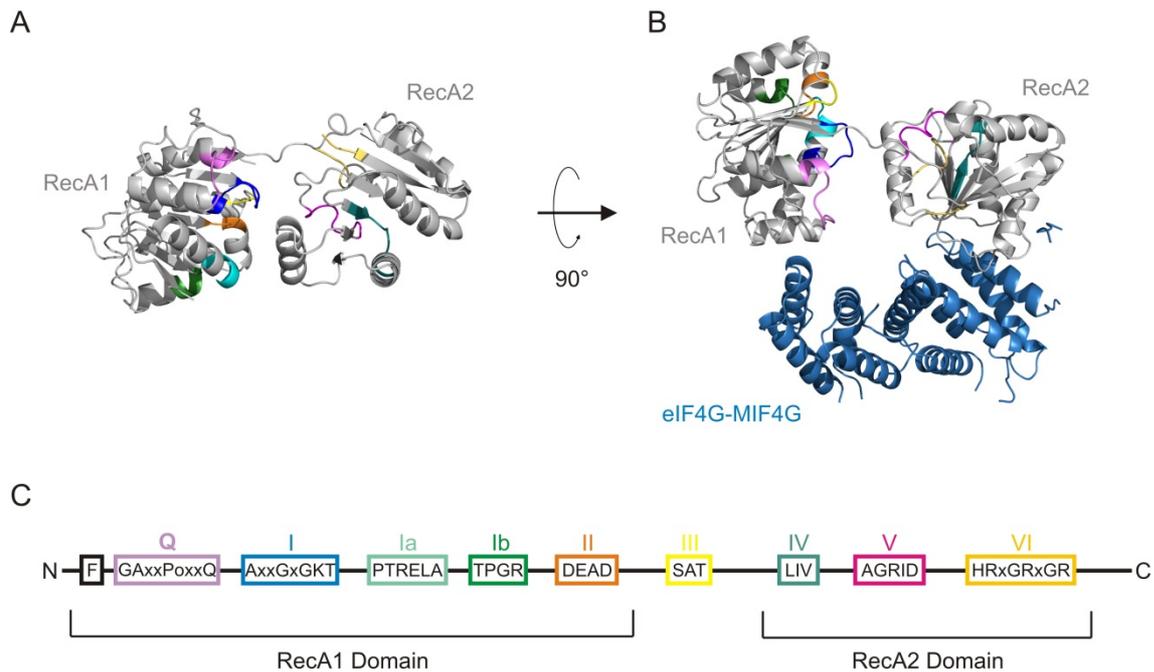
Most aspects of mRNA metabolism involve extensive remodeling of RNA-RNA and RNA-protein interactions. These structural rearrangements are usually catalyzed by a class of enzymes referred to as RNA helicases. Helicases are defined by their ability to unwind DNA or RNA duplexes at the expense of NTP hydrolysis. According to structural similarities, helicases can be grouped into 5 superfamilies (SF1-5). SF3, SF4 and SF5 helicases are multimeric enzymes that are predominantly found in bacteria and viruses (Patel and Picha, 2000). In contrast, SF1 and SF2 helicases function as monomers or dimers and contain an overall similar structure (Tuteja and Tuteja, 2004a, b). Most RNA helicases belong to superfamily 2 (SF2) and are further categorized as DEAD-box, DEAH and DExH helicases (collectively referred to as DExD/H-box helicases) depending on certain conserved sequence elements along their helicase core (Fairman-Williams et al., 2010; Jankowsky and Fairman, 2007; Linder et al., 1989). The conserved structure of DExD/H-box helicases and their function during mRNA metabolism will be further discussed in chapter 2.4.1 and 0, respectively.

### **2.4.1 Structure of DExD/H-box helicases**

DExD/H-box helicases are conserved across all phyla and involved in virtually all aspects of RNA metabolism (Cordin et al., 2006). They share a conserved helicase core of around 350-400 residues, consisting of two globular RecA-like domains (the N-terminal RecA1 domain and C-terminal RecA2 domain) that are connected by a flexible hinge region (Figure 5A (Caruthers and McKay, 2002)). RecA is a bacterial DNA-binding protein involved in homologous recombination. The RecA-like domain consists of 5  $\beta$ -strands surrounded by 5  $\alpha$ -helices (Story and Steitz, 1992). In the absence of ATP and RNA, the relative orientation of the two domains is not well defined. This variability is

caused by the flexible interdomain linker and is probably important for the enzymatic activity of the protein. Upon binding of ATP and RNA, the two domains are brought in close proximity to each other and form a deep interdomain cleft (Bono et al., 2006; Sengoku et al., 2006). The nucleotide binds inside the cleft and the RNA interacts with both domains at the base of the helicase core, almost perpendicular to the cleft. This ATP- and RNA-bound state is often referred to as the 'closed' conformation of the helicase (as opposed to the 'open' conformation in the absence of ATP and RNA).

Several short, highly conserved sequence motifs line the interdomain cleft and coordinate the binding of RNA and ATP (I, Ia, Ib, II, III, IV, V, VI) (Figure 5, (Cordin et al., 2006)). Among the conserved motifs are the well-defined Walker A and Walker B motifs required for ATPase activity (Walker et al., 1982). The amino acid sequence of the Walker B motif determines the helicase subclass (e.g. Asp-Glu-Ala-Asp (D-E-A-D) for DEAD-box helicases). Whereas motif I (Walker A), motif II (Walker B) and motif VI regulate ATP-binding and hydrolysis, motif III (S-A-T) couples ATP hydrolysis to RNA unwinding and motif Ia, Ib, IV and V bind the RNA substrate (mostly through interaction with the sugar-phosphate backbone) (Cordin et al., 2006). Interactions between the different motifs ensure that nucleotide hydrolysis is coupled to changes in RNA affinity, thus connecting ATPase and helicase activity. Additional motifs (e.g. the Q motif and the GG motif) are present in some but not all DExD/H-box helicases and further fine-tune the regulatory network (Cordin et al., 2004; Linder and Jankowsky, 2011; Tanner et al., 2003). Even though DExD/H-box proteins are classified as helicases, most members of this group display very weak strand separation activity *in vitro*. In contrast to the highly processive duplex unwinding activity of many DNA helicases, DExD/H-box proteins rarely unwind duplexes exceeding two helical turns (Jankowsky and Fairman, 2007). However, this lack of processivity is not likely to limit the activity of DExD/H-box proteins *in vivo*, since most cellular substrates are short RNA duplexes, such as rRNA-snRNA interactions during ribosomal biogenesis or snRNA-snRNA and snRNA-mRNA interactions during pre-mRNA splicing (Yang and Jankowsky, 2006). Hence, DExD/H-box proteins might be more appropriately considered as ATP-dependent RNA-binding proteins (Jankowsky and Fairman, 2007). How these RNA-binding proteins can shape the gene regulatory network of a cell will be discussed in the following chapter.



**Figure 5: Structural features of DEAD-box helicases.** (A-B) Crystal structure of the DEAD-box helicase eIF4A as part of the eIF4A/eIF4G complex (PDB file 2VSO). The conserved sequence motifs that line the interdomain cleft are color-coded as depicted in (C).

#### 2.4.2 DExD/H-box proteins function as components of large RNP complexes

DExD/H-box proteins regulate processes as diverse as mRNA transcription, pre-RNA splicing, nucleocytoplasmic export, ribosome biogenesis, translation initiation and mRNA degradation (Silverman et al., 2003). To catalyze the remodeling of large RNP complexes, they unwind short RNA duplexes and resolve RNA secondary structures, but also act as RNPsases in displacing proteins from mRNA or as stable mRNA clamps, as exemplified by the function of eIF4A3 within the EJC (see chapter 2.3).

Interestingly, the ATPase activity of most isolated DExD/H-box proteins is rather weak (Ballut et al., 2005; Li et al., 1999). Considering the large number of cellular DExD/H-box proteins and the high conservation of the helicase core, it seems reasonable that the activity of DExD/H-box proteins is tightly regulated. Thus, most DExD/H-box proteins do not function alone, but in the context of larger RNP complexes. The interaction with both stimulatory and inhibitory proteins ensures the spatial and temporal control of ATPase and helicase activity (Silverman et al., 2003).

For example, the weak ATPase activity of eIF4A is stimulated through interactions with the large scaffolding protein eIF4G. Together with the cap-binding protein eIF4E, they form the translation initiation complex eIF4F. Within this complex, the helicase activity of eIF4A is restricted to the cap-proximal 5' UTR of mRNAs, where it causes melting of secondary structures and thereby facilitates scanning of the small ribosomal subunit (Svitkin et al., 2001). Another example of regulated ATPase activity is the mRNA export factor Dbp5 (Ddx19 in humans). Dbp5 confers directionality to mRNA export by removing the export receptor Mex67:Mtr2 (NXF1:p15 in humans) from mRNA. The RNPase activity of Dbp5 is restricted to the cytoplasmic face of the nuclear pore complex (NPC) through interaction with the regulatory protein Gle1. Gle1 interacts with cytoplasmic nucleoporins and locally activates the ATPase activity of Dbp5 (Hodge et al., 2011; Montpetit et al., 2011; Noble et al., 2011). Thus, Dbp5 only remodels mRNPs that have exited the nuclear pore.

Interactions with regulatory proteins are usually mediated through highly variable N- and C-terminal domains that flank the conserved helicase core of most DExD/H-box proteins. However, not all helicases possess N- and C-terminal extensions to regulate their activity. Among the 'minimal' RNA helicases are the aforementioned eIF4A and Dbp5. To regulate the ATPase activity of eIF4A and Dbp5, interacting proteins need to directly bind to the conserved helicase core. Interestingly, both eIF4G and Gle1 use similar  $\alpha$ -helical structures to interact with eIF4A and Dbp5, respectively (Montpetit et al., 2011). In eIF4G this domain is called MIF4G (middle domain of initiation factor 4G). MIF4G domains are found in several other proteins that bind DExD/H-box helicases. Their conserved structure and function will be further introduced in the following chapter.

### **2.4.3 The MIF4G-domain has evolved as a helicase interaction domain**

The MIF4G domain was named after the translation initiation factor eIF4G, where it was first described. It belongs to the HEAT repeat protein family. The HEAT domain is a solenoid protein domain named after 4 HEAT repeat containing proteins (Huntingtin, elongation factor 3 (EF3), protein phosphatase 2a (PP2A), yeast kinase TOR1). It consists of tandem repeats of antiparallel  $\alpha$ -helices (the HEAT repeats) linked by flexible loops and arranged around a common axis to form a rod-like structure

(Andrade et al., 2001). HEAT repeat domains are often involved in protein-protein interactions.

The MIF4G domain of eIF4G contains 10  $\alpha$ -helices that form a right-handed solenoid. The convex surface of the MIF4G domain interacts with both RecA-like domains of eIF4A. Whereas the N-terminal helices of the MIF4G domain bind to the RecA2 domain of the helicase, the C-terminal helices contact the RecA1 domain (Schutz et al., 2008). This interaction holds eIF4A in an extended conformation. In this conformation, both RecA-like domains are in close proximity and the conserved DExD/H-box motifs face towards the inner side of the helicase core. Thus, the extended helicase is poised for ATP and RNA binding, explaining the stimulatory effect of eIF4G on the ATPase activity of eIF4A. Interestingly, the affinity of eIF4A for eIF4G is mostly determined by the interaction of the RecA2 domain with the N-terminal part of the MIF4G domain (Schutz et al., 2008). The weaker interactions between RecA1 and the C-terminus of MIF4G are important to maintain the extended conformation, but can be easily resolved when the helicase adopts its closed conformation upon ATP and RNA binding (Schutz et al., 2008). Apart from the MIF4G domain, eukaryotic eIF4G contains two more HEAT repeat domains, a MA3 domain and a W2 domain. The closely related MA3 domain was shown to also contribute to eIF4A binding of eIF4G. Interestingly, the tumor suppressor Pdc4 (programmed cell death 4) also contains two MA3 domains and competes with eIF4G for eIF4A binding. But whereas eIF4G stimulates the ATPase activity of eIF4A, Pdc4 traps the helicase in an inhibited conformation and thereby blocks translation initiation (Loh et al., 2009). When bound to Pdc4, eIF4A adopts an open conformation with several of its DEAD-box motifs facing the solvent rather than the interdomain cleft (Loh et al., 2009). This example demonstrates that similar interaction domains can stabilize DExD/H-box helicases in activated or inhibited conformations. Very subtle differences in the binding motifs determine whether the interaction has a stimulatory or an inhibitory effect. This high degree of specificity is also reflected in the choice of binding partners. For example eIF4A3, a very close homolog of eIF4A does not bind the MA3 domain of eIF4G (Li et al., 1999).

Several other proteins have been described to bind DExD/H-box helicases through a very similar mechanism. The export mediator Gle1 comprises an all  $\alpha$ -helical HEAT

repeat protein that binds the DEAD-box helicase Dbp5 (Montpetit et al., 2011). Analogous to the interaction between eIF4G and eIF4A, the stable interaction between both proteins primarily involves the RecA2 domain of Dbp5, whereas secondary transient interactions between RecA1 and Gle1 regulate the activity of the helicase (Montpetit et al., 2011). A unique feature of the interaction between Gle1 and Dbp5 is the requirement for the small molecule inositol hexakisphosphate (IP<sub>6</sub>), which bridges the protein-protein interaction (Montpetit et al., 2011). Because Gle1 stabilizes Dbp5 in a partially open conformation it was suggested that the interaction stimulates RNA release and enzyme recycling (Montpetit et al., 2011).

A recent study identified yet another protein interaction that is mediated via a similar mechanism. NOM1 binds the DEAD-box helicase eIF4A3 through its MIF4G-domain (Alexandrov et al., 2011). Mutational analyses suggested that the interaction between both proteins resembles the interaction between eIF4G and eIF4A. Even though biochemical and structural studies of the NOM1-eIF4A3 interaction are still pending, the genetic interaction between both proteins during rRNA biogenesis in yeast and humans implied that NOM1 is an evolutionary conserved regulator of the DEAD-box protein eIF4A3 (Alexandrov et al., 2011).

The introductory section of this thesis focused on the function of RNA-binding protein complexes during eukaryotic gene expression; starting from the general role of RBPs in mRNA metabolism and ending with the particular example of the EJC as a splicing dependent protein complex. Special emphasis was placed on the role of the DEAD-box helicase eIF4A3 within the EJC. To better understand this protein, the general structural and functional characteristics of DExD/H-box helicases were discussed in more detail. The section ended with the description of the MIF4G domain and related structures and their role in regulating the enzymatic activity of DExD/H-box helicases. In the experimental part of this thesis, I will now describe the discovery and characterization of another MIF4G-domain containing protein that interacts with eIF4A3. The results presented in the following chapters will shed light on the relationship between the EJC and the splicing machinery and further expand our knowledge of the RNP interaction network.



### 3 Aims and Significance of the project

The EJC is a key regulator of mRNA metabolism that assembles on mRNA during splicing. It binds the mRNA in a sequence-independent manner 20-24 nt upstream of exon-exon junctions. The description of the EJC crystal structure and subsequent biochemical studies have led to a comprehensive model for the stepwise assembly of the EJC core on the mRNA. Although it is well established that EJC assembly is a splicing-dependent process, the factors that mediate the interaction between the EJC and the splicing machinery are still unknown.

The aim of this study is to establish a physical link between the spliceosome and the EJC. To this end, I seek to identify a direct spliceosomal interaction partner of eIF4A3, the DEAD box helicase that anchors the complex on the mRNA. A very promising candidate for this function is the spliceosomal protein CWC22. CWC22 displays significant sequence identity with the recently identified eIF4A3 binding partner NOM1. Moreover, CWC22 features a MIF4G and a MA3 domain, both of which have been previously shown to bind DEAD-box helicases.

The initial goal of this study is to verify the interaction between CWC22 and eIF4A3 in mammalian cells. Furthermore, the direct interaction between both proteins will be investigated using purified proteins *in vitro*. Subsequently, the interaction site will be further characterized through deletion mutants and site-directed mutagenesis. The overall aim is to identify CWC22 binding-deficient point mutants of eIF4A3 that can be employed to analyze the influence of CWC22 on EJC assembly during pre-mRNA splicing. To complement the *in vitro* findings, siRNA mediated knockdown experiments will further clarify the function of CWC22 in living cells.

Based on these experiments, I intend to generate a model for the molecular mechanism of EJC assembly in human cells, which can explain the splicing-dependent nature of this process.



## 4 Results

The aim of this work was the identification of a direct link between the EJC and the spliceosome. For structural reasons, the *Results* section is divided into three main parts. Section 4.1 describes the identification of CWC22 as a direct spliceosomal interaction partner of the EJC core protein eIF4A3. Functional data presented in section 4.2 demonstrate that the interaction between CWC22 and eIF4A3 is crucial for splicing-dependent EJC assembly. The role of CWC22 in pre-mRNA splicing is analyzed in section 4.3, concluding with the observation that pre-mRNA splicing and EJC assembly are two distinct functions of CWC22 that can be experimentally uncoupled.

### 4.1 CWC22 is a direct interaction partner of eIF4A3

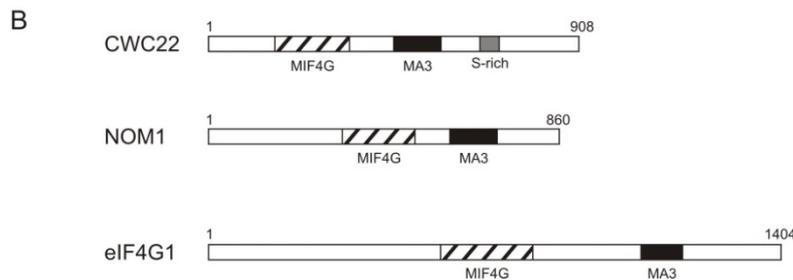
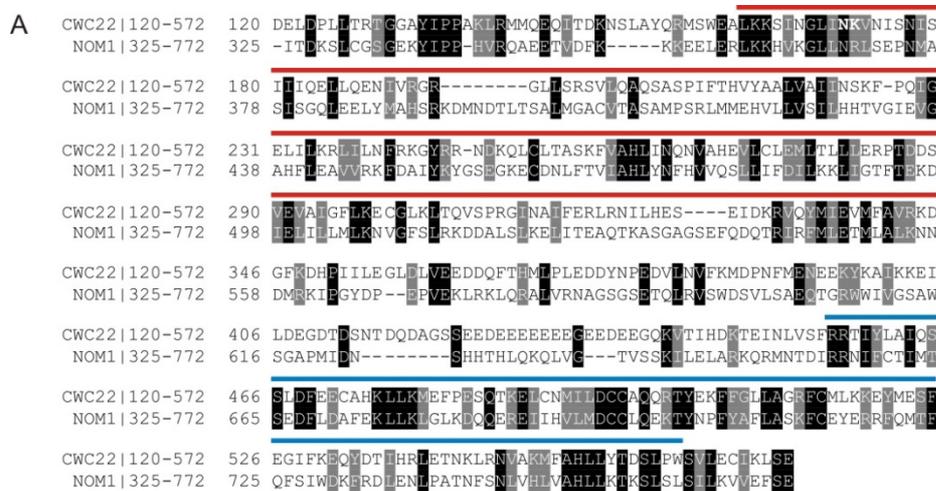
The nuclear EJC consists of the three proteins eIF4A3, MAGOH and Y14 that assemble on pre-mRNA during splicing (Gehring et al., 2009a). Previous work has shown that the interaction of eIF4A3 with the early spliceosome is independent of MAGOH and Y14, leading to the hypothesis that it is the first EJC protein to engage the splicing machinery (Gehring et al., 2009a). As a DEAD-box helicase, eIF4A3 binds directly to mRNA, thus forming an assembly platform for the other EJC factors (Le Hir et al., 2001). This distinguished role during EJC assembly prompted me to search for a spliceosomal interaction partner of eIF4A3 that would explain the splicing-dependent nature of the assembly process.

#### 4.1.1 An *in silico* search identifies CWC22 as a putative eIF4A3 binding protein

In order to identify putative new interaction partners of eIF4A3, a BLAST (Basic Local Alignment Search Tool) search against the non-redundant human genome database was performed to find spliceosomal factors with homology to any of the known eIF4A3-binding proteins. Interestingly, NOM1, a protein that interacts with yeast and human eIF4A3 during rRNA biogenesis (Alexandrov et al., 2011), shared significant sequence identity with the spliceosomal protein Complexed With Cef1 (also known as KIAA1604, Figure 6A). The more detailed analysis of the alignment revealed that the

## Results

sequence conservation is restricted to the central part of both proteins (amino acid residues 325-772 of NOM1 aligned with amino acid residues 120-572 of CWC22).



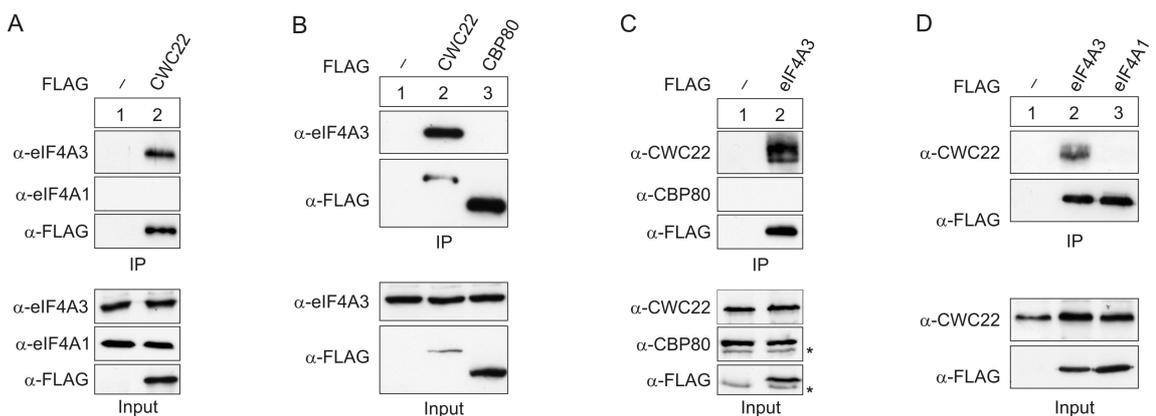
**Figure 6: CWC22 is a homolog of NOM1.** (A) Alignment of the protein sequences of human CWC22 (Q9HCG8) and human NOM1 (Q5C9Z4). The alignment was generated with ClustalW and boxshade 3.21. The MIF4G domain is indicated by the red line and the MA3 domain by the blue line. (B) Domain structure of CWC22, NOM1 and eIF4G1.

Interestingly, structural predictions (using SMART (Schultz et al., 1998)) revealed that both proteins harbor a MIF4G domain (middle domain of translation initiation factor 4G) (approximately between residues 362-559 of NOM1 and 163-346 of CWC22) followed by an MA3 domain (between residues 655-761 of NOM1 and 455-561 of CWC22), giving them an overall similar domain structure (Figure 6B). Both MIF4G and MA3 domain belong to the family of  $\alpha$ -helical HEAT repeat domains and are often involved in binding to DExD/H-box helicase. They were first described in the translation initiation factor eIF4G, where they facilitate the interaction with the RNA helicase eIF4A (Schutz et al., 2008). Moreover, similar domains are present in Gle1 and the apoptotic regulator Pcd4 (programmed cell death 4), where they mediate the

interaction with the nuclear export factor Dbp5 and the translation initiation factor eIF4A1, respectively (Dossani et al., 2009; Yang et al., 2004).

#### 4.1.2 CWC22 interacts with eIF4A3 in cell extracts

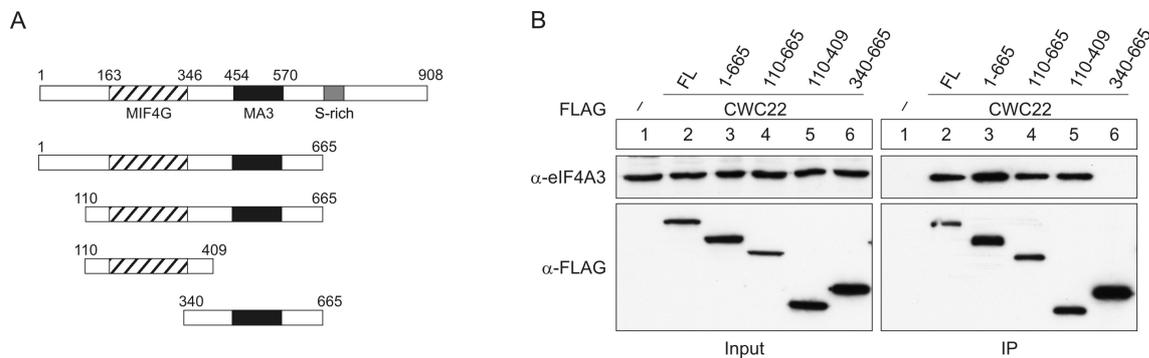
To test whether CWC22 is a true interaction partner of eIF4A3, FLAG-tagged CWC22 was cloned into a eukaryotic expression vector and expressed in HeLa cells. The interaction with endogenous eIF4A3 was analyzed by co-immunoprecipitation (Co-IP) with anti-FLAG beads followed by western blotting (Figure 7). Both RNase A and RNase I were included in the IP-buffer to prevent unspecific interactions through RNA-bridging. FLAG-CWC22 co-precipitated endogenous eIF4A3 in an RNase-insensitive manner but failed to bind the closely related protein eIF4A1 (Figure 7A). To verify the specificity of the assay, the interaction between FLAG-CBP80 and eIF4A3 was tested. CBP80 is a component of the cap-binding complex and contains several MIF4G-like domains, but unlike FLAG-CWC22 it did not co-precipitate endogenous eIF4A3 (Figure 7B). To further confirm the interaction between eIF4A3 and CWC22, FLAG-eIF4A3 was expressed in HeLa cells and the interaction with endogenous CWC22 was analyzed by co-immunoprecipitation with anti-FLAG beads. Indeed, FLAG-eIF4A3 co-precipitated endogenous CWC22 but not CBP80 (Figure 7C), whereas the negative control FLAG-eIF4A1 failed to interact with CWC22 (Figure 7D). Taken together, these results show that CWC22 interacts with eIF4A3 in a highly specific manner.



**Figure 7: CWC22 interacts with eIF4A3.** FLAG-immunoprecipitation from cell lysates treated with RNase A (A, C) or RNase I (B, D). (A-B) FLAG-CWC22 and FLAG-CBP-80 were immunoprecipitated and co-precipitated endogenous eIF4A3 and eIF4A1 were detected by immunostaining with specific antibodies. (C-D) FLAG-eIF43 and FLAG-eIF4A1 were immunoprecipitated and co-precipitated endogenous CWC22 and CBP80 were detected with specific antibodies. Unfused FLAG served as a negative control. 10% of the lysate was loaded as input.

#### 4.1.3 eIF4A3 binds the MIF4G domain of CWC22

The role of MIF4G and MA3 domains in binding to DExD/H-box helicases suggested that either one or both domains mediate the interaction between CWC22 and eIF4A3. To test this hypothesis, several truncations of FLAG-CWC22 were tested in Co-IP experiments for their ability to interact with endogenous eIF4A3 (Figure 8A). As expected, deletion of either C-terminal (CWC22(1-665)) or both C- and N-terminal domains of the protein (CWC22(110-665)) had no effect on eIF4A3 precipitation (Figure 8B, lane 3 and 4). Interestingly, the MA3 domain could also be deleted (CWC22(110-409)) without compromising eIF4A3 binding (Figure 8B, lane 5). In contrast, a CWC22 deletion mutant that lacked the MIF4G domain (CWC22(340-665)) failed to interact with eIF4A3 (Figure 8B, lane 6). Hence, the minimal eIF4A3-interaction domain comprised of amino acid residues 110-409 (Figure 8B, lane 5), indicating that the MIF4G domain of CWC22 is both required and sufficient for the interaction with eIF4A3.



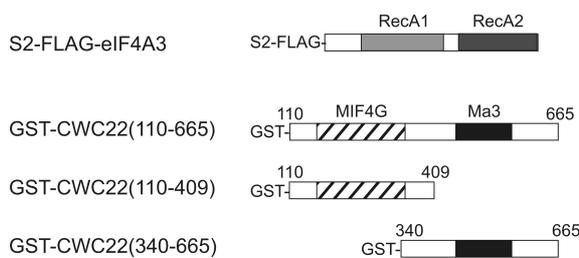
**Figure 8: eIF4A3 interacts with the MIF4G domain of CWC22.** (A) Schematic representation of truncated CWC22 used for FLAG-immunoprecipitation. (B) FLAG-tagged CWC22 was immunoprecipitated from RNase A treated cell lysates and co-precipitation of endogenous eIF4A3 was detected by immunostaining with a specific antibody. Unfused FLAG served as a negative control. 10% of the lysate was loaded as input.

#### 4.1.4 Pulldown experiments confirm a direct interaction of eIF4A3 and CWC22

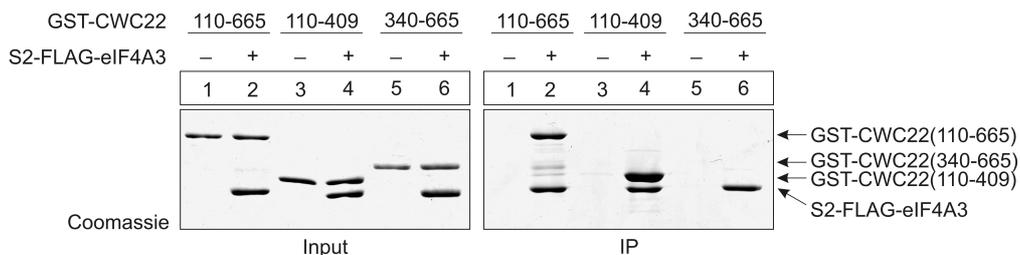
All interactions described until now were analyzed in cell lysates. Even though RNase was included in the assay buffer to prevent non-specific RNA-bridging, this does not exclude the possibility that other proteins in the cell lysate mediated the interaction between CWC22 and eIF4A3. To test whether CWC22 directly binds to eIF4A3, both proteins were purified from *E.coli* and analyzed through *in vitro* pulldown experiments

(Figure 9). Full-length eIF4A3 with an N-terminal Strep (S2) and N-terminal FLAG-tag (S2-FLAG-eIF4A3) was purified from bacterial cultures on streptactin resin (Figure 9A). CWC22 was expressed as a GST-fusion protein and purified via Glutathion-Sepharose. Since full-length CWC22 was poorly expressed in *E.coli*, three truncated versions of CWC22 were used, GST-CWC22(110-665), comprising both MIF4G and MA3 domain, GST-CWC22(110-409), which contains only the MIF4G domain and GST-CWC22(340-665), which contains only the MA3 domain (Figure 9A). To analyze complex formation, purified S2-FLAG-eIF4A3 was pre-incubated with each of the CWC22 variants, followed by FLAG-immunoprecipitation. The precipitated proteins were resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by coomassie staining (Figure 9B).

A



B



**Figure 9: EIF4A3 directly interacts with the MIF4G domain of CWC22.** (A) Schematic representation of the recombinant proteins expressed in *E.coli*. (B) *In vitro* pulldown experiment. S2-FLAG-eIF4A3 was immobilized on anti-FLAG beads and incubated with CWC22 pieces. Co-precipitated proteins were resolved by SDS-PAGE and stained with coomassie. Samples without S2-FLAG-eIF4A3 served as a negative control. 10% of the assembly reaction was loaded as input.

In line with the previously described Co-IP experiments, FLAG-tagged eIF4A3 co-precipitated GST-CWC22(110-665) and GST-CWC22(110-409) but failed to bind GST-CWC22(340-665), demonstrating that the MIF4G domain of CWC22 is required for

direct binding to eIF4A3. The MA3 domain did not contribute to eIF4A3 binding, and the construct lacking the MA3 domain (CWC22(110-409)) even exhibited slightly better co-precipitation efficiencies (Figure 9B, compare lane 2 and lane 4).

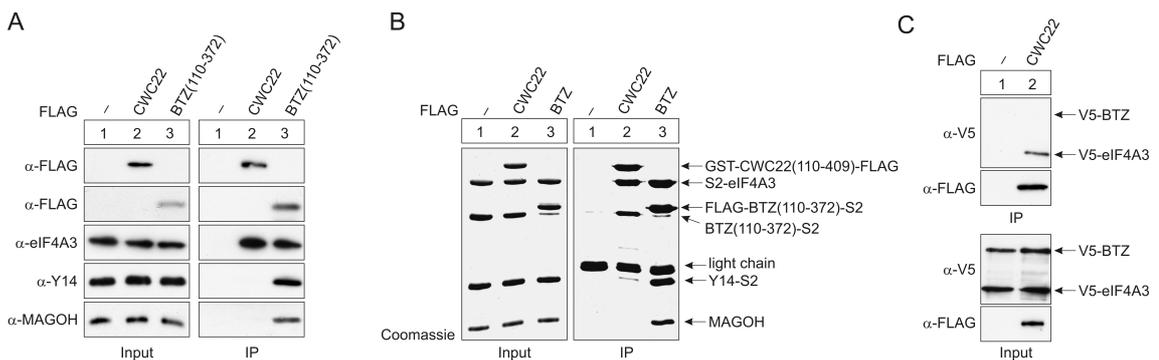
#### 4.1.5 CWC22 does not interact with the other EJC proteins

In living cells, eIF4A3 binds stably to Y14, MAGOH and Barentsz (BTZ) to form the core of the EJC. With regard to its role in complex formation, it is important to understand whether CWC22 binds eIF4A3 alone or as part of the EJC. In order to address this question, FLAG-CWC22 and FLAG-BTZ were expressed in HeLa cells and Co-IP experiments were performed as described in paragraph 4.1.2. As expected, FLAG-BTZ co-precipitated the endogenous EJC proteins eIF4A3, Y14 and MAGOH from cell lysates, demonstrating that an EJC was formed under these conditions (Figure 10A, lane 3). In contrast, FLAG-CWC22 co-precipitated comparable amounts of eIF4A3, but failed to interact with Y14 and MAGOH, indicating that CWC22 only interacts with eIF4A3 but not with the assembled EJC (Figure 10A, lane 2).

To further confirm these observations, the interaction between purified GST-CWC22(110-409)-FLAG and a recombinant EJC was analyzed *in vitro*. To this end, the EJC proteins eIF4A3, Y14, MAGOH and the central domain of BTZ were expressed in bacterial lysates and affinity-purified on streptactin resin. Since Y14 and MAGOH form a stable heterodimer, they can be co-expressed in bacteria and un-tagged MAGOH is co-purified together with C-terminally Strep-tagged Y14 (Y14-S2). The BTZ construct used in this study comprised of the SELOR (speckle localizer and RNA binding module) domain and was expressed with C-terminal Strep-tag and either with or without a N-terminal FLAG-tag (BTZ(110-372)-S2 and FLAG-BTZ(110-372)-S2). Previous studies have shown that the SELOR domain is sufficient to bind eIF4A3 and the EJC (Degot et al., 2004). Furthermore, the SELOR domain is required *in vitro* to stabilize the interaction between eIF4A3, Y14 and MAGOH (Ballut et al., 2005).

To form the recombinant EJC eIF4A3, Y14, MAGOH and BTZ(110-372) were pre-incubated together with poly(U) RNA and the non-hydrolysable ATP analog ADPNP in the presence (Figure 10B, lane 2) or absence (Figure 10B, lane 1 and 3) of CWC22(110-409). The complexes were then precipitated with immobilized FLAG-tagged CWC22(110-409) (Figure 10B, lane 2) or FLAG-tagged BTZ(110-372) (Figure 10B, lane

3), respectively. As a negative control, no FLAG-tagged protein was included in lane 1 to exclude unspecific interactions with the beads used for immunoprecipitation. The precipitated protein complexes were resolved by SDS-PAGE and visualized with coomassie. In line with published data (Ballut et al., 2005), FLAG-BTZ(110-372)-S2 co-precipitated eIF4A3, Y14 and MAGOH, proving that the recombinant proteins can form an EJC *in vitro* (Figure 10B, lane 3). GST-CWC22(110-409)-FLAG on the other hand did interact with eIF4A3 and BTZ(110-372), but not with Y14 and MAGOH (Figure 10B, lane 2). These results corroborate previous Co-IP experiments and suggest that the inability of CWC22 to interact with EJC-bound eIF4A3 stems from a general inaccessibility of the binding site within the complex.



**Figure 10: CWC22 interacts with eIF4A3 but not with Y14 and MAGOH.** (A) Immunoprecipitation of FLAG-CWC22 and FLAG-BTZ from RNase A treated cell lysates. Co-precipitated endogenous eIF4A3, Y14 and MAGOH were detected by staining with specific antibodies. Unfused FLAG served as a negative control. 10% of the lysate were loaded as input. (B) Interaction between recombinant CWC22 and EJC proteins *in vitro*. S2-eIF4A3, BTZ(110-372)-S2, Y14-S2 and MAGOH were incubated under EJC assembly conditions and the complexes were immunoprecipitated with immobilized GST-CWC22(110-409)-FLAG (lane 2) or FLAG-BTZ(110-372)-S2 (lane 3). BTZ(110-372)-S2 without a FLAG tag served as a negative control (lane 1). Protein complexes were resolved by SDS-PAGE and visualized with coomassie. 10% of the assembly reaction was loaded as input. (C) Immunoprecipitation of FLAG-CWC22 from RNase A treated cell lysates. Co-precipitation of V5-eIF4A3 and V5-BTZ was detected by immunostaining with a V5-specific antibody. Unfused FLAG served as a negative control. 10% of the cell lysate was loaded as input.

Interestingly, GST-CWC22(110-409)-FLAG co-precipitated BTZ(110-372) together with eIF4A3, demonstrating that simultaneous binding of eIF4A3 to CWC22 and BTZ is possible. Since eIF4A3 can bind BTZ in the absence of Y14 and MAGOH (Andersen et

al., 2006; Bono et al., 2006), it is conceivable that such a CWC22/eIF4A3/BTZ complex could also form in living cells. To test this hypothesis, the interaction between FLAG-CWC22 and BTZ was analyzed through Co-IP experiments (Figure 10C). Because no good antibody against endogenous BTZ was available, V5-tagged BTZ and V5-tagged eIF4A3 were co-expressed with FLAG-CWC22 in HeLa cells. Surprisingly, FLAG-CWC22 co-precipitated V5-eIF4A3 but failed to interact with V5-BTZ. The discrepancy between Co-IP and *in vitro* pulldown experiments could be explained by the fact that CWC22 is a nuclear protein whereas full length BTZ resides predominantly in the cytoplasm. Hence, even though CWC22, eIF4A3 and BTZ can form a complex *in vitro*, this complex is most likely not formed in living cells.

#### 4.1.6 CWC22 is no component of the post-splicing mRNP

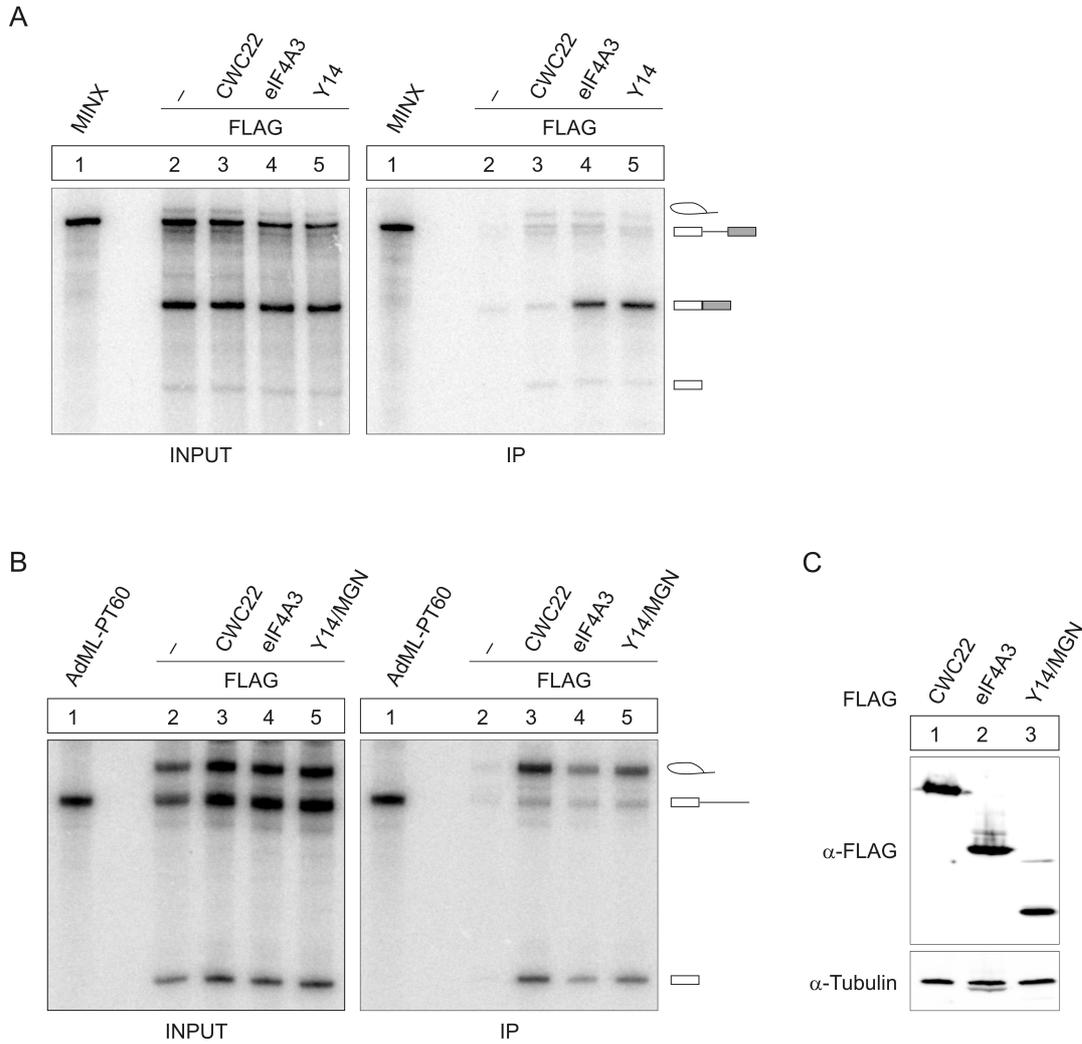
Mass spectrometric studies revealed that CWC22 is part of the human spliceosome (Agafonov et al., 2011). Moreover, the yeast homolog of CWC22 is involved in the first catalytic step of splicing through interaction with the DExD/H-box helicase Prp2 (Yeh et al., 2011). However, detailed information on the role of human CWC22 during pre-mRNA splicing is still missing. To shed further light on the function of CWC22 during splicing, the interaction of CWC22 with spliced mRNA was studied in an *in vitro* splicing system. This assay allows for the splicing of an *in vitro* transcribed mRNA substrate that is body-labeled with radioactive  $\alpha$ -<sup>32</sup>P-GTP in HeLa nuclear extracts (for a detailed description of the assay see *Material and Methods*). The splicing reaction is supplemented with whole cell extracts from HEK293 cell that express the FLAG-tagged protein of interest, so that the interaction between FLAG-tagged protein and spliced mRNA can be analyzed by FLAG-immunoprecipitation. After RNA extraction and denaturing PAGE, the co-precipitated mRNA can be visualized by autoradiography. This method was applied to analyze the interaction of FLAG-CWC22 with spliced mRNA and to compare it to the mRNA-interaction pattern of its binding partner eIF4A3, as well as the non-interacting EJC protein Y14 (Figure 11).

The first mRNA substrate used in this study was MINX, a derivative of the viral mRNA Adenovirus major late (AdML), that contains two exons and one intron and is very efficiently spliced *in vitro* (Figure 11A, left panel) (Bell et al., 2002; Zillmann et al., 1988). As expected, both FLAG-eIF4A3 and FLAG-Y14 specifically precipitated the

spliced mRNA, but no pre-mRNA, indicating that they have been successfully incorporated into a splicing-dependent EJC (Figure 11A, lane 4 and 5). Despite being well expressed (Figure 11C), FLAG-CWC22 bound neither pre-mRNA nor spliced mRNA, demonstrating that it is no part of the post-splicing mRNP (Figure 11A, lane 3). This is not surprising since spliceosomal subunits dissociate from the mRNA after the splicing process is completed. Indeed, splicing is a very dynamic process that requires many rearrangements, and spliceosomal proteins rarely stay attached to the spliceosome throughout the whole splicing process (Wahl et al., 2009).

#### **4.1.7 CWC22, eIF4A3 and Y14 interact with the C-complex spliceosome**

To further delineate at which step during splicing CWC22 engages the human spliceosome, a different mRNA substrate was used. AdML-PT60 is a version of AdML mRNA that lacks the second exon. Instead, AdML-PT60 has an elongated polypyrimidine tract consisting of 60 pyrimidines (C and U). This elongated polypyrimidine tract enhances the overall splicing efficiency and thereby allows the first transesterification reaction to occur even in the absence of a second exon. Without this second exon however, the splicing reaction cannot proceed and instead the spliceosome arrests in a stage known as C-complex (Bessonov et al., 2010). The mRNA bands that can be visualized on the gel after splicing of AdML-PT60 correspond to lariat mRNA (top band), unspliced AdML (middle band) and first exon (bottom band). Both FLAG-eIF4A3 and FLAG-Y14 precipitated C-complex mRNA (top and bottom band), which shows that they already interact with the spliceosome before the second catalytic step of splicing (Figure 11B, lane 4 and 5). Of note, FLAG-CWC22 precipitated comparable amounts of mRNA, marking it as a component of the spliceosomal C-complex (Figure 11B, lane 3). These results are in accordance with published data from mass spectrometric analyses and a yeast-2-hybrid screen, which recently defined CWC22 as a constituent of activated B complexes ( $B_{act}$ ) and C-complexes (Agafonov et al., 2011; Hegele et al., 2012). Moreover, the data show that CWC22 and the EJC proteins eIF4A3 and Y14 interact with the spliceosome at the same time, which is a prerequisite for a possible role of CWC22 during EJC assembly.

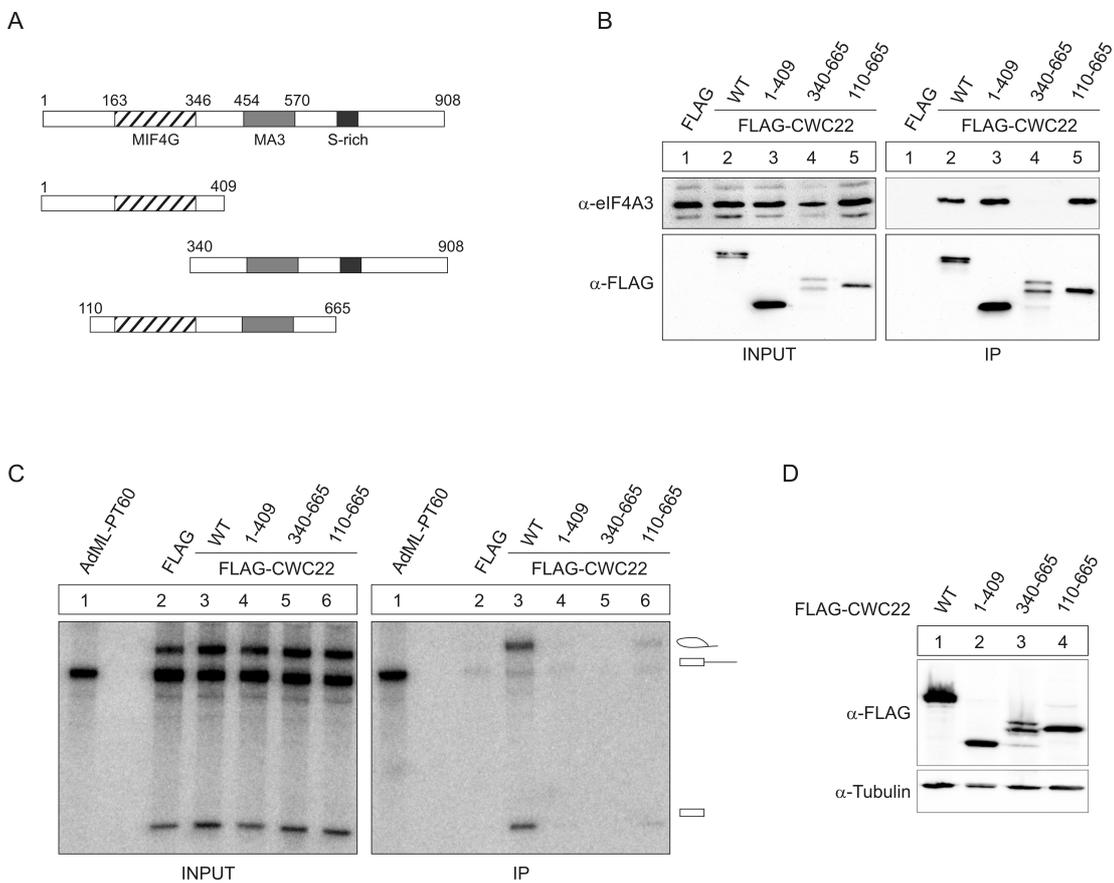


**Figure 11: Interaction of CWC22, eIF4A3 and Y14 with spliced mRNA.** (A)  $^{32}$ P-body-labeled MINX mRNA was spliced in the presence of FLAG-CWC22, FLAG-eIF4A3, FLAG-Y14 or unfused FLAG as a negative control. After FLAG-immunoprecipitation of mRNPs, the RNA was extracted, resolved on a denaturing urea-gel and visualized by phosphor imaging autoradiography. 10% of the splicing reaction was loaded as input. Schemes on the right site of the panels depict the splicing products. (B) Splicing and immunoprecipitation of AdML-PT60 mRNA as described in (A). (C) Expression of FLAG-tagged proteins in HEK293 extracts was detected by immunostaining with a FLAG-specific antibody. Tubulin served as a loading control.

#### 4.1.8 CWC22 uses multiple domains to interact with the spliceosome

In order to identify the part of CWC22 that mediates the interaction with the spliceosome, splicing of AdML-PT60 was repeated in the presence of different truncated versions of FLAG-CWC22 (Figure 12A). Surprisingly, only full-length CWC22 interacted well with spliced AdML-PT60 (Figure 12C, lane 3). In contrast, deletion of the C-terminal part of the protein including the MA3 domain (CWC22(1-409)) or an N-terminal deletion that removed the MIF4G domain (CWC22(340-908)) completely

eliminated mRNA binding (Figure 12C, lane 4 and 5). Furthermore, a construct that contained both MIF4G and MA3 domain but lacked the N- and C-terminal parts of the protein (CWC22(110-665)) showed a consistent, but strongly reduced interaction with AdML-PT60 (Figure 12C, lane 6). Taken together, these results show that even though the MIF4G domain of CWC22 is sufficient to bind eIF4A3 (Figure 12B), the MA3 domain, as well as elements in the N- and C-terminus of the protein, is important for the function of CWC22 in the spliceosome.



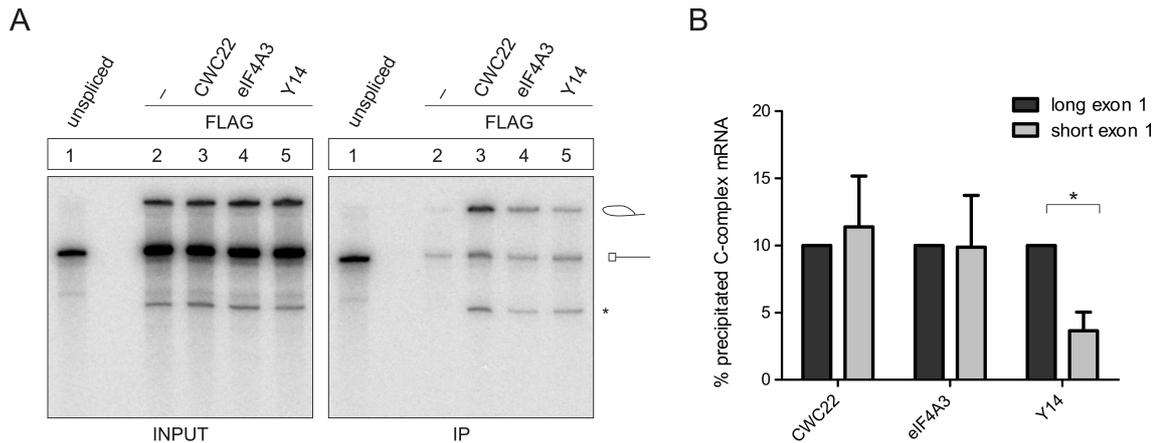
**Figure 12: CWC22 uses multiple domains to interact with AdML-PT60 mRNA.** (A) Schematic representation of the deletion mutants used in this assay. (B) FLAG-immunoprecipitation of CWC22 deletion mutants from RNase A treated cell lysates. Co-precipitated endogenous eIF4A3 was detected by staining with a specific antibody. Unfused FLAG served as a negative control. 10% of the cell lysate was loaded as input. (C) Splicing of AdML-PT60 mRNA in the presence of FLAG-CWC22 deletion mutants or unfused FLAG as a negative control. FLAG-containing mRNP complexes were immunoprecipitated, the extracted RNA was resolved by urea-PAGE and visualized by autoradiography. 10% of the splicing reaction was loaded as input. Schemes on the right site depict the splicing products. (D) Expression of FLAG-tagged proteins in HEK293 extracts was detected by immunostaining with a FLAG-specific antibody. Tubulin served as a loading control.

#### **4.1.9 Initial spliceosome interaction of CWC22 and eIF4A3 is independent of the EJC deposition site on spliced mRNA**

The EJC always assembles on the mRNA at a fixed distance (20-24nt) upstream of the splice site. Shortening exon 1 of the mRNA substrate to less than 20 nt will therefore prevent EJC assembly (Gehring et al., 2009a; Le Hir et al., 2000a). The role of this EJC deposition site during spliceosomal recruitment of CWC22, eIF4A3 and Y14 can be analyzed using an AdML-PT60 mRNA substrate with a first exon that consists of only 18 nucleotides (AdML-PT60/e1(18)). To be able to compare the pulldown of AdML-PT60/e1(18) to AdML-PT60 mRNA, the intensity of precipitated lariat mRNA (top band) relative to the input was analyzed. Since 10% of the total reaction was loaded as input, equal band intensities in input and IP lanes corresponded to 10% pulldown efficiency (Figure 13B).

Shortening the first exon had no effect on the interaction of FLAG-CWC22 with AdML-PT60 (Figure 13A, lane 3; compare to Figure 11B, lane 3; quantification Figure 13B). The same was true for FLAG-eIF4A3 (Figure 13A, lane 4; compare to Figure 11B, lane 4; quantification Figure 13B), which indicates that the initial interaction of both proteins with the spliceosome is independent of the EJC-binding site on the pre-mRNA. In contrast, FLAG-Y14 displayed a strongly reduced ability to interact with AdML-PT60/e1(18) compared to AdML-PT60 (Figure 13A, lane 5; compare to Figure 11B, lane 5; quantification Figure 13B). Thus, Y14 can only interact with the spliceosome when the EJC binding site on the mRNA is already present. This observation is in accordance with a model for EJC assembly where eIF4A3 is initially recruited to the spliceosome in an RNA-independent manner, then binds the mRNA 20-24nt upstream of the splice site and thereby forms a platform for the subsequent recruitment of the Y14/MAGOH heterodimer.

The analogous spliceosomal interaction pattern of CWC22 and eIF4A3 is compatible with a putative role for CWC22 in recruiting eIF4A3 to the spliceosome. To test this hypothesis, a comprehensive mutational analysis of the CWC22/eIF4A3 interaction surface was conducted and the mutants were functionally analyzed *in vitro* and in cell extracts. The results of this investigation will be presented in the next chapter.



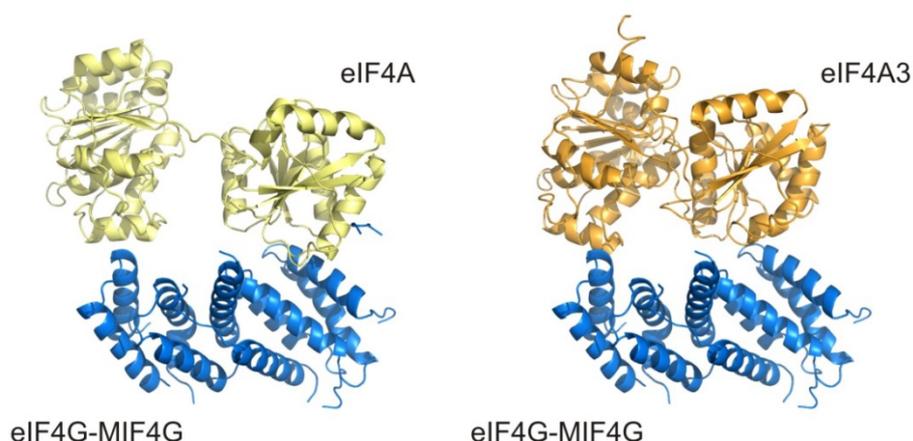
**Figure 13: Y14, but not CWC22 and eIF4A3, needs the EJC-assembly site to interact with AdML-PT60.** (A) Splicing of AdML-PT60/e1(18) mRNA in the presence of FLAG-CWC22, FLAG-eIF4A3, FLAG-Y14 or unfused FLAG as a negative control. After immunoprecipitating FLAG-containing mRNP complexes, the co-precipitated RNA was resolved by urea-PAGE and visualized by autoradiography. 10% of the splicing reaction was loaded as input. Schemes on the right site depict the splicing products. The asterisk (\*) denotes an unspecific splicing product. The corresponding western blot for protein expression is shown in **Figure 11C**. (B) Quantification of precipitated AdML-PT60 mRNA (long exon 1, Figure 11B) compared to AdML-PT60/e1(18) (short exon 1, Figure 13B). Error bars represent standard deviation (n=3, p = 0.0156).

## 4.2 CWC22 recruits eIF4A3 to the spliceosome and mediates EJC assembly

### 4.2.1 The CWC22/eIF4A3-interaction can be modeled with an existing eIF4G/eIF4A crystal structure

Co-IP and *in vitro* pulldown experiments revealed that eIF4A3 interacts directly with the MIF4G domain of CWC22. At the time when the initial experiments were performed, no structural information on CWC22 was available that would allow a more detailed analysis of the interaction surface. In several other protein complexes, however, MIF4G domains mediate the interaction with DExD/H-box helicases, among them the well-characterized eIF4A/eIF4G complex. Human eIF4A1 and eIF4A3 share 65% sequence identity (Li et al., 1999). Because of this high degree of sequence conservation, an existing X-ray structure of the yeast eIF4A/eIF4G-MIF4G complex (Schutz et al., 2008) can be used to predict a model for the interaction between eIF4A3 and CWC22 (Figure 14). To this end, the structures of eIF4A3 (as part of the EJC; PDB 2JOS) and eIF4A (in complex with the eIF4G; PDB 2VSO) were superimposed using DaliLite (Holm and Park, 2000) and the MIF4G domain of eIF4G served as a surrogate for the MIF4G domain of CWC22 (Figure 14). Of note, the structures of eIF4A and

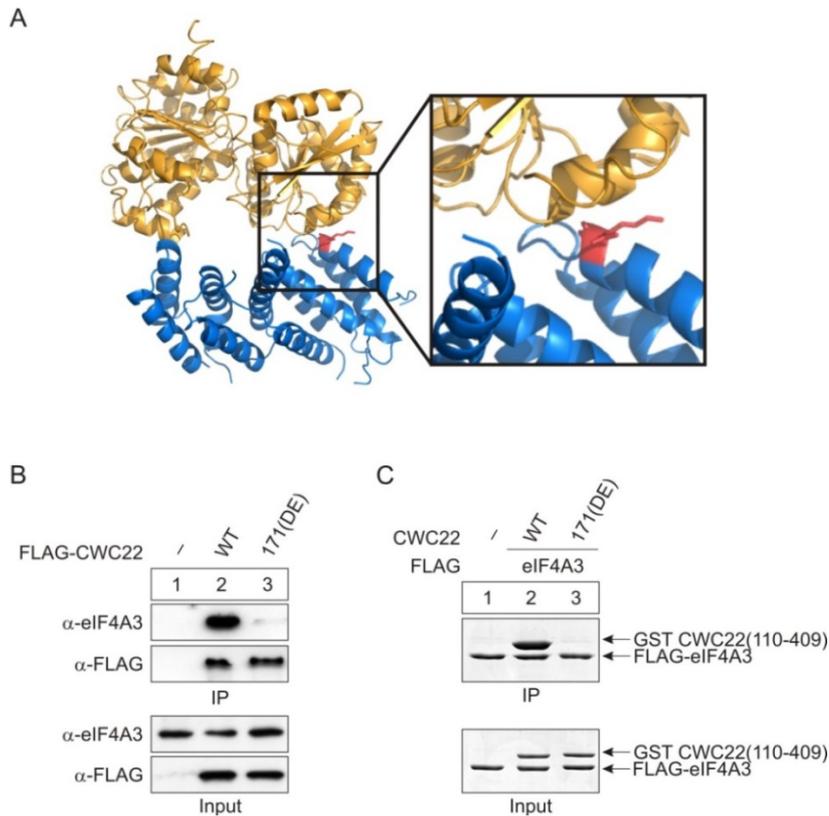
eIF4A3 do not completely align in this model, because eIF4A adopts an extended conformation in complex with eIF4G, whereas eIF4A3 is in its closed conformation when bound to the EJC (Bono et al., 2006; Schutz et al., 2008). However, the RecA-like domains of both eIF4A and eIF4A3 display an overall similar orientation, with the DEAD-box motifs facing each other at the active site cleft between RecA1 and RecA2 (Bono et al., 2006; Schutz et al., 2008).



**Figure 14: Modeling the interaction between eIF4A3 and CWC22 with an existing eIF4A/eIF4G crystal structure.** Left side: X-ray structure of the yeast eIF4A/eIF4G-MIF4G interaction (Schutz et al., 2008). Right side: A model for the interaction between eIF4A3 and CWC22 was generated through superimposition of eIF4A (PDB: 2VSO) and eIF4A3 (as part of the EJC; PDB 2JOS). The MIF4G-domain of eIF4G served as a surrogate for the MIF4G domain of CWC22. Note that eIF4A and eIF4A3 used for modeling adopt different conformations (extended vs. closed).

The model was used for structure-guided mutagenesis of residues at the putative eIF4A3-CWC22 interaction surface. Subsequently, the mutants were analyzed in Co-IP and *in vitro* pulldown experiments (Figure 15 and Figure 16). I was able to identify two residues (Asn-171 and Lys-172) on the surface of CWC22 that were crucial for eIF4A3 binding. Mutating Asn-171 and Lys-172 to Aspartate and Glutamate respectively (NK-171/2-DE) completely abolished the interaction between FLAG-CWC22 and endogenous eIF4A3 in cell lysates (Figure 15A). Likewise, recombinant CWC22(NK-171/2-DE) failed to interact with eIF4A3 *in vitro* (Figure 15B). Interestingly, Asn-171 and Lys-172 map to a 12 amino acid conserved sequence motif of yeast eIF4G that forms the largest contiguous interaction surface with eIF4A in the co-crystal (KSLLNKLTLEMF) (Schutz et al., 2008). Moreover, two mutations within the same sequence motif of the yeast NOM1 homolog Sgd1p (NSSLNKLSDSNI) mediated its

genetic interaction with the yeast homolog of eIF4A3, Fal1 (Alexandrov et al., 2011). These observations imply that eIF4G, NOM1 and CWC22 use similar residues to interact with eIF4A and eIF4A3 respectively, and suggest that the overall mode of interaction between DEAD-box helicases and their MIF4G interaction partners is conserved.

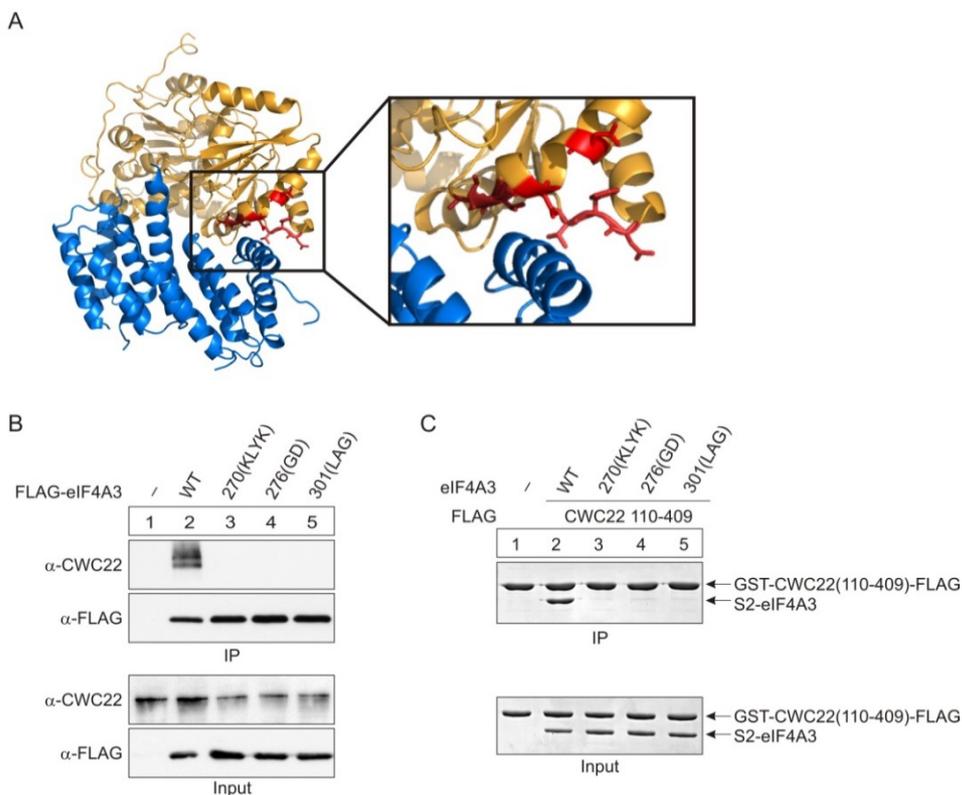


**Figure 15: Mutations at the putative eIF4A3/CWC22 interaction surface disrupt binding in cell lysates and *in vitro*.** (A) Modeled interaction between eIF4A3 and the MIF4G-domain of eIF4G. The respective mutated residues on the MIF4G-domain (Asp-171 and Lys-172) are depicted in red. (B) Immunoprecipitation of FLAG-CWC22(WT) and FLAG-CWC22(NK-171/2-DE) from RNase A treated cell lysates. Co-precipitated endogenous eIF4A3 was detected by immunostaining with a specific antibody. Unfused FLAG served as a negative control. 10% of the cell lysate was loaded as input. (C) Interaction between recombinant proteins *in vitro*. Co-precipitation of GST-CWC22(110-409) or GST-CWC22(110-409; NK-171/2-DE) with immobilized S2-FLAG-eIF4A3. The protein complexes were resolved by SDS-PAGE and visualized with coomassie. 10% of the assembly reaction was loaded as input.

Three point mutations were identified in the RecA2 domain of eIF4A3 (DLYD-270-KLYK, TI-276-GD and NFT-301-LAG) that impaired both the interaction with endogenous CWC22 in Co-IP experiments (Figure 16A) and the binding of recombinant eIF4A3 to CWC22 *in vitro* (Figure 16B). Interestingly, mutation of residues in the RecA1 domain of eIF4A3 or the corresponding surface of CWC22 did not impair the interaction between

eIF4A3 and CWC22 (Supplementary Figure 1 and Figure 21C). This observation is analogous to published data on the eIF4A/eIF4G and Dbp5/Gle1 complex, which showed that the C-terminal domains (RecA2) of the helicases display similar affinities for the MIF4G domain as the full-length protein (Dossani et al., 2009; Schutz et al., 2008). The mutational analysis of eIF4A3 thus corroborates results from the CWC22 mutagenesis, further strengthening the hypothesis that eIF4A3 and CWC22 form an eIF4A/eIF4A-like interaction pair.

A year after our initial report of the interaction between CWC22 and eIF4A3 (Steckelberg et al., 2012), a crystal structure of eIF4A3 and the MIF4G domain of CWC22 was published (Buchwald et al., 2013), which confirmed the interaction between the N-terminal helices of CWC22-MIF4G and the RecA2 domain of eIF4A3 proposed by this model (discussed in paragraph 4.2.6 and chapter 5).



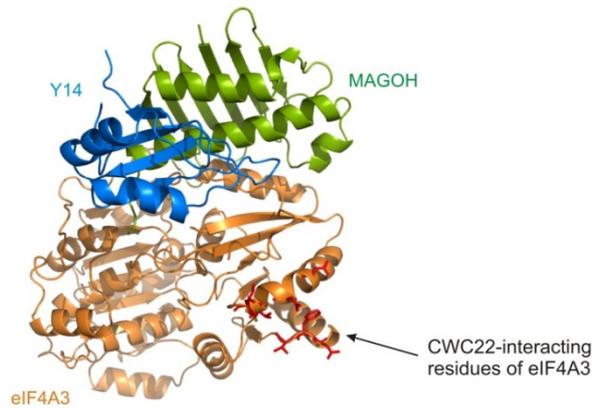
**Figure 16: Mutations at the putative eIF4A3/CWC22 interaction surface disrupt binding in cell lysates and *in vitro*.** (A) Modeled interaction between eIF4A3 and the MIF4G-domain of eIF4G. The mutated residues on eIF4A3 (DLYD-270-KLYK, TI-276-GD and NFT-301-LAG)) are depicted in red. (B) Immunoprecipitation of FLAG-eIF4A3 and the respective mutants from RNase A treated cell lysates. Co-precipitated endogenous CWC22 was detected by immunostaining with a specific antibody. Unfused FLAG served as a negative control. 10% of the cell lysate was loaded as input. (C) Interaction between recombinant proteins *in vitro*. Co-precipitation of recombinant S2-eIF4A3 and the indicated mutants with immobilized GST-CWC22(110-409)-FLAG. The protein complexes were resolved by SDS-PAGE and visualized with coomassie. 10% of the assembly reaction was loaded as input.

#### 4.2.2 CWC22-binding deficient eIF4A3 mutants form an EJC *in vitro*

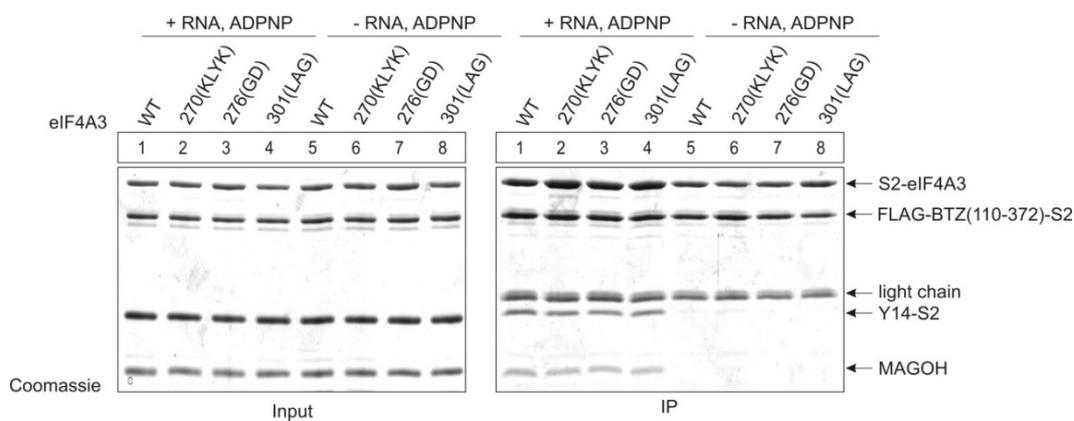
CWC22 is the first reported spliceosomal interaction partner of eIF4A3. As such, it is an ideal candidate for the recruitment eIF4A3 to the spliceosome and the subsequent initiation of EJC assembly. The eIF4A3 mutants described in paragraph 4.2.1 provide an elegant tool to test the putative role of CWC22 during splicing-dependent EJC assembly. But before the mutants can be analyzed in splicing reactions, their functionality and proper folding should be verified *in vitro*. The mutated residues do not overlap with binding sites for Y14/MAGOH or BTZ on eIF4A3 (Figure 17A); nevertheless, the general ability of eIF4A3 mutants to interact with the EJC proteins Y14, MAGOH and BTZ was confirmed in pulldown experiments. To this end, Y14-S2/MAGOH, FLAG-BTZ(110-372)-S2 and either WT or mutant S2-eIF4A3 were expressed in *E.coli* and purified as described in paragraph 4.1.5. EJC assembly was initiated by incubating the proteins in the presence of poly(U) RNA and ADPNP and complex formation was analyzed through FLAG-immunoprecipitation (Figure 17B, lane 1-4).

Importantly, all eIF4A3 versions precipitated comparable amounts of Y14 and MAGOH, which demonstrates that the mutants were able to incorporate into recombinant EJCs as efficiently as the WT protein (Figure 17B, lane 1-4). Hence, the mutations did not perturb the overall folding of the protein. To confirm that the complexes exhibit the characteristics of an EJC, the experiment was repeated in the absence of RNA and ADPNP (Figure 17B, lane 5-8). EIF4A3 can only interact with the Y14/MAGOH heterodimer when it adopts its closed conformation in a nucleotide- and RNA-bound state. Omitting RNA and ADPNP from the reaction prevented EJC assembly and served as a specificity control of the assay (Figure 17B, lane 5-8). Of note, FLAG-BTZ(110-372)-S2 can interact with S2-eIF4A3 in the absence of RNA and ADPNP (Figure 17B, lane 5-8), but no Y14-S2 and MAGOH was co-precipitated under these conditions, demonstrating that the observed protein complexes indeed possess the characteristics of EJCs.

A



B

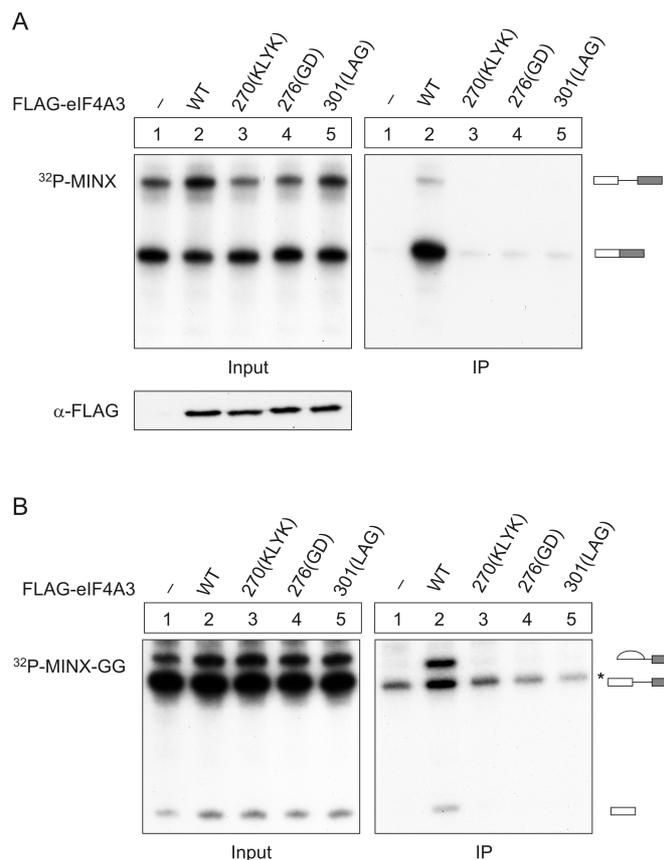


**Figure 17: CWC22-binding deficient eIF4A3 mutants can form an EJC *in vitro*.** (A) Crystal structure of eIF4A3 in complex with Y14 and MAGOH (PDB 2JOS). The mutated residues are depicted in red. (B) *In vitro* EJC assembly with WT and mutant eIF4A3. FLAG-BTZ(110-372)-S2, Y14-S2, MAGOH and S2-eIF4A3 (WT or mutant) were purified from *E.coli* and incubated under EJC assembly conditions. The protein complexes were co-precipitated with immobilized FLAG-BTZ(110-372)-S2, resolved by SDS-PAGE and visualized with coomassie. EJC assembly in the absence of RNA and ADPNP served as a specificity control (lane 5-8). 10% of the assembly reaction was loaded as input.

#### 4.2.3 CWC22-binding deficient eIF4A3 mutants fail to form splicing-dependent EJCs

Whereas the EJC core assembles *in vitro* without the help of additional factors, EJC assembly in living cells and cell extracts is strictly splicing-dependent (Gehring et al., 2009a; Le Hir et al., 2000a; Le Hir et al., 2000b). The CWC22-binding deficient eIF4A3 mutants were therefore analyzed using the *in vitro* splicing system introduced in paragraph 4.1.6. Splicing of the MINX mRNA substrate was carried out in the presence of HEK293 extracts expressing either WT or mutant FLAG-eIF4A3. Subsequent FLAG-immunoprecipitation revealed that FLAG-eIF4A3 specifically precipitated spliced mRNA, marking it as part of the splicing-dependent EJC (Figure 18A, lane 2). However, none of the FLAG-eIF4A3 mutants was able to precipitate MINX (Figure 18A, lane 3-5),

demonstrating that the CWC22-binding site on eIF4A3 is crucial for its splicing-dependent interaction with mRNA. To further delineate which step of eIF4A3 recruitment is impaired, the splicing reaction was repeated using a MINX-version with an AG→GG mutation of the splice acceptor site (MINX-GG, Figure 18B). Like the AdML-PT60 construct introduced in paragraph 4.1.7, MINX-GG leads to an arrest of the splicing reaction after the first catalytic step (during C-complex formation). Whereas MINX-GG could be precipitated with FLAG-eIF4A3(WT), none of the eIF4A3 mutants showed any interaction with the mRNA (Figure 18B, lane 3-5). These data imply that the mutations compromise the initial interaction between eIF4A3 and the spliceosome, and suggest a role for CWC22 during the early recruitment of eIF4A3 to spliced mRNA. Of note, mutations in the RecA1 domain of eIF4A3 did not impair the splicing-dependent EJC assembly (Supplementary Figure 4).



**Figure 18: CWC22-binding deficient eIF4A3 mutants fail to form a splicing-dependent EJC.** (A) *In vitro* splicing of <sup>32</sup>P-MINX was carried out in the presence of FLAG-eIF4A3, the indicated mutants of eIF4A3 or unfused FLAG as a negative control. After immunoprecipitation of FLAG-containing mRNPs, the co-precipitated mRNA was resolved by urea-PAGE and visualized by autoradiography. 10% of the splicing reaction was loaded as input. The western blot (bottom) shows the expression of FLAG-tagged proteins in HEK293 whole cell extracts. (B) *In vitro* splicing and immunoprecipitation of <sup>32</sup>P-MINX-GG mRNA as described in (A). The corresponding western blot for protein expression is shown in (A).

The results described here provide strong evidence for a role of CWC22 during spliceosomal recruitment of eIF4A3. Nevertheless, the use of CWC22 binding deficient mutants is an indirect proof and does not exclude the possibility that the mutated residues have a different, CWC22-independent effect on eIF4A3. Further evidence that CWC22 is indeed the factor linking eIF4A3 to the spliceosome will therefore be provided in the following paragraphs.

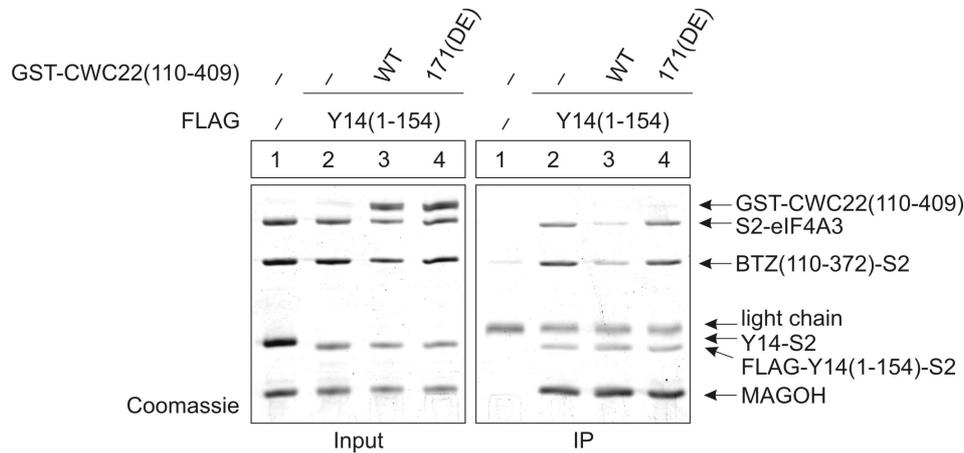
#### **4.2.4 Recombinant CWC22(110-409) disturbs EJC assembly *in vitro***

Previous experiments showed that recombinant CWC22 is unable to interact with the assembled EJC *in vitro* (Figure 10B). However, it remains to be elucidated, whether recombinant CWC22 has an effect on *in vitro* EJC assembly itself. Hence, a recombinant EJC was assembled from purified S2-eIF4A3, BTZ(110-372)-S2, MAGOH and FLAG-Y14(1-154)-S2 in the presence or absence of GST-CWC22(110-409). Assembled complexes were precipitated with anti-FLAG beads and analyzed by SDS-PAGE and coomassie staining (Figure 19).

Earlier experiments had revealed the mutually exclusive binding of eIF4A3 to either EJC or CWC22 (Figure 10B). To selectively analyze EJC-bound eIF4A3, the complex was precipitated with FLAG-tagged Y14 instead of FLAG-tagged BTZ(110-372), as before. A slightly shorter version of Y14 (comprising of residues 1-154) was used to prevent the co-migration of FLAG-Y14 with the antibody light chain on the SDS-gel. This should not affect complex formation, since Y14(1-154) possesses identical EJC binding affinities as the full-length protein (Bono et al., 2006).

Interestingly, addition of GST-CWC22(110-409) to EJC assembly reactions severely reduced the ability of FLAG-Y14(1-154)-S2 to co-precipitated S2-eIF4A3 and BTZ(110-372)-S2, indicating that *in vitro* EJC assembly is impaired in the presence of recombinant CWC22 (Figure 19, lane 3). To prove that this effect is specific, the experiment was repeated with the mutant GST-CWC22(110-409(NK-171/2-DE)). In contrast to the WT protein, GST-CWC22(110-409(NK-171/2-DE) had no effect on the co-precipitation of S2-eIF4A3 and BTZ(110-372)-S2, demonstrating that a direct interaction between CWC22 and eIF4A3 is required to block EJC assembly (Figure 19, lane 4). This observation suggests that GST-CWC22(110-409) competes with the EJC

proteins for eIF4A3 binding, and addition of excess CWC22(110-409) can outcompete EJC assembly *in vitro*.



**Figure 19: Recombinant CWC22(110-409) impairs EJC assembly *in vitro*.** Recombinant EJCs were assembled in the presence of GST-CWC22(110-409) (lane 3) or GST-CWC22(110-409; NK-171/2-DE) (lane 4). Complexes were immunoprecipitated by anti-FLAG beads, resolved by SDS-PAGE and stained with coomassie. An EJC containing no FLAG tag served as a negative control. 10% of the assembly reaction was loaded as input.

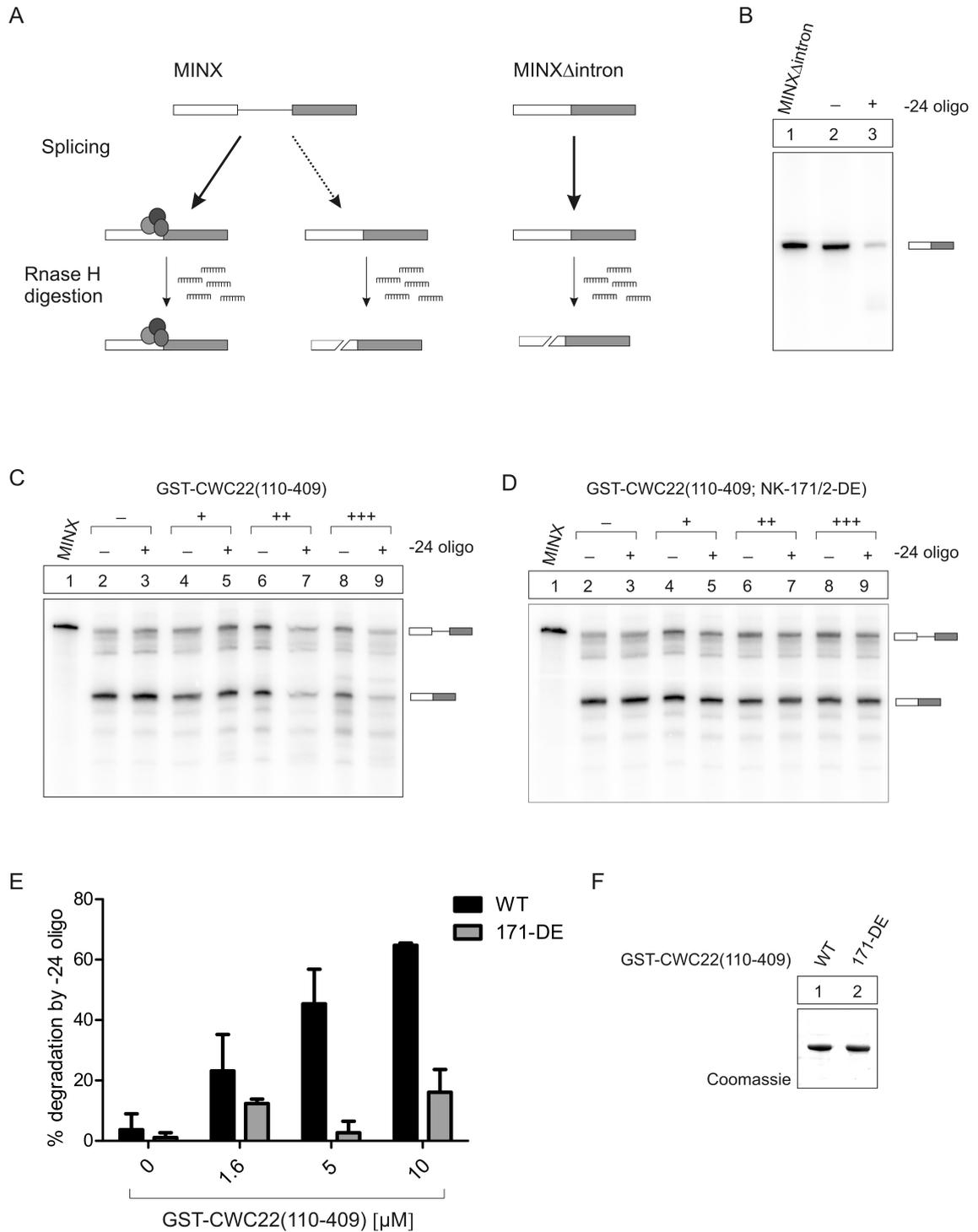
#### 4.2.5 Recombinant CWC22(110-409) disturbs EJC assembly in cell extracts

The experiments described so far demonstrate that the interaction between CWC22 and eIF4A3 is necessary for EJC assembly in cell extracts. But the same interaction is dispensable or even deleterious for the assembly of recombinant EJCs *in vitro*. This discrepancy reflects the differential dependence of EJC assembly on pre-mRNA splicing. It also implies that the interaction between CWC22 and the spliceosome is absolutely essential for EJC assembly in living cells and cell extracts. To test this hypothesis, I performed *in vitro* splicing reactions of MINX mRNA, supplemented with increasing amounts of recombinant GST-CWC22(110-409) (Figure 20F). This piece of CWC22 is sufficient to bind eIF4A3 but does not interact with the spliceosome (Figure 12B, lane 3 and Figure 12C, lane 4).

EJC assembly on spliced mRNA was analyzed through oligonucleotide-directed RNase H digestion (Figure 20A). After splicing, a DNA-oligonucleotide complementary to the EJC-deposition site on MINX (-24nt upstream of the splice site) was added to the reaction. This oligonucleotide forms a DNA/RNA heteroduplex with MINX and thereby activates RNase H, present in nuclear extracts, to cleave the RNA. The cleavage is visible as a decrease of full-lengths mRNA. However, if an EJC is bound to the mRNA upstream of the exon-exon junction, this sequence is protected from RNase H cleavage, and the reduced cleavage activity serves as a direct indicator for EJC formation (Le Hir et al., 2000a). The maximal cleavage activity of RNase H was determined using an intronless version of MINX (MINX $\Delta$ intron) (Figure 20B). MINX $\Delta$ intron is not spliced, and accordingly no EJC protects the mRNA from RNase H cleavage. The maximal cleavage activity thus achieved ranged from 80 to 90 % (Figure 20B). In the absence of GST-CWC22(110-409), splicing of MINX initiated very efficient EJC assembly, so that almost no RNase H cleavage of spliced mRNA was detected (Figure 20C, lane 1-2). Supplementing the splicing reaction with increasing amounts of GST-CWC22(110-409) led to a concentration-dependent deprotection of the spliced mRNA, indicating that GST-CWC22(110-409) reduces the number mRNA molecules that carry an EJC (Figure 20C, lane 4-9 and quantification Figure 20E). To test the specificity of this assay, the RNase H-protection assay was repeated in the presence of GST-CWC22(110-409(NK-171/2-DE)) (Figure 20D). Of note, no increase in RNase H digestion could be observed in the presence of mutated CWC22 (Figure 20D, lane 4-9), demonstrating that a direct interaction between eIF4A3 and recombinant GST-CWC22(110-409) is required to prevent EJC assembly.

These data are in accordance with the previously described *in vitro* results, showing that binding of eIF4A3 to CWC22 in the absence of additional spliceosomal factors adversely affects EJC assembly. A piece of CWC22 that binds eIF4A3 but fails to interact with the spliceosome thus represents a dead end and inhibits EJC assembly.

## Results

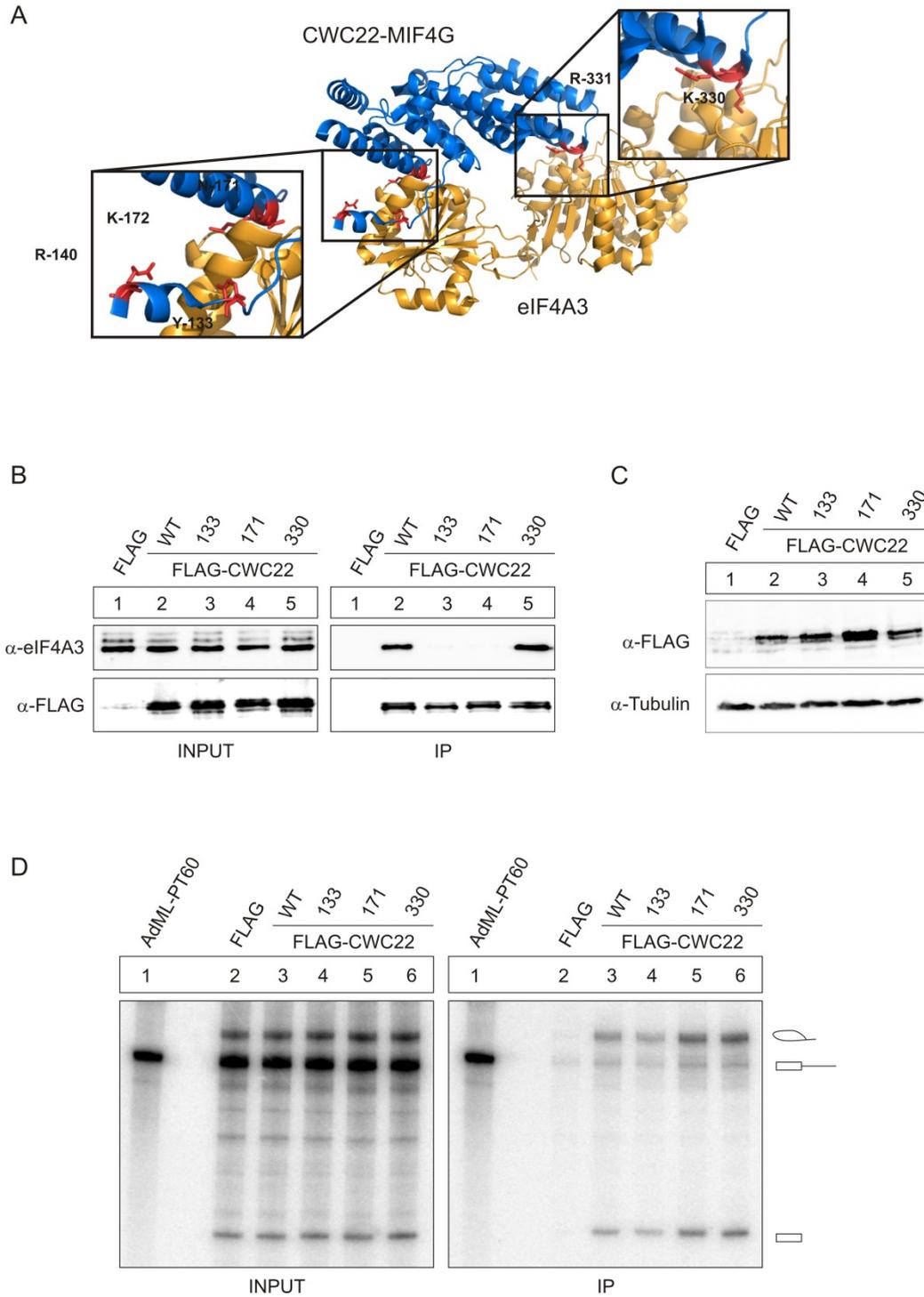


**Figure 20: GST-CWC22(110-409) prevents splicing-dependent EJC assembly.** (A) Schematic representation of the RNase H digestion assay. (B) Oligonucleotide-directed RNase H digestion of  $^{32}$ P-MINX $\Delta$ intron mRNA. (C-D) Oligonucleotide-directed RNase H digestion of  $^{32}$ P-MINX mRNA that was spliced in the presence of increasing amounts of recombinant GST-CWC22(110-409) (C) or GST-CWC22(110-409; NK-171/2-DE) (D). (E) Quantification of RNase H degradation activity in the presence of GST-CWC22(110-409) (WT) or GST-CWC22(110-409; NK-171/2-DE) (171-DE). The cleavage activity on intronless mRNA (MINX $\Delta$ intron) was set to 100. (F) 5  $\mu$ g of GST-CWC22(110-409) and GST-CWC22(110-409; NK-171/2-DE) were resolved by SDS-PAGE and visualized with coomassie.

#### 4.2.6 CWC22 interaction with the spliceosome is independent of eIF4A3

*In vitro* splicing experiments presented in paragraph 4.1.8 showed that a CWC22-construct, which lacked the MIF4G domain, failed to precipitate AdML-PT60 mRNA (Figure 12C, lane 5). Since the MIF4G domain mediates the binding to eIF4A3, it should be elucidated whether the interaction with eIF4A3 is required for the spliceosomal recruitment of CWC22. To address this question, CWC22 with mutated eIF4A3-binding sites was analyzed through *in vitro* splicing reactions (Figure 21). Apart from the CWC22(NK-171/2-DE) mutant introduced in paragraph 4.2.1, two new CWC22 mutants were analyzed in this assay (CWC22(Y-133-E/R-140-E) and CWC22(KR-330/1-DE)). Both mutants were designed on the basis of a recently published co-crystal structure of eIF4A3 and the MIF4G domain of CWC22 (Buchwald et al., 2013) (Figure 21A). CWC22(Y-133-E/R-140-E) was directly adopted from this publication. Isothermal titration calorimetry (ITC) experiments revealed that this mutant possesses a 100 fold lower affinity for eIF4A3 than the wild type protein (3  $\mu$ M compared to 30 nM (Buchwald et al., 2013). In Co-IP experiments, this decrease in affinity resulted in a complete loss of binding to endogenous eIF4A3, comparable to that observed for CWC22(NK-171/2-DE) (Figure 21B, lane 3-4). The second mutant was designed to disrupt the interaction between the N-terminal part of CWC22 and the RecA1 domain of eIF4A3. According to the crystal structure, amino acid residues Lys-330, Tyr-334 and Val-338 of CWC22 interact with Cys-99 of eIF4A3. Cys-99 is one of the few interacting amino acid residues that are not conserved between eIF4A3 and eIF4A1, making it a potential candidate for the specificity of CWC22-binding. Interestingly, mutating Lys-330 and Arg-331 did not affect the binding to endogenous eIF4A3 in Co-IP experiments (Figure 21B, lane 5) and resulted in similar binding affinities in ITC experiments (Supplementary Figure S2).

The analysis of all three CWC22 mutants revealed that neither the interaction with RecA2, nor the binding to the RecA1 domain of eIF4A3 is required for the interaction of CWC22 with AdML-PT60 mRNA (Figure 21D). Since the mutations NK-171/2-DE and Y-133-E/R-140-E both completely inhibit the ability of CWC22 to interact with eIF4A3 in cell lysates (Figure 21B), this data clearly demonstrate that the interaction of CWC22 with the spliceosome is independent from its binding to eIF4A3.



**Figure 21: Interaction between CWC22 and the spliceosome is independent of eIF4A3-binding.** (A) X-ray structure of eIF4A3 in complex with the CWC22 MIF4G-domain (PDB: 4C9B, (Buchwald et al., 2013). The mutated residues are depicted in red. (B) FLAG-immunoprecipitation of CWC22 mutants from RNase A treated cell lysates. Co-precipitated endogenous eIF4A3 was detected by staining with a specific antibody. Unfused FLAG served as a negative control. 10% of the cell lysate was loaded as input. (C) Expression of FLAG-tagged proteins in HEK293 extracts was detected by immunostaining with a FLAG-specific antibody. Anti-Tubulin served as a loading control. (D) Splicing of AdML-PT60 mRNA in the presence of FLAG-CWC22, the indicated mutants or unfused FLAG as a negative control. FLAG-containing mRNP complexes were immunoprecipitated, the extracted RNA was resolved by urea-PAGE and visualized by autoradiography. 10% of the splicing reaction was loaded as input. Schemes on the right site depict the splicing products.

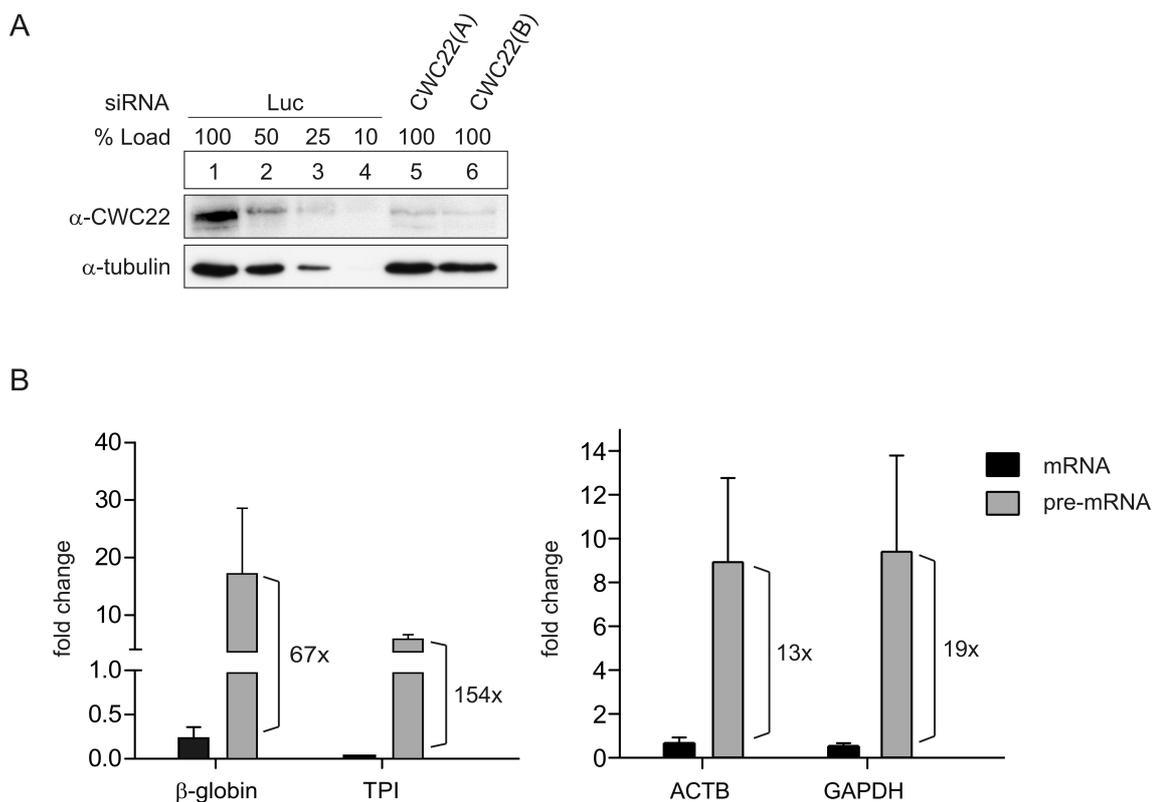
### 4.3 CWC22 is required for pre-mRNA splicing in human cells

#### 4.3.1 siRNA knock down of CWC22 leads to an accumulation of pre-mRNA

CWC22 has been characterized as an essential splicing factor in *Saccharomyces cerevisiae* (Yeh et al., 2011). It is recruited to the pre-catalytic spliceosome and stays associated with the splicing machinery throughout the whole splicing reaction. Interestingly, the DExD/H-box helicase Prp2 was identified as an interaction partner of yeast CWC22. Functionally, CWC22 is required for the first catalytic step of splicing by coupling the ATPase activity of Prp2 to the release of the U2 components SF3a and SF3b from the branch site of the mRNA (Yeh et al., 2011). In contrast, the role of human CWC22 during pre-mRNA splicing is only poorly understood.

Since many splicing factors are conserved from yeast to humans it is very likely that CWC22 also functions during pre-mRNA splicing in human cells. To verify this assumption, CWC22 was depleted from HeLa cells using two different short interfering RNAs (siRNAs) (Figure 22A). The CWC22-depletion efficiency was assessed through immunoblotting with an antibody against endogenous CWC22. Both siRNA constructs led to a reduction of protein levels to approximately 20% (Figure 22A, lane 5-6). An siRNA targeting luciferase (Luc) mRNA served as a negative control (Figure 22A, lane 1-4). The effect of CWC22 depletion on pre-mRNA splicing was subsequently investigated by quantitative real time PCR (qRT-PCR). To this end, two different reporter mRNAs were transfected in siRNA-treated cells. The  $\beta$ -globin (HBB) reporter contains two introns whereas the triose phosphate isomerase (TPI) reporter has six introns, but both mRNAs are very efficiently spliced in control cells. Mature mRNA levels were specifically amplified with intron-spanning primer pairs whereas pre-mRNA was detected with primers that targeted intronic sequences (Supplementary Figure 3). All results were normalized to the expression of the intronless neomycin resistance cassette (APH 3' II) of the eukaryotic expression vector pCI-neo. Upon depletion of CWC22, pre-mRNA levels were 10 to 30-fold up-regulated and the corresponding mature mRNAs were reduced down to 5%, demonstrating that splicing is severely impaired in CWC22-depleted cells (Figure 22B, left graph). To further confirm these results, the effect of CWC22 depletion on endogenous mRNAs was analyzed. Both, splicing of actin  $\beta$  (ACTB) and glyceraldehyde-3-phosphate-dehydrogenase (GAPDH)

was investigated using specific primer pairs designed to target pre-mRNA and mature mRNA, respectively. As previously observed for the reporter mRNAs, both endogenous mRNAs displayed a strong splicing defect upon CWC22 knock down (Figure 22B, right graph). On average, the pre-mRNA was approximately 10-fold up-regulated and the mature mRNA was reduced to 50% compared to control cells (Figure 22B, right graph). The effect on the endogenous transcripts was less severe, which can be explained by the long half-life of both mRNAs. Whereas the reporter constructs were only expressed 12 hours after depletion of CWC22, endogenous mRNAs detected in this assay represent a mixed population of mRNAs that were transcribed (and spliced) before and after siRNA-induced depletion of CWC22. Nevertheless, the overall results for both types of mRNA are similar and provide very strong evidence for an essential role of CWC22 during pre-mRNA splicing in human cells.



**Figure 22: siRNA-mediated knockdown of CWC22 impairs pre-mRNA splicing.** (A) HeLa cells were transfected with two different siRNAs (CWC22(A) and CWC22(B)) targeting CWC22 mRNA. siRNA against luciferase mRNA (Luc) served as a negative control. The knock down efficiency was assessed by immunoblotting with an antibody against endogenous CWC22. Tubulin served as a loading control. (B) qRT-PCR analysis showing the fold change of pre-mRNA and mRNA levels upon CWC22-depletion for  $\beta$ -globin and TPI reporter mRNAs (left side) and endogenous ACTB and GAPDH transcripts (right side). Numbers beside the bars represent the ratio between pre-mRNA and mRNA. Error bars represent the standard deviation between 3 individual experiments.

#### **4.3.2 MIF4G and MA3 domains of CWC22 are required for pre-mRNA splicing**

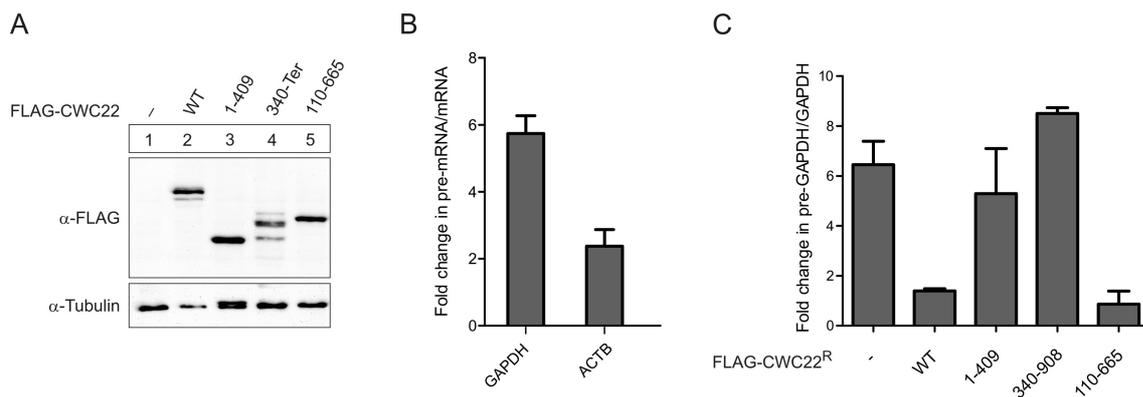
In yeast, a segment of Cwc22 consisting of the MA3 domain and a conserved downstream region is sufficient to rescue growth in CWC22-deficient cells (Yeh et al., 2011). Moreover, the corresponding recombinant piece of CWC22 could complement splicing in CWC22-depleted cell extracts (Yeh et al., 2011). Taken together, these results demonstrate that the MA3, but not the MIF4G domain of Cwc22, is sufficient to support pre-mRNA splicing in yeast.

In order to decipher which parts of CWC22 are required for pre-mRNA splicing in human cells, I performed complementation experiments with siRNA-resistant CWC22-constructs in CWC22-depleted cells. To circumvent the need for transient transfection, stable HEK293 Flp-In T-REx cell lines (Invitrogen) expressing doxycycline-inducible CWC22 were generated. This system deploys the FLP DNA recombination system derived from *S.cerevisiae* to facilitate the targeted integration of a transgene into the genome of mammalian cells (see Material and Methods (Craig, 1988; Sauer, 1994)). The inserted gene is expressed under the control of the human cytomegalovirus immediate-early (CMV) promoter and two tetracycline operator (TO) sites (derived from the *E.coli* tetracycline resistance operon (Hillen and Berens, 1994)) so that protein expression is constitutively repressed by a Tet-repressor (TetR) and only induced when doxycycline (a tetracycline-derivative) is added (see Materials and Methods).

A total of six stable cell lines was generated, expressing siRNA-resistant FLAG-CWC22(WT), FLAG-CWC22(1-409), FLAG-CWC22(340-908), FLAG-CWC22(110-665), as well as the mutant versions FLAG-CWC22(NK-171/2-DE) and FLAG-CWC22(110-665; NK-171/2-DE). Flp-In 293 cells without an integrated transgene served as a negative control. Protein expression upon doxycycline treatment was verified by immunoblotting with an antibody against the FLAG-tag (Figure 23A, Figure 24A, Figure 26A). Subsequently, the effect of CWC22-depletion was analyzed by qRT-PCR as described in paragraph 4.3.1. Prior to this, CWC22 had only been depleted in HeLa cells (Figure 22); to test if Flp-In 293 cells are equally amenable to siRNA-mediated knock down, the effect of CWC22-depletion on splicing of GAPDH and ACTB mRNA was analyzed in control cells (Figure 23B). To avoid the need for an internal unspliced

control, the ratio of pre-mRNA/mRNA was calculated for each sample. Notably, the change in the pre-mRNA/mRNA ratio upon CWC22-depletion was less pronounced in Flp-In 293 cells compared to HeLa cells (2-6 fold compared to 13-19 fold; Figure 23B and Figure 22B). Flp-In 293 cells tend to grow in heaps, which possibly resulted in a lower knock down efficiency compared to HeLa cells. Since stronger effects were observed for GAPDH mRNA compared to ACTB, all subsequent complementation experiments were performed using GAPDH primer pairs.

Importantly, expression of siRNA-resistant FLAG-CWC22 completely restored pre-mRNA splicing in knock down cells, which confirms the specificity of the knock down and demonstrates the feasibility of the complementation experiment (Figure 23C). Analyzing three truncated versions of CWC22 revealed that neither the N-terminal half of CWC22 (FLAG-CWC22(1-409)), nor the C-terminal segment of the protein (FLAG-CWC22(340-908)) were able to restore splicing in CWC22-depleted cells (Figure 23C). In contrast, the central piece of CWC22 that contains both MIF4G and MA3 domain (FLAG-CWC22(110-665)) displayed a similar rescuing potential as the WT protein (Figure 23C). This observation is surprising, since only the full length protein, but not the truncated version (FLAG-CWC22(110-665)) was able to efficiently precipitate AdML-PT60 in *in vitro* splicing experiments (Figure 12C). The results show that even though CWC22(110-665) interacts less efficiently with the spliceosome, it is sufficient to rescue pre-mRNA splicing in an over-expression system. The central piece of CWC22, containing both MIF4G and MA3 domain, thus represents the minimal fragment that can support pre-mRNA splicing in human cells.



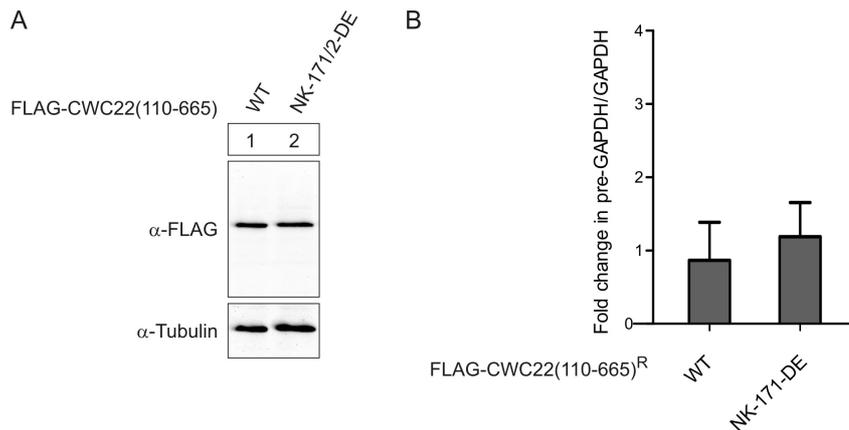
**Figure 23: The MIF4G and MA3 domain of CWC22 are required for pre-mRNA splicing.** (A) Stable Flp-In 293 cells were induced with doxycycline and expression of FLAG-CWC22 constructs was assessed by immunoblotting with an FLAG-specific antibody. (B) Fold change in pre-mRNA/mRNA ratio of GAPDH and ACTB upon CWC22-depletion in Flp-In control cells was quantified by qRT-PCR. Bars represent the mean of three independent experiments; error bars depict the standard deviation (SD). (C) CWC22 complementation assay. Protein expression in Flp-IN 293 cells was induced by doxycycline and the change in pre-GAPDH/GAPDH ratio upon CWC22-depletion was analyzed as described in (B).

#### 4.3.3 The interaction between CWC22 and eIF4A3 is dispensable for splicing

*In vitro* splicing experiments showed that the binding of CWC22 to eIF4A3 is not required for the interaction between CWC22 and AdML-PT60 mRNA (Figure 21). Whether this interaction is equally dispensable for the function of CWC22 during pre-mRNA splicing, remains elusive. Hence, the link between pre-mRNA splicing and eIF4A3 interaction was investigated using an eIF4A3-binding deficient mutant of CWC22.

Surprisingly, the over-expression of FLAG-CWC22(NK-171/2-DE) induced massive cell death, even in the absence of CWC22-depletion, indicating that the protein has a dominant negative effect on cell viability (discussed in paragraph 4.3.4). Depleting endogenous CWC22 in FLAG-CWC22(NK-171/2-DE)-expressing cells further aggravated the effect and made it impossible to analyze pre-mRNA splicing by qRT-PCR.

Since FLAG-CWC22(110-665) had a similar rescuing potential as full-length FLAG-CWC22 (Figure 23C), the experiment was repeated with this shorter version of CWC22. Of note, over-expression of FLAG-CWC22(110-665; NK-171/2-DE) did not promote cell death, so that the mutant could be compared to FLAG-CWC22(110-665) in complementation experiments. Interestingly, both FLAG-CWC22(110-665) and FLAG-CWC22(110-665; NK-171/2-DE) were able to restore splicing in CWC22-depleted cells (Figure 24B), indicating that the interaction with eIF4A3 is not required for the function of CWC22 during pre-mRNA splicing.

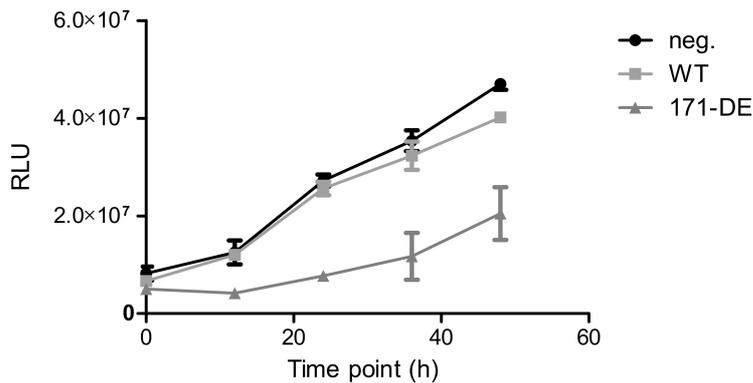


**Figure 24: The interaction between CWC22 and eIF4A3 is not required for pre-mRNA splicing.** (A) Expression of FLAG-CWC22 constructs in doxycycline-induced stable Flp-In 293 cells was assessed by immunoblotting with an antibody against FLAG. (B) CWC22 complementation assay. Protein expression in Flp-IN 293 cells was induced with doxycycline and the change in pre-GAPDH/GAPDH ratio upon CWC22-depletion was analyzed by qRT-PCR. Bars represent the mean of three independent experiments; error bars depict the standard deviation (SD).

#### 4.3.4 Over-expression of FLAG-CWC22(NK-171/2-DE) is toxic for HEK293 cells

Over-expression of FLAG-CWC22(NK-171/2-DE) in Flp-In 293 cells led to massive cell death. The observed effect was quantified with a luminescence-based cell viability assay. The CellTiter-Glo assay (Promega) uses ATP from metabolically active cells to catalyze the mono-oxygenation of luciferin, which results in a stable luminescent signal. The measured luminescence is therefore directly proportional to the number of viable cells.

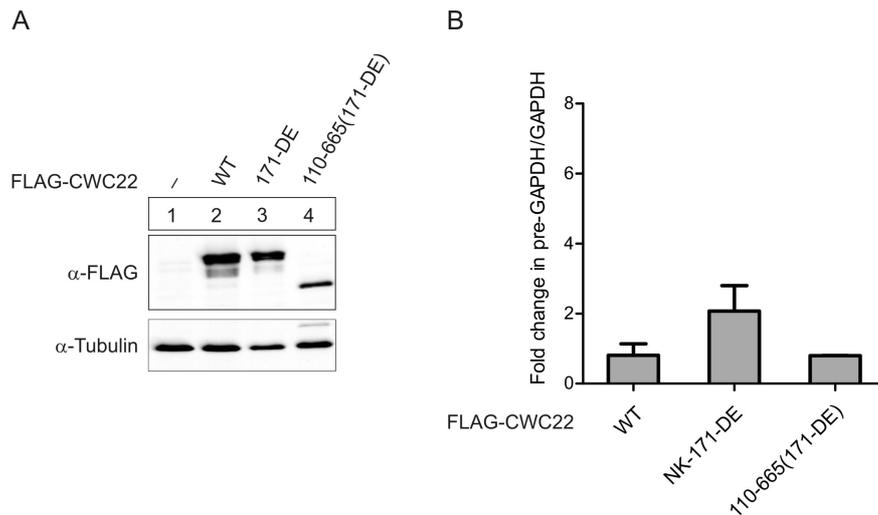
Equal numbers of Flp-In cells with integrated FLAG-CWC22(WT), FLAG-CWC22(NK-171/2-DE) or without an integrated transgene were grown in 96-well plates in 100  $\mu$ l cell culture medium. Twelve hours after seeding, transgene expression was induced by adding doxycycline and the number of viable cells was determined after 0, 12, 24, 36 and 48 hours using the CellTiter-Glo assay (Figure 25). Interestingly, Flp-In cells expressing FLAG-CWC22(NK-171/2-DE) already displayed severely reduced numbers of viable cells after 12 hours. In contrast, over-expression of FLAG-CWC22(WT) had no effect on the survival of cells. These observations demonstrate that over-expression of FLAG-CWC22(NK-171/2-DE) has a very rapid effect on cell viability and suggest that the FLAG-CWC22(NK-171/2-DE) might impinge on both proliferation and survival of cells. Additional tests that distinguish between cytotoxic and cytostatic effects will be necessary to identify the exact mechanism by which over-expression of FLAG-CWC22(NK-171/2-DE) impairs cell viability.



**Figure 25: Over-expression of FLAG-CWC22 (NK-171/2-DE) decreases the number of viable cells.** CellTiter-Glo assay on Flp-In cells over-expressing FLAG-CWC22 (WT), FLAG-CWC22(NK-171/2-DE) (171-DE) or no protein (neg.) Luminescent signal (relative light units, RLU) is proportional to the number of viable cells per well. Error bars represent the standard deviation (SD) of 4 experiments.

#### 4.3.5 Over-expression of FLAG-CWC22(NK-171/2-DE) inhibits EJC assembly without impairing pre-mRNA splicing

Over-expression of FLAG-CWC22(NK-171/2-DE) in Flp-In 293 cells induced cell death. Since this effect was detectable in the presents of endogenous CWC22, it implied that FLAG-CWC22(NK-171/2-DE) exerts a dominant negative effect on vital cellular processes. The crucial role of CWC22 during pre-mRNA splicing and EJC assembly suggested that either one or both of these functions are impaired in FLAG-CWC22(NK-171/2-DE) over-expressing cells. To identify the cause of cell death, splicing in Flp-In 293 cells was analyzed by qRT-PCR. To this end, expression of FLAG-CWC22, FLAG-CWC22(NK-171/2-DE) and FLAG-CWC22(110-665; NK-171/2-DE) was induced with doxycycline (Figure 26A). 48 hours after transgene expression, the change in the pre-GAPDH/GAPDH ratio compared to un-induced control cells was determined by qPCR (Figure 26B). Neither the expression of FLAG-CWC22, nor of FLAG-CWC22(110-665; NK-171/2-DE) had any effect on pre-mRNA splicing (Figure 26B). Over-expression of FLAG-CWC22(NK-171/2-DE) led to a very small increase (~2-fold) in pre-mRNA/mRNA ratio, indicating that splicing is slightly impaired (Figure 26B). However, this small splicing defect is not likely to cause the observed cell death, since depletion of CWC22 results in a more pronounced splicing defect (Figure 22-23) but appeared to be less toxic.

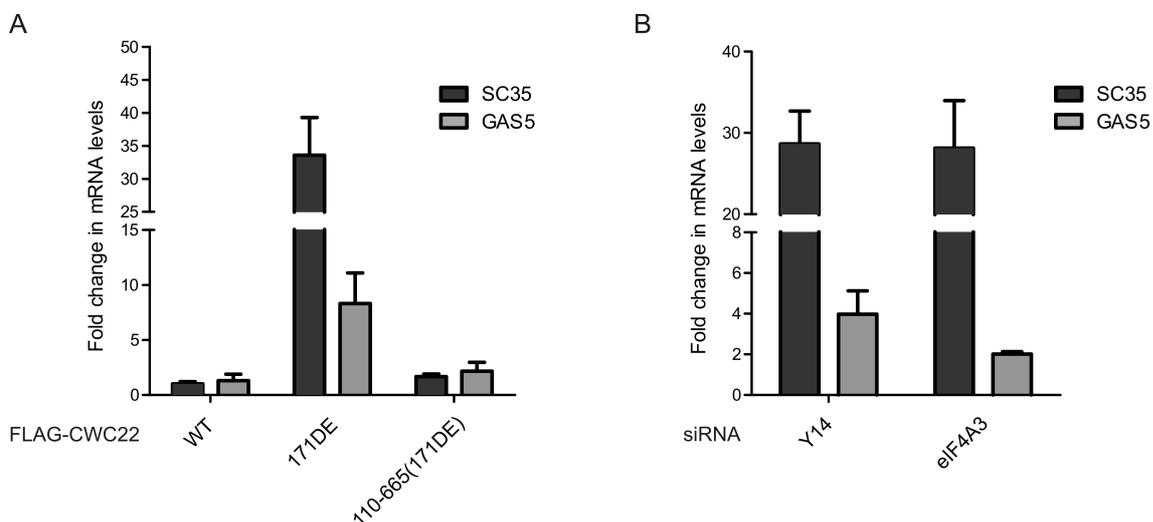


**Figure 26: Over-expression of FLAG-CWC22(NK-171/2-DE) does not affect pre-mRNA splicing.** (A) Expression of FLAG-CWC22 constructs in doxycycline-induced stable Flp-In 293 cells was assessed by immunoblotting with an antibody against FLAG. (B) Flp-IN 293 cells were induced with doxycycline and the change in pre-GAPDH/GAPDH compared to un-induced control cells was analyzed by qRT-PCR. Bars represent the mean of three independent experiments; error bars depict the standard deviation (SD).

The effect of FLAG-CWC22(NK-171/2-DE) over-expression on EJC assembly was analyzed next. Since there is no reliable assay to quantify the number of EJCs in living cells, I measured the expression of endogenous NMD targets as an indirect marker for EJC assembly. Endogenous NMD targets are naturally occurring transcripts, whose expression is regulated by the NMD machinery. In most cases, the degradation of endogenous NMD targets is initiated by an EJC downstream of the termination codon. Thus, inhibition of EJC assembly will prevent NMD and lead to an up-regulation of endogenous NMD targets.

I determined mRNA levels of the endogenous NMD targets SC35 (1.7 kb version) and GAS5, normalized to an internal control (tata box binding protein (TBP)). The 1.7 kb transcript of splicing factor SC35 (also known as SRSF2) is part of an auto-regulatory feedback loop that controls SC35 expression by PTC-activated NMD (Sureau et al., 2001). Likewise, the snoRNA-encoding pseudogene GAS5 (growth arrest-specific 5) contains a PTC and is rapidly degraded by NMD. As expected, siRNA mediated knock down of the EJC proteins Y14 and eIF4A3 resulted in a drastic increase in SC35 and GAS5-mRNA levels (Figure 27B), demonstrating that both mRNAs are negatively regulated by EJC-dependent NMD. Interestingly, over-expression of FLAG-CWC22(NK-171/2-DE), but not FLAG-CWC22 or FLAG-CWC22(110-665;NK-171/2-DE), caused an

equally strong up-regulation of SC35 (1.7kd) and GAS5 mRNA (Figure 27A). These results suggest that over-expression of CWC22(NK-171/2-DE) impairs EJC assembly, thereby causing the observed toxic effect. It is interesting to note that over-expression of the shorter version FLAG-CWC22(110-665;NK-171/2-DE) did not change the expression levels of SC35 and GAS5 mRNA. This could be due to a weaker expression of FLAG-CWC22(110-665;NK-171/2-DE) compared to FLAG-CWC22(NK-171/2-DE) (Figure 26A). Alternatively, it might indicate that only the full-length version CWC22, which very efficiently interacts with the spliceosome (Figure 12), can exert a dominant negative effect over endogenous CWC22.



**Figure 27: Over-expression of FLAG-CWC22(NK-171/2-DE) leads to an up-regulation of endogenous NMD targets similar to the effect of EJC-protein depletion.** (A) Flp-In 293 cells were induced with doxycycline and the change in SC35 (1.7kb) and GAS5 mRNA levels compared to uninduced control cells were quantified by qRT-PCR. (B) Quantification of the change in SC35 (1.7kb) and GAS5 mRNA levels upon depletion of EJC proteins Y14 and eIF4A3. All bars represent the mean of three independent experiments. Error bars = SD.

The results presented in the paragraph 4.3.3 and 4.3.5 indicated that the functions of CWC22 during pre-mRNA splicing and EJC assembly are not directly linked. An eIF4A3-binding deficient mutant of CWC22 was able to support pre-mRNA splicing in CWC22-depleted cells (Figure 24), whereas the dominant negative mutant of CWC22 inhibited EJC assembly without impairing pre-mRNA splicing. The implications of these findings for gene regulation in human cells will be discussed in the next section of this thesis.

## 5 Discussion

### 5.1 CWC22 links pre-mRNA splicing and EJC assembly

The EJC is a key regulator of mRNA metabolism that assembles preferentially on spliced mRNA in the nucleus. Even though the composition and function of EJCs have been studied extensively, the mechanism of splicing-dependent EJC assembly remained elusive. In this study, I describe the identification of CWC22 as a direct spliceosomal interaction partner of the EJC protein eIF4A3, thus establishing a first link between the splicing machinery and the EJC.

Using a combination of biochemical and cell biological assays, I could demonstrate that

- (a) CWC22 directly interacts with eIF4A3 via its MIF4G domain;
- (b) the interaction between CWC22 and eIF4A3 is required for splicing-dependent EJC assembly;
- (c) CWC22 is an essential pre-mRNA splicing factor in human cells; and
- (d) the functions of CWC22 during pre-mRNA splicing and EJC assembly can be experimentally uncoupled.

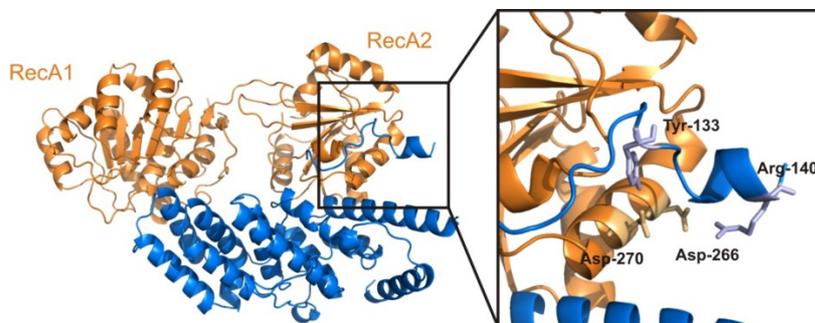
The data of this study is supported and complemented by three recent reports that describe the biochemical and structural characterization of the CWC22-eIF4A3 interaction *in vitro* and in cell extracts (Alexandrov et al., 2012; Barbosa et al., 2012; Buchwald et al., 2013).

### 5.2 The eIF4A3-CWC22 complex shows a conserved interaction pattern

#### 5.2.1 The CWC22/eIF4A3 structure confirms a eIF4G/eIF4A-like interaction

Immunoprecipitation experiments with deletion mutants of CWC22 revealed that the MIF4G domain of CWC22 directly interacts with eIF4A3 (Figure 8, Figure 9). This highly conserved domain is also found in other proteins that bind DEAD-box helicases, such as eIF4G, NOM1 and Gle1, where it mediates the interaction with eIF4A1, eIF4A3 and

Dbp5, respectively. Successful modeling of the interaction between CWC22 and eIF4A3 with an existing eIF4G-eIF4A X-ray structure suggested that both protein complexes use a similar interaction pattern (Figure 14-16). This hypothesis was supported by the recently solved crystal structure of eIF4A3 in complex with the MIF4G domain of CWC22 (Buchwald et al., 2013). Importantly, the crystal structure confirmed several of the interacting residues previously identified by site-directed mutagenesis based on the CWC22-eIF4A3 model (Figure 28-29). The MIF4G domain of CWC22 comprises amino acid residues 149–345. Similar to the MIF4G domain of eIF4G, it features 5 HEAT repeats (10  $\alpha$ -helices) that form a crescent shaped molecule. In the co-crystal structure, helix  $\alpha$ 1 and  $\alpha$ 4 of CWC22-MIF4G contact the RecA2 domain of eIF4A3, whereas helix  $\alpha$ 10 interacts with RecA1 (Buchwald et al., 2013). Moreover, an N-terminal extension (residues 123–141) of the molecule is involved in binding to eIF4A3. Tyr-133 and Arg-140 of the N-terminal extension contact Asp-270 and Asp-266 of eIF4A3, explaining why mutating Asp-270 of eIF4A3 strongly reduced the interaction between both proteins (Figure 16).

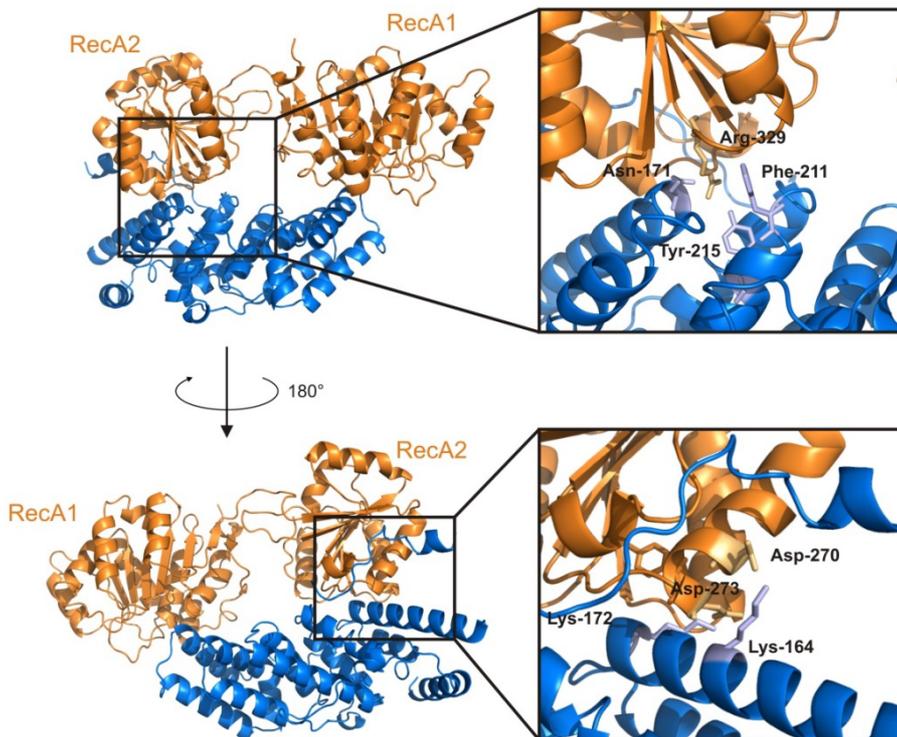


**Figure 28: An N-terminal extension of the CWC22 MIF4G domain interacts with the RecA2 domain of eIF4A3.** Crystal structure of eIF4A3 in complex with the MIF4G domain of CWC22 (PDB file 4C9B). The interacting residues are highlighted in light blue (CWC22) and light orange (eIF4A3). The interacting residue Asp-270 has been identified by homology modeling.

Furthermore, the crystal structure revealed that Asn-171 on helix  $\alpha$ 1 of CWC22 (together with Phe-211 and Tyr-215 on helix  $\alpha$ 4) contacts eIF4A3 Arg-329, whereas Lys-172 and Lys-164 of CWC22 form hydrogen bonds with Asp-270, Asp-273 and Ala-300 of eIF4A3. In line with the structural data, I could show that mutating CWC22 Asn-171 and Lys-172, as well as eIF4A3 Asp-270, Asp-273 and the amino acids adjacent to Ala-300 (NFT-301-LAG) strongly reduced the interaction between CWC22 and eIF4A3 in Co-IP and *in vitro* pulldown experiments (Figure 15-16). In summary, all point

mutants identified through homology modeling are directly involved in the interaction between CWC22 and eIF4A3, or in very close proximity to interacting residues (Figure 28, 29).

Interestingly, Asp-270 of eIF4A3 is involved in two intermolecular interactions at once. The importance of this residue is further highlighted by the fact that an Asp→Gly mutation has been found in a patient with Richieri-Costa-Pereira syndrome, an autosomal-recessive craniofacial disorder (Favaro et al., 2014).



**Figure 29: A crystal structure confirms interacting residues identified by homology modeling.** Crystal structure of eIF4A3 in complex with the MIF4G domain of CWC22 (PDB file 4C9B). The interacting residues are highlighted in light blue (CWC22) and light orange (eIF4A3). Asp-270 and Asp-273 of eIF4A3 and Lys 172 of CWC22 have been identified by homology modeling.

### 5.2.2 The MIF4G domain of CWC22 regulates the conformation of eIF4A3

The co-crystal structure revealed that the major interaction surface of eIF4A3 and CWC22 is composed of the RecA2 domain of eIF4A3 and the N-terminal helices (HEAT1 and HEAT2) of CWC22-MIF4G (Buchwald et al., 2013). This observation was further supported by ITC experiments, showing that the affinity of eIF4A3 for CWC22 is mostly determined by the RecA2 domain (Buchwald et al., 2013). These findings explain why I

was unable to identify point mutants in the RecA1 domain of eIF4A3 or the C-terminal part of CWC22-MIF4G that impair the interaction between both proteins (Figure 21C and Supplementary Figure 1). Whereas strong interactions anchor the RecA2 domain of eIF4A3 to the N-terminal helices of CWC22-MIF4G, additional weak interactions between the RecA1 domain of eIF4A3 and helix  $\alpha$ 10 of CWC22-MIF4G determine the relative orientation of the RecA-like domains within the complex. By interacting with both RecA-like domains of eIF4A3, the MIF4G domain fixes the helicase in an extended conformation (Buchwald et al., 2013). A similar mechanism has been described for the interaction of eIF4G with eIF4A and Gle1 with Dbp5 (Montpetit et al., 2011; Schutz et al., 2008). In all three cases, the N-terminal helices of the MIF4G domain bind the RecA2 domain of the helicase, whereas the C-terminal helices are involved in a 'latching' mechanism that determines the orientation of RecA1 relative to RecA2. But while eIF4G and Gle1 stabilize their respective DEAD-box helicases in an activated conformation, CWC22-MIF4G has the opposite effect on eIF4A3. Upon binding to CWC22-MIF4G, the RecA-like domains of eIF4A3 are rotated approximately 90° with respect to each other, adopting a conformation that is incompatible with RNA and ATP binding (Buchwald et al., 2013). It has been suggested that this conformation is responsible for the inhibitory effect of CWC22 on the ATPase activity of eIF4A3 observed *in vitro* (discussed in chapter 5.3 (Barbosa et al., 2012; Buchwald et al., 2013)).

Taken together, eIF4G, Gle1 and CWC22 use similar mechanisms to interact with eIF4A, Dbp5 and eIF4A3 respectively, suggesting that the overall mode of interaction between DEAD-box helicases and their MIF4G interaction partners is conserved. Subtle differences in the 'latching' mechanism determine whether the interaction has a stimulatory or inhibitory effect on the helicase (Buchwald et al., 2013; Montpetit et al., 2011; Schutz et al., 2008). Despite the overall conserved interaction pattern, MIF4G-domain proteins display a high degree of specificity for their respective helicases. This is exemplified by the fact that I was unable to co-precipitate eIF4A3 together with CBP80 or eIF4A1 together with CWC22 (Figure 7). Small variations at the interaction surface confer specificity and thus ensure that only the appropriate helicase is regulated.

### 5.2.3 The MIF4G-domain emerges as a common regulator of RNA helicases

Considering the large number of DExD/H-box helicases involved in RNA metabolism, it is intriguing to speculate that other RNA helicases are regulated in a similar way. Several proteins involved in mRNA metabolism contain one or more MIF4G domains, among them the NMD factor Upf2 (3 MIF4G domains), the cap-binding protein CBP80 (1 MIF4G and 2 MIF4G-like domains), the CBP80/20-dependent translation initiation factor (CTIF, 1 MIF4G domain) and the polyadenylate-binding protein-interacting protein 1 (PAIP1, 1 MIF4G domain). The interaction of PAIP1 with the DEAD-box helicase eIF4A1 exerts a stimulatory effect on translation initiation (Craig et al., 1998). The crystal structure of the MIF4G domain of PAIP1 revealed that it adopts a similar fold as the MIF4G domain of eIF4G (Lei et al., 2011). Most importantly, the major eIF4A-interaction surface on helix  $\alpha$ 1 of eIF4G is conserved in PAIP1, suggesting that both proteins use equivalent residues to interact with the DEAD-box helicase (Lei et al., 2011). Further studies will be required to verify whether PAIP1 uses a similar mechanism as eIF4G to regulate the ATPase activity of eIF4A1. So far, no direct DExD/H-box interaction partners have been described for Upf2, CBP80 or CTIF, but their intimate connection to mRNP remodeling makes them likely candidates for the regulation of RNA helicases. Moreover, it is conceivable that some MIF4G-domain proteins have more than one target helicase. In particular, CWC22 might be involved in the regulation of spliceosomal DExD/H-box helicases. In yeast, Cwc22 controls the DEAH-box helicase Prp2 by coupling ATP hydrolysis to spliceosomal remodeling (Yeh et al., 2011). To date, no similar function has been described for human CWC22, but the high degree of conservation between yeast and human spliceosome suggests that CWC22 might also regulate the human Prp2 homolog DHX16. Further studies will be required to identify novel MIF4G-DExD/H-box interaction pairs. The available structural information on eIF4G-eIF4A, Gle1-Dbp5 and CWC22-eIF4A3 complexes will likely help to generate models for the interaction and aid the structure-guided mutagenesis as described for CWC22 and eIF4A3 in this study. The discovery of other MIF4G-proteins and their respective helicases will expand our understanding of the regulatory network that controls the spatial and temporal activation of RNA helicases during mRNP remodeling and eukaryotic gene expression.

### 5.3 CWC22 mediates splicing-dependent EJC assembly

The direct interaction of eIF4A3 with the spliceosomal protein CWC22 pointed to a role for CWC22 during EJC assembly. And indeed, I was able to show that eIF4A3 mutants that are unable to interact with CWC22 fail to form EJCs during *in vitro* splicing reactions (Figure 18). This effect was strictly splicing-dependent, since the same point mutations did not impair the assembly of EJCs from recombinant components (Figure 17). Taken together, these results strongly suggest that CWC22 has a prominent role during EJC assembly. This hypothesis was further supported by two recent reports (Alexandrov et al., 2012; Barbosa et al., 2012). They showed that co-expression of recombinant CWC22 promotes the incorporation of eIF4A3 into splicing-dependent EJCs (Barbosa et al., 2012). Furthermore, knockdown of CWC22 impaired the interaction of eIF4A3 with cellular mRNAs and triggered the up-regulation of endogenous NMD targets (Alexandrov et al., 2012; Barbosa et al., 2012). Finally, mutating Thr-304 on eIF4A3 decreased the affinity for spliced mRNA, an effect that was exacerbated by the co-expression of mutant CWC22(G168Y) (Alexandrov et al., 2012).

#### 5.3.1 CWC22 recruits eIF4A3 to the spliceosome

CWC22 is an abundant component of the activated spliceosome and spliceosomal C-complex (Bessonov et al., 2008). Together with the finding that CWC22-binding deficient eIF4A3 mutants fail to interact with early splicing intermediates (Figure 18B), this suggested that CWC22 mediates EJC assembly by recruiting eIF4A3 to the early spliceosome. To carry out this function, CWC22 needs to interact with the splicing machinery. The domains of CWC22 that mediate spliceosomal interaction are not well defined, but *in vitro* splicing experiments revealed that only full-length CWC22 interacted strongly with the splicing machinery (Figure 12). Whereas CWC22(110-665) showed residual binding to splicing intermediates, all shorter pieces of CWC22 completely failed to interact with the spliceosome (Figure 12). In line with this observation, recombinant GST-CWC22(110-409), which binds eIF4A3 but does not interact with the splicing machinery, had a dominant negative effect on splicing-dependent EJC assembly (Figure 20). In contrast, a slightly larger piece of CWC22, comprising MIF4G and MA3 domain (amino acid residues 100-665), was able to promote EJC assembly in cell extracts, demonstrating that in an over-expression

system, the residual spliceosomal interaction of CWC22(100-665) is sufficient to recruit eIF4A3 to the spliceosome (Barbosa et al., 2012). Notably, size fractionation of cell lysates revealed that CWC22 and eIF4A3 exist in large RNase-resistant protein complexes together with the spliceosomal protein Prp19 (Barbosa et al., 2012). It will be interesting to determine which other proteins associate with CWC22 and eIF4A3. The identification of spliceosomal interaction partners of CWC22 will be instrumental in unraveling the mechanism of eIF4A3 recruitment and EJC assembly.

### **5.3.2 The core EJC proteins are recruited to the spliceosome via different routes**

In cell extracts, CWC22 does not interact with the MAGOH/Y14 heterodimer or BTZ (Figure 10A, C), supporting the prior notion that the EJC core components are recruited to the spliceosome via separate routes (Gehring et al., 2009a). Whereas BTZ might be able to associate with the pre-EJC after disassembly of the spliceosome, incorporation of MAGOH/Y14 into EJCs is strictly splicing-dependent (Gehring et al., 2009a). To fully understand the mechanism of splicing-dependent EJC assembly, it is therefore important to identify factors that mediate the spliceosomal recruitment of MAGOH/Y14. Immunoprecipitation of endogenous EJCs identified a large number of interacting proteins, among them many spliceosomal proteins and splicing factors (Singh et al., 2012). Especially SR proteins were found to be over-represented in the EJC interactome. However, the high complexity of the splicing machinery hampers the identification of individual proteins that might mediate the spliceosomal recruitment of MAGOH/Y14. In the future, more detailed interaction studies and *in vitro* binding assays will be required to identify direct spliceosomal interaction partners of MAGOH and Y14.

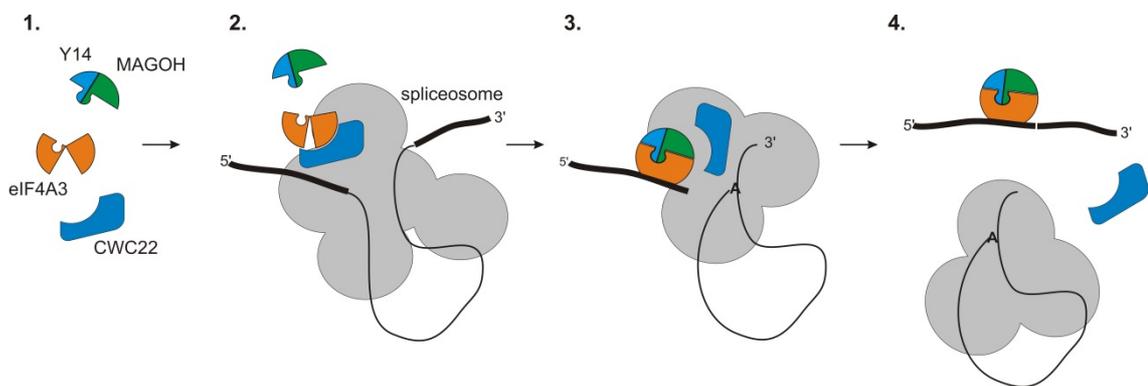
### **5.3.3 EJC assembly requires a rearrangement of the eIF4A3-CWC22 interaction**

*In vitro* pulldown experiments revealed that binding of eIF4A3 to MAGOH/Y14 and CWC22 is mutually exclusive (Figure 10B). This observation implies that during EJC assembly, the interaction between CWC22 and eIF4A3 has to be resolved in order to allow binding of MAGOH/Y14. Accordingly, CWC22 is not found in post-splicing mRNPs (Figure 11A). *In vitro*, MAGOH/Y14 and CWC22 compete for the binding to eIF4A3, so that an excess of recombinant CWC22 impairs the assembly of recombinant EJCs

(Figure 19), while the presence of MAGOH/Y14 reduces the affinity of eIF4A3 for CWC22 (Barbosa et al., 2012). It remains to be elucidated whether a similar competition mechanism mediates EJC assembly in living cells. It should be noted that MAGOH/Y14 only interact with the closed conformation of eIF4A3, whereas CWC22 holds eIF4A3 in an open conformation (Andersen et al., 2006; Bono et al., 2006; Buchwald et al., 2013). It is possible that MAGOH/Y14 play an active role in displacing CWC22 from eIF4A3. However, because the heterodimer can only interact with the closed helicase, this would require eIF4A3 to switch from open to closed conformation while still bound to CWC22. Binding of MAGOH/Y14 to eIF4A3 would then displace CWC22 and fix the assembled complex on the mRNA. Alternatively, an additional (spliceosomal) factor might regulate the rearrangement. This factor would actively break the interaction between CWC22 and eIF4A3, thereby allowing the helicase to adopt its close conformation and facilitating the binding of MAGOH/Y14. Many spliceosomal proteins could be involved in remodeling the eIF4A3-CWC22 interaction during splicing. In particular, spliceosomal DExD/H-box helicases might compete with eIF4A3 for CWC22-binding. An interesting candidate is the spliceosomal DEAH-box helicase Prp22. Prp22 is involved in the late stages of splicing by mediating exon ligation and the release of mRNA from the intron-lariat-spliceosome (ILS) (Company et al., 1991; Fourmann et al., 2013; Schwer, 2008). Moreover, it has been suggested that the ATPase activity of Prp22 (HRH1 in human cells) is required for the assembly of EJCs on spliced mRNA (Zhang and Krainer, 2007). In line with a possible role for Prp22 in breaking the association between eIF4A3 and CWC22, is the observation that Prp22 triggers the release of CWC22 from the post-catalytic spliceosome (Fourmann et al., 2013). It is therefore conceivable that the individual EJC components are recruited to the spliceosome prior to exon ligation (as part of the C-complex), but stable complex assembly takes place after the second catalytic step of splicing, upon Prp22-mediated release of CWC22 from eIF4A3. Even though this model would provide a plausible explanation for the role of Prp22 during EJC assembly, it is difficult to reconcile with previous data showing that MAGOH/Y14 require the interaction with eIF4A3 and the EJC deposition site to bind the early spliceosome (Gehring et al., 2009a). Further studies will be required to unravel how and when the EJC assembles on spliced mRNA

and whether DExD/H-box helicases or additional spliceosomal proteins play a regulatory role in the process.

Based on the current knowledge, I propose a four-step model for the CWC22-mediated assembly of EJCs (Figure 30). The initial event in EJC assembly is the binding of eIF4A3 to CWC22 (step 1). It remains to be elucidated whether both proteins engage the spliceosome individually or as a heterodimer. The spliceosomal interactions of CWC22 anchor eIF4A3 to the splicing machinery (step 2). In the following step, the interaction between CWC22 and eIF4A3 is broken to allow MAGOH/Y14 to bind to the helicase (step 3). Finally, spliceosomal remodeling leads to the release of the intron-lariat spliceosome and CWC22 from the spliced mRNA (step 4).

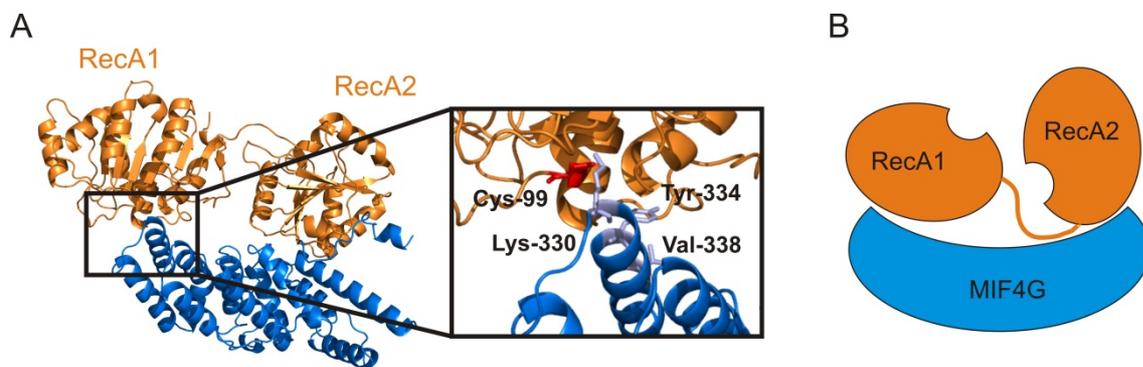


**Figure 30: A 4-step model for the CWC22-mediated assembly of EJCs.** For further description, see paragraph 5.3.3.

#### 5.3.4 The ATPase activity of eIF4A3 is dispensable for spliceosomal recruitment

A recent study showed that CWC22 reduces the affinity of eIF4A3 for ssRNA *in vitro* (Barbosa et al., 2012). Likewise, CWC22 inhibits the RNA-dependent ATPase activity of eIF4A3 (Barbosa et al., 2012). It was suggested that CWC22 prevents RNA-binding and ATP hydrolysis by fixing eIF4A3 in an inactive conformation (Buchwald et al., 2013). This inactive conformation was proposed to prevent premature RNA binding of eIF4A3. It should be noted however, that the ATPase activity of eIF4A3 in the absence of BTZ is very weak (Ballut et al., 2005). Moreover, the weak latching interaction between the RecA1 domain of eIF4A3 and the MIF4G domain of CWC22 might allow the helicase to adopt its closed conformation while still being tightly anchored to CWC22 via its RecA2

domain (Buchwald et al., 2013). It therefore remains to be elucidated whether the inhibition of ATPase activity is required for EJC assembly in living cells. Notably, mutating Cys-99 on the RecA1 domain of eIF4A3 did not affect the incorporation of FLAG-eIF4A3 into splicing-dependent EJCs (Supplementary Figure 4). In the crystal structure, Cys-99 contacts Lys-330, Tyr-334 and Val-338 of CWC22, thus composing the main CWC22-interacting residue of the RecA1 domain (Figure 31 (Buchwald et al., 2013)). It remains to be experimentally verified whether the eIF4A3(C-99-Q) mutant is indeed immune to ATPase inhibition through CWC22. However, the observation that this mutant displays no EJC assembly defect is a first hint that ATPase inhibition might not be required for EJC assembly *in vivo*.



**Figure 31: Interactions between the RecA1 domain of eIF4A3 and helix  $\alpha$ -10 of CWC22 fix the helicase in an inactive conformation.** (A) Crystal structure of eIF4A3 in complex with the MIF4G domain of CWC22 (PDB file 4C9B). Cys-99 of eIF4A3 is depicted in red, the interacting residues of CWC22 are highlighted in light blue. (B) Scheme of the relative orientation of the RecA-like domains of eIF4A3 in complex with the MIF4G domain of CWC22. Note the solvent-exposed ATP-binding site of RecA1.

Whereas the importance of ATPase inhibition is still under debate, there is ample evidence that CWC22 does not stimulate the ATPase activity of eIF4A3 (Barbosa et al., 2012; Buchwald et al., 2013). This is in line with a previous report, showing that ATP hydrolysis is not required for eIF4A3-incorporation into spliceosomes (Shibuya et al., 2006). Instead of regulating the ATPase activity of eIF4A3, CWC22 might simply function as an adapter protein that mediates the association of eIF4A3 with the spliceosome.

## 5.4 CWC22 is required for pre-mRNA splicing in human cells

Cwc22 was originally described in *Saccharomyces cerevisiae* as one of 26 proteins that associate with the splicing factor Cef1/Ntc85 (named CWC for complexed with Cef1) (Ohi and Gould, 2002). Cef1/Ntc85 is an integral component of the NTC (nineteen complex) that is named for its association with the spliceosomal protein Prp19. The NTC is involved in spliceosomal activation by stabilizing the interaction of U5 and U6 snRNPs with the pre-mRNA (Chan and Cheng, 2005). A more recent study reported that Cwc22 is not tightly associated with the NTC, but requires the presence of the NTC for the stable association with the spliceosome (Yeh et al., 2011). Like the NTC, Cwc22 associates with the pre-catalytic spliceosome and stays associated with the splicing machinery until the completion of the catalytic reaction (Yeh et al., 2011). In yeast, Cwc22 functions in the first catalytic step of splicing by coupling the ATP hydrolysis of the DEAH-box helicase Prp2 to the release of the U2 components SF3a and SF3b from the branch point sequence of the pre-mRNA (Yeh et al., 2011). In human cells, CWC22 is found in the activated spliceosome and spliceosomal C-complex, but the exact function during pre-mRNA splicing remained elusive (Bessonov et al., 2010; Bessonov et al., 2008).

Using siRNA-mediated depletion of CWC22, I could show that CWC22 is an essential splicing factor in human cells (Figure 22). Knockdown of CWC22 led to severe up-regulation of pre-mRNAs and concomitant reduction of mature mRNA levels. This observation was supported by an independent report that showed a down-regulation of spliced mRNA upon CWC22 depletion (Alexandrov et al., 2012). In both cases, the effect could be rescued by over-expressing siRNA-resistant CWC22, which confirmed the specificity of the knockdown. Surprisingly, a third report did not observe a splicing defect upon CWC22 depletion (Barbosa et al., 2012). This discrepancy might be due to an insufficient knockdown in the latter study.

### 5.4.1 MIF4G and MA3 domain are required for pre-mRNA splicing in humans

Interestingly, a fragment of CWC22 comprising of MIF4G and MA3 domain was required for pre-RNA splicing in human cells (Figure 23). This is in contrast to data obtained in yeast, where a segment of Cwc22 containing the MA3 domain and a short

conserved downstream region is sufficient to support pre-mRNA splicing and cellular growth (Yeh et al., 2011). The disparity in domain requirement demonstrates that, even though the overall role of CWC22 in pre-mRNA splicing is conserved from yeast to humans, the mechanism of splicing regulation differs between species. In yeast, the function of CWC22 during splicing has been linked to the regulation of the DEAH-box helicase Prp2 (Yeh et al., 2011). It will be interesting to determine if this function requires a direct interaction between the MA3 domain of Cwc22 and Prp2, and whether the mechanism is conserved in human cells. And if so, is binding of CWC22-MA3 to Prp2 compatible with a concomitant binding of CWC22-MIF4G to eIF4A3? Yeast cells do not form splicing-dependent EJCs and accordingly, eIF4A3 is not found in the yeast spliceosome (Fabrizio et al., 2009). It is tempting to speculate that the requirement for the MIF4G domain during pre-mRNA splicing in human cells, but not in yeast, is directly linked to the function of CWC22-MIF4G during EJC assembly. However, my finding that a CWC22-mutant, which is unable to interact with eIF4A3, can rescue splicing in CWC22-depleted cells, implies that the MIF4G domain might have an EJC-independent role during pre-mRNA splicing (Figure 24). Notably, a segment of yeast Cwc22 containing MIF4G and MA3 domain, but not the conserved downstream sequence, supported cellular growth in yeast, suggesting a functional redundancy between the MIF4G domain and the C-terminal extension of the MA3 domain (Yeh et al., 2011). Further studies will be required to unravel the role of CWC22 during pre-mRNA splicing in yeast and humans. In this regard, it will be particularly interesting to identify direct interaction partners of MIF4G and MA3 domain. Both eIF4A3 and Prp2 belong to the family of DExD/H-box helicases. Considering the conserved function of MIF4G and MA3 domains in regulating RNA helicases, it is reasonable to speculate that other DExD/H-box helicases might interact with CWC22. Eight conserved DExD/H-box helicases are involved in pre-mRNA splicing in human and yeast, all of which represent possible candidates for an interaction with CWC22.

It will be equally important to understand whether MIF4G and MA3 domain have cooperative or independent functions during pre-mRNA splicing. A cooperative regulation has been described for the protein eIF4G, where both MIF4G and MA3 domain are involved in regulating eIF4A (Marintchev et al., 2009). Alternatively, MIF4G

and MA3 domain might control different steps of pre-mRNA splicing. Of note, CWC22 binds to the pre-catalytic spliceosome and remains associated with the splicing machinery throughout the whole splicing reaction (Fourmann et al., 2013). It is therefore conceivable that CWC22 regulates several consecutive steps of splicing.

#### **5.4.2 N- and C-terminal domains of CWC22 regulate spliceosomal recruitment**

Whereas the central part of CWC22 is highly conserved, the N- and C-terminal regions are more variable, and their size can differ considerably between species. Interestingly, the N- and C-terminal domains of CWC22 were not required to rescue pre-mRNA splicing in CWC22-depleted cells (Figure 23). This is surprising, since only the full-length protein successfully bound the spliceosomal C-complex (Figure 12). Most likely, over-expression of the truncated protein could compensate for the reduced spliceosomal affinity. Nevertheless, it will be interesting to determine how additional sequences in the N- and C-terminus regulate the interaction of CWC22 with the spliceosome. Of note, a dominant negative point mutation of CWC22 only impaired cell viability in context of the full-length protein, suggesting that the truncated version is unable to displace endogenous CWC22 (Figure 25).

A yeast-two-hybrid screen (Y2H) recently identified two putative spliceosomal interaction partners of CWC22 (Hegele et al., 2012). Like CWC22, FRG1 is a component of the activated spliceosome ( $B^{act}$ ), whereas FAM32A is found in the spliceosomal C-complex. To date, the functions of both proteins remain elusive. Further research will be required to verify the interaction of FAM32A and FRG1 with CWC22 and to determine whether they cooperate with CWC22 to regulate pre-mRNA splicing and EJC assembly.

Pre-mRNA splicing and EJC assembly are closely linked processes. But whereas it is well established that splicing promotes EJC assembly, it remains unknown to what extent the EJC influences splicing. The interconnection of both events will be discussed in the following chapter.

## **5.5 The roles of CWC22 in pre-mRNA splicing and EJC assembly can be functionally uncoupled**

### **5.5.1 Pre-mRNA splicing and EJC assembly are closely linked processes**

Recently it was discovered that EJCs do not only regulate post-splicing mRNA metabolism, but can also influence the splicing process itself. Initially described for a small subset of genes in *Drosophila*, EJC proteins have also been found to regulate the splicing of certain mRNAs in *Xenopus laevis* and in human cells (Ashton-Beaucage et al., 2010; Haremakei and Weinstein, 2012; Michelle et al., 2012; Roignant and Treisman, 2010). To date it is unknown how EJC proteins influence splicing and whether this function represents a general mechanism or is restricted to specific mRNAs. The dual function of CWC22 in EJC assembly and pre-mRNA splicing suggested a possible role for this protein in linking both processes. It has been proposed that EJC proteins might influence splicing by stabilizing the interaction between the spliceosome and the mRNA (Ashton-Beaucage et al., 2010). Accordingly, a direct interaction between the splicing factor CWC22 and RNA-bound eIF4A3 could facilitate the detection of 5' splice sites in a mechanism known as exon definition (Fox-Walsh et al., 2005). It should be noted however, that several nuclear EJC proteins regulate pre-mRNA splicing, but only eIF4A3 directly interacts with CWC22. This implies that additional factors are involved, or else the splicing function of EJC proteins is independent of CWC22.

### **5.5.2 EIF4A3 does not influence the splicing function of CWC22**

The MIF4G domain of CWC22 is involved in both eIF4A3 binding and pre-mRNA splicing. In order to analyze whether both functions are directly linked, I aimed to experimentally uncouple pre-mRNA splicing from EJC assembly. To this end, endogenous CWC22 was replaced by a mutant protein (NK-171/2-DE), which is unable to bind eIF4A3 (Figure 24). Interestingly, the CWC22-mutant was as efficient in restoring pre-mRNA splicing in CWC22-depleted cells as the wild type protein. This demonstrates that the interaction between CWC22 and eIF4A3 is not required for pre-mRNA splicing, and implies that the MIF4G domain has additional, EJC-independent functions. Consistent with this finding, the interaction with eIF4A3 was not required for the incorporation of CWC22 into the spliceosome (Figure 21). Taken together,

these results show that eIF4A3 does not significantly influence the spliceosomal recruitment or function of CWC22.

### **5.5.3 The selective inhibition of EJC assembly is toxic in human cells**

Notably, the complementation assays with mutant CWC22 could only be performed with a truncated version of CWC22 (amino acid residues 110-665), because the mutated full-length protein had a dominant negative effect on cell viability (Figure 25). The toxic effect was most likely caused by an inhibition of EJC assembly, as shown by a drastic up-regulation of endogenous NMD targets upon expression of CWC22(NK-171/2-DE) (Figure 27). Even though CWC22(110-665; NK-171/2-DE) is equally deficient in binding eIF4A3, it had no negative effect on cell viability and EJC assembly. This can be explained by the lower affinity of truncated CWC22 for the spliceosome (Figure 12). Although over-expressed CWC22(110-665; NK-171/2-DE) can replace the function of endogenous CWC22 during pre-mRNA splicing, it is less efficiently recruited to the splicing machinery. Residual endogenous CWC22 could therefore maintain basal EJC levels and prevent the cytotoxic effect.

To date it remains unknown why the inhibition of EJC assembly causes cytotoxicity. A recent report linked the growth arrest observed upon NMD inhibition to the up-regulation of the non-coding RNA GAS5 (growth-arrest specific transcript 5) (Mourtada-Maarabouni and Williams, 2013). Importantly, I observed a significant up-regulation of GAS5 upon expression of CWC22(NK-171/2-DE) (Figure 27). Considering the key role of EJCs during mammalian gene expression, it is very likely that additional cellular processes are impaired upon CWC22(NK-171/2-DE) expression. On note, the growth arrest was already apparent after 12-24 hours, suggesting that fast cellular processes dictate the toxic response to CWC22(NK-171/2-DE) expression.

Interestingly, expression of CWC22(NK-171/2-DE) caused an increase of SC35 and GAS5, but did not affect the mRNA levels of another NMD target, called UHG (Figure 27 and Supplementary Figure 5A). In contrast, the knockdown of CWC22 led to an up-regulation of UHG, but had a less pronounced effect on SC35 and GAS5 (Supplementary Figure 5B (Alexandrov et al., 2012)). In the latter case, impaired pre-mRNA splicing most likely masked the effect of defective EJC assembly. This example highlights the importance of distinguishing between the two functions of CWC22 and

demonstrates that EJC inhibition cannot be faithfully recapitulated by depleting CWC22. In contrast, expressing the dominant negative CWC22(NK-171/2-DE) caused a similar effect as the depletion of the integral EJC proteins eIF4A3 and Y14 (Figure 27). This confirms that the CWC22-mutant can indeed be used to selectively disrupt EJC assembly without changing the expression levels of endogenous EJC proteins.

## **5.6 Concluding remarks**

In this study I presented the identification and characterization of CWC22 as an essential splicing factor that mediates EJC assembly in human cells. The finding that CWC22 directly binds the core EJC protein eIF4A3 and mediates its recruitment to spliced mRNA is a first step to understanding the splicing-dependant nature of EJC assembly. Nevertheless, more extensive research is required to fully unravel the molecular mechanism of EJC assembly in living cells. In particular, it will be important to identify factors that recruit the EJC proteins MAGOH and Y14 to the spliceosome. Moreover, it remains to be elucidated how and when the rearrangements of the catalytic spliceosome initiate the EJC assembly on spliced mRNAs. Finally, the current model does not explain why EJCs assemble invariantly 20-24 nt upstream of the exon-exon junction. It will therefore be especially interesting to determine the factors that regulate the spacial control of EJC assembly. A comprehensive molecular model of splicing-dependent EJC assembly will eventually help to explain how mammalian cells integrate the individual processing steps of gene expression.

Apart from identifying a key regulator of splicing-dependent EJC assembly, my work has uncovered a novel DEAD-box/MIF4G-protein interaction pair. Together with previous data on similar protein complexes, this suggests a conserved role for MIF4G domain proteins in regulating the function of RNA helicases. This study is a good example for the identification and characterization of novel interaction partners based on homology modeling. With regard to the structural conservation of the described protein-protein interaction, it is conceivable that other interactions can be found in a similar way. The identification of additional MIF4G proteins and their DExD/H-box interaction partners could explain the spacial and temporal control of RNA helicases and thereby extend our knowledge on mammalian gene expression regulation.

## **6 Material and Methods**

### **6.1 Plasmids**

#### **6.1.1 Mammalian expression plasmids**

For protein expression in mammalian cell culture, modified versions of the pCI-neo expression vector (Promega) were used, that harbor an N-terminal FLAG (DYKDDDDK) tag (pCI-FLAG) or N-terminal V5 (GKPIPPLLGLDST) tag (pCI-V5). Transgene expression was initiated by the CMV immediate-early enhancer/promoter region upstream of the multiple cloning site.

#### **6.1.2 Bacterial expression plasmids**

Strep-tagged proteins were expressed from pET51b (Novagen), GST-tagged proteins from pGEX6p3 (GE Healthcare) and the pETDuet-1 vector (Novagen) was used for the co-expression of two proteins (MAGOH and Y14). All three vectors contain an ampicillin resistance gene for positive selection and the T7 system that allows controlled protein expression upon induction with IPTG (Isopropyl- $\beta$ -D-thiogalactopyranosid).

#### **6.1.3 Plasmids for in vitro transcription**

The templates for in vitro transcription of radiolabeled mRNA were cloned with EcoRI/BamHI into pGEM-4Z (Promega). pGEM-4Z harbors an SP6 promotor on the 5' side of the inserted template DNA.

### **6.2 Mammalian cell culture**

#### **6.2.1 Mammalian cell lines**

All mammalian cell culture experiments were performed in HeLa Tet-Off (HTO) cells (Clontech) or human embryonic kidney (HEK) 293 cells (life technologies). The cells were grown in Dulbecco's Modified Eagle Medium (DMEM) GlutaMAX (Gibco),

supplemented with 1% penicillin/streptomycin and 10% fetal bovine serum (FBS) at 37°C and 5% CO<sub>2</sub>. The cells were passaged every 2-3 days (at approximately 90% confluency) by trypsinizing.

### 6.2.2 Generation of stable cell lines

HEK293 cells stably expressing doxycycline-inducible FLAG-tagged proteins were generated using the Flp-In™ system and Flp-In™ T-REx™ 293 cells from life technologies. This system allows the targeted integration of a gene of interest into a single site of the host genome. Integration of the new gene is achieved through FLP recombinase mediated homologous recombination between a single Flp Recombination Target (FRT) site in the genome of the host cell line and two FRG sites flanking the gene of interest in the vector pcDNA5/FRT/TO. A hygromycin resistance gene is inserted into the host genome together with the gene of interest, allowing the selection for successful recombination on the basis of hygromycin resistance.

To generate stable cell lines, FLAG-tagged CWC22 or the respective mutants were cloned between the FRT sites of pcDNA5/FRT/TO using XhoI and NotI restriction enzymes. The siRNA CWC(A) targeting sequence (5'-AAAAGTAGTGTGGCACAGATAAAA-3') (codons 2-9 of the CWC22 ORF) was replaced by the resistant sequence 5'-AAATCATCAGTGGCCCAAATCAAA-3' to allow the rescue of siRNA-mediated CWC22-knockdown. 2x10<sup>5</sup> 293 Flp-In T-REx cells (life technology) were seeded into 6-well plates containing 2 ml DMEM (supplemented with FBS and antibiotics). 24 hours after seeding, the cells were transfected with 0.5 µg pcDNA5/FRG/TO, 4.5 µg pOG44 (the expression vector for the FLP recombinase) and 0.3 µg pCI-Venus as a transfection control. 24 hours after transfection, the cells were splitted 1:3 into new 6-well plates containing fresh medium supplemented with 150 µg/ml Hygromycin B. Selection in hygromycin-containing medium was continued for approximately 3 weeks until single hygromycin-resistant colonies became visible in the cell culture dish. Single colonies were picked, expanded and tested for expression of the FLAG-tagged protein by western blot analysis. To induce protein expression, the Flp-In cells were cultured in the presence of 1 µg/ml doxycycline for 48 hours.

### 6.2.3 Transfection of plasmid DNA

Plasmid DNA was transfected with the BBS/Calciumphosphate method. Cells were seeded 24 hours prior to transfection at the appropriate density to achieve ~60% confluency on the next day (see Table). The medium was changed once 2-6 hours before the transfection. To prepare the transfection mix, the plasmid DNA was diluted in ultra pure water (90  $\mu$ l, 180  $\mu$ l or 450  $\mu$ l for transfections in 6-well plates, 6cm dishes or 10cm dishes, respectively). CaCl<sub>2</sub> was added to the diluted DNA to a final concentration of 250 mM. The DNA-CaCl<sub>2</sub> solution was mixed with an equal volume of 2xBBS and vortexed vigorously. After 15 minutes incubation at room temperature (RT), the transfection mix was added to the cell culture dish and distributed by gentle shaking. On the following day, the medium was changed to remove calcium-phosphate precipitates. The cells were harvested 48 hours after transfection. The amount of transfected plasmid DNA depended on the particular experiment and will be indicated for each type of experiment separately. A plasmid expressing mVenus (pCI-mVenus) was included in each experiment and served as a transfection control.

#### Overview of transfection mix

	6-well plate	6cm dish	10cm dish
<b>Number of cells</b>	2.4x10 <sup>5</sup>	6.5x10 <sup>5</sup>	1.3x10 <sup>5</sup>
<b>Volume of medium (ml)</b>	2	4	10
<b>Volume of H<sub>2</sub>O + DNA (<math>\mu</math>l)</b>	90	180	450
<b>Volume of 2.5M CaCl<sub>2</sub> (<math>\mu</math>l)</b>	10	20	50
<b>Volume of 2xBBS</b>	100	200	500
<b>Volume of total mix (<math>\mu</math>l)</b>	200	400	1000

#### 2x BBS

50 mM        BES (N, N-bis(Hydroxyethyl)-2-Aminoethansulfonate)  
 1.5 mM        Na<sub>2</sub>HPO<sub>4</sub>  
 280 mM        NaCl  
 adjust pH to 6.96

#### 6.2.4 Transfection of siRNA

For transfection of siRNAs, approximately  $2 \times 10^5$  cells were seeded in 6-well plates 24 hours prior to transfection. On the next day the medium was changed to DMEM without antibiotics and siRNAs were transfected with Lipofectamine RNAiMAX (life technologies) according to the manufacturers instructions. In brief, 30pmol siRNA and 5  $\mu$ l RNAiMAX were each diluted in 250  $\mu$ l OptiMEM. Both dilutions were mixed, vortexed briefly, incubated for 15 minutes at RT and added to the cells. 24 hours after transfection the medium was replaced by DMEM supplemented with FBS and antibiotics and the cells were harvested after 48 hours. siRNA targeting luciferase mRNA was used as a negative control.

#### List of siRNAs used in this study

Name	Sequence
siCWC22(A)	5'-AAAGTAGTGTGGCACAGATAA-3'
siCWC22(B)	5'-CTCGCACTGGTGGAGCATATA-3'
eIF4A3	5'- CGAGCAATCAAGCAGATCA-3'
Y14	5'-ATATGAAACATACAAGGAA-3'
Luciferase	5'-AACGUACGCGGAAUACUUCGATT-3'

#### 6.2.5 Cell viability assay

The CellTiter-Glo<sup>TM</sup> Luminescent cell viability assay (Promega) was applied to analyze the number of viable cells per well. To this end, 10000 Flp-In<sup>TM</sup> T-REx<sup>TM</sup> 293 cells expressing FLAG-CWC22, FLAG-CWC22(NK-171-DE) or no protein were seeded into 96-well plates in 100  $\mu$ l DMEM (supplemented with FBS and antibiotics). 12 hours after seeding, protein expression was induced with 1  $\mu$ g/ml doxycycline and the number of viable cells per well was determined after 0, 12, 24, 36 and 48 hours using CellTiter-Glo reagent according to the manufactures instructions. In brief, 100  $\mu$ l CellTiter-Glo<sup>TM</sup>-reagent was added directly to the cells. After shaking for 2 minutes at 700 rpm and incubation at RT for 10 minutes the luminescent signal was measured using the Centro XS<sup>3</sup> LB 960 Luminometer (Berthold Technologies). Four replicates were measured for

each condition and time point. DMEM without cells served as a control for background luminescence.

### **6.2.6 Preparation of cell lysates**

To prepare cell lysates for western blot analysis or immunoprecipitation, confluent cells were washed once with PBS and scraped from the cell culture dish in PBS. The cell suspension was transferred to 1.5 ml reaction tubes and pelleted for 5 min at 200xg. The PBS was removed, the cell pellet resuspended in IP lysis buffer (1x TBS, 0.5% Triton X-100 (Sigma-Aldrich), 50 µg/ml RNaseA (Applichem), protease inhibitor cocktail (Sigma-Aldrich #P2714-1BTL, diluted 1:200)) and the lysate frozen at -20°C for at least 4 hours.

## **6.3 Molecular biology**

### **6.3.1 Cloning**

The full length CWC22 gene was PCR-amplified from HeLa cDNA using Phusion DNA polymerase (Finnzymes). 5' XhoI and 3' NotI restriction sites were introduced with the primer sequences. The PCR product was separated through agarose gel electrophoresis and purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). The purified PCR product was digested with XhoI and NotI (NEB) over night at 37°C, purified again with the NucleoSpin Gel and PCR Clean-up kit. Using T4 ligase (Promega), the PCR product was ligated into a XhoI/NotI-linearized pCI-neo vector (Promega) containing a FLAG-tag (DYKDDDDK) (pCI-neo-FLAG). Ligated plasmids were transformed into the chemically competent *E.coli* strain XL1-blue and plasmid DNA was subsequently extracted from single clones using the ZR Plasmid Miniprep kit (Zymo Research). Deletion mutants and point mutants of CWC22 were generated by PCR, using the pCI-FLAG-CWC22 plasmid as a template. All sequences were verified by Sanger sequencing. Mammalian expression plasmids for eIF4A3, eIF4A1, CBP80, MAGOH, Y14, BTZ, and BTZ(110-372) were already present in the laboratory.

### 6.3.2 Site-directed mutagenesis

To induce point mutations in pCI-FLAG-CWC22 and pCI-FLAG-eIF4A3, a two-step PCR-based site directed mutagenesis strategy was applied. The wild type expression plasmid served as a template in both PCRs. The first PCR was performed with gene-specific primers harboring the desired point mutations and the primer pCI-antisense that anneals to the vector sequence at the 3' end of the gene. Thus, the PCR generates a product with the point mutation at the 5' end and a short sequence of vector DNA on the 3' end. This PCR product was used as a reverse primer in the second PCR. In combination with the pCI-sense primer, the second PCR generated a product that covers the whole gene and has short stretches of vector sequence on both ends. The short vector sequences contain the restriction sites that were used for the initial cloning of the gene. The PCR product was digested with XhoI/NotI and ligated into linearized pCI-neo-FLAG as described before.

### 6.3.3 Plasmid preparation

Plasmid DNA was amplified in *E.coli* XL1-blue and extracted with alkaline lysis using the ZR Plasmid Miniprep kit (Zymo Research) or the NucleoBond Xtra Midiprep kit (Macherey Nagel) according to the manufacturer's instructions. In general, 4 ml bacterial over night culture in LB medium supplemented with 100 µg/ml ampicillin was used for a mini-scale plasmid preparation and 100 ml overnight culture was used for midi-scale plasmid preparations.

### 6.3.4 RNA extraction from cell lysates

Total RNA was extracted from cells using the phenol/guanidine-based Isol-RNA Lysis Reagent™ (5 Prime). Cells were washed once with PBS and harvested directly in Isol-RNA Lysis Reagent™. 1 ml Isol-RNA Lysis Reagent™ was added to each well of a 6-well plate. The cells were lysed through pipetting up and down and the suspension was transferred to a 1.5 ml reaction tube. Phase separation was induced by adding 100 µl (1/10) 1-Bromo-3-chloropropane (BCP) followed by vigorous vortexing and 5 min incubation at RT. Afterwards, the mix was centrifuged for 10 min at 20000xg and 4°C. After centrifugation, the upper, aqueous phase (~400 µl) was transferred to a new 1.5 ml reaction tube. One volume 2-propanol was added, mixed by vortexing and incubated at least 20 min on ice. Subsequently,

the RNA was precipitated by centrifuging for 20 min at 20000xg and 4°C. The supernatant was removed and the pellet was washed twice with 1 ml 75% ethanol (centrifuging 10 minutes at 20000xg and 4°C). After the last washing step, the ethanol was removed completely and the RNA pellet was air dried for a few minutes. The pellet was resuspended in ultra pure H<sub>2</sub>O (20 µl) through shaking for 15 minutes at 700rpm and 65°C. The RNA concentration was determined by measuring the absorption at 260nm (A<sub>260</sub>) using a Nanodrop spectrophotometer (Thermo Scientific).

### **6.3.5 Removal of genomic DNA**

Prior to cDNA synthesis (chapter 6.2.6) the purified total RNA was incubated in the presence of DNase I (NEB) to remove residual contaminants of genomic DNA. 1 µl DNase (2 units) was added to up to 5 µg of total RNA and incubated for 30 min at 37°C. Afterwards, the RNA was recovered using the RNA clean and Concentrator-5 kit (Zymo Research) according to the manufacturer's instructions.

### **6.3.6 cDNA synthesis**

cDNA was generated by reverse transcribing total RNA using the ProtoScript II reverse transcriptase (NEB). 2-4 µg of total RNA (after DNase I-digest, see chapter 6.2.5) was mixed with 1 µl oligo dT-primer (dT(18), 100 µM) and 1 µl dNTPs (10 mM each) in a total volume of 12.5 µl. In order to unfold secondary structures, the mix was heated for 5 min at 70°C, followed by chilling on ice for 2 minutes. The transcription was started by adding 2 µl DTT (0.1M), 4 µl ProtoScript II reaction buffer, 1 µl ProtoScript II and 0.5 µl RNasin (Promega), followed by incubation at 42°C for one hour. Afterwards, the enzyme was heat-inactivated at 80°C for 5 minutes.

### **6.3.7 Quantitative real-time PCR**

Quantitative real time PCR (qRT-PCR) was performed using the iQ SYBR green supermix (Bio-rad) and the CFX96 touch thermal cycler. DNase I-treated cDNA served as a template for the qPCR. The qPCR was performed in a 96-well clear optical reaction plates. For each reaction 5 µl of the 2x iQ SYBR green supermix was mixed with 2 pmol of each primer and 4 µl of dilution cDNA (1:5 dilution of the cDNA described in chapter

6.2.6). The reaction mix was filled up to 10  $\mu$ l with ultra pure H<sub>2</sub>O and analyzed on a CFX96 touch thermal cycler.

Transcription program for the CFX96 touch thermal cycler

- (1) 95°C                      3 min
- (2) 95°C                      15 sec
- (3) 60°C                      15 sec                      + Plate read
- (4) 72°C                      10 sec                      →to (2) x39
- (5) Melting curve 65°C – 95°C, increment 0.5°C, 5 sec, + Plate read

After each round of amplification (after the annealing step (3)), the light emitted from the SYBR green dye bound to double-stranded DNA was detected as a direct indicator for the number of transcripts in the sample. The amplification cycle at which the fluorescence crosses a certain background level (cycling threshold, C<sub>T</sub>) directly corresponds to the number of mRNA-copies in the original cDNA sample. Differences in mRNA expression levels were calculated using the  $\Delta\Delta C_T$  method:

$$\text{Relative expression (fold change)} = 2^{-\Delta\Delta C_T}$$

with                       $\Delta C_T = C_T(\text{target}) - C_T(\text{housekeeping gene})$

and                       $\Delta\Delta C_T = \Delta C_T(\text{sample}) - \Delta C_T(\text{reference})$

A melting curve was determined after the last amplification step to ensure that the amplification products represent a homogenous population.

**List of qPCR primers used in this study**

Gene	Forward primer	Reverse Primer
GAPDH	5'-GAGTCAACGGATTTGGTCGT-3'	5'-TTGATTTTGGAGGGATCTCG-3'
Pre-GAPDH	5'-GAGCTGGGGAATGGGACT-3'	5'-TGATGGCATGGACTGTGG-3'
ACTB	5'-GGACTTCGAGCAAGAGATGG-3'	5'-AGCACTGTGTTGGCGTACAG-3'
Pre-ACTB	5'-AGAAAATCTGGCACCACACC-3'	5'-AACGGCAGAAGAGAGAACCA-3'
β-globin mRNA	5'-AAGGCTCATGGCAAGAAAG-3'	5'-ACACCAGCCACCACTTTCT-3'
β-globin pre-mRNA	5'-AGTCCAAGCTAGGCCCTTT-3'	5'-ACACCAGCCACCACTTTCT-3'
TPI mRNA	5'-AGTTCTTCGTTGGGGGAAAC-3'	5'-CCACAGCAATCTTGGGATCT-3'
TPI pre-mRNA	5'-CTGGAAGGCTCTTCGAGTTG-3'	5'-CCACAGCAATCTTGGGATCT-3'
APH 3' II	5'-ATACTTTCTCGGCAGGAGCA-3'	5'-TGAATGAACTGCAGGACGAG-3'

## 6.4 Protein biochemistry

### 6.4.1 Immunoprecipitation

Co-immunoprecipitations with anti-FLAG (M2) magnetic beads (Sigma-Aldrich) were performed to analyze the interaction of proteins in cell lysates. To this end HTO or HEK293 cells were seeded in 6cm plates and transfected on the next day with 1.5 μg of plasmid expressing the FLAG-tagged protein (in some cases together with 0.5 μg of plasmid expressing a V5-tagged protein) and 0.5 μg of plasmid expressing mVenus as a transfection control. The cells were harvested 48 hours after transfection in 400 μl IP lysis buffer (see chapter 6.1.6). The cell lysates were cleared by centrifugation (10 min, 10000xg). 12 μl anti-FLAG (M2) magnetic beads were washed once with IP wash buffer (1xTBS, 0.5% Triton X-100). 40 μl of the cleared cell lysate was set aside as input and the rest was mixed with the washed anti-FLAG magnetic beads and rotated in an overhead shaker at 4°C for two hours or overnight. Subsequently the beads were washed three times with 600 μl IP wash buffer (each time 5 min incubation in the overhead shaker at 4°C) and eluted by adding 50 μl 1x SDS loading buffer to the beads.

12 µl of the IP and 12 µl of the input were loaded onto an SDS-poly acrylamide gel for further analysis.

#### **6.4.2 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)**

SDS-polyacrylamide gels were cast with the standard mini-gel system from Bio-Rad using 1 mm combs with 10 or 15 wells. The polyacrylamide gel mixes were prepared from Rotiphorese Gel 30 (37.5:1) stock solution (Roth) diluted in 2x stacking gel buffer (0.5M Tris-HCl, pH 6.8, 0.4% SDS) or separating gel buffer (1.5M Tris-HCl, pH 8.8, 0.4% SDS). Stacking gels contained 5% polyacrylamide and separating gels 10 or 12% polyacrylamide. Polymerization was induced by adding ammonium persulfate (100 µl of a 10% stock solution per 10 ml gel mix) and TEMED (Tetramethylethylenediamine, 15 µl per 10 ml gel mix).

Protein samples were diluted in SDS-loading buffer, boiled for 3-5 minutes at 95°C and loaded onto the gel. The gels were run 1x SDS-running buffer (25 mM Tris-HCl, 20 mM Glycine, 0.1% SDS) in vertical electrophoresis chambers at 200V for approximately one hour.

#### **6.4.3 Western blot**

After SDS-PAGE, the gel was equilibrated shortly with 1x Transfer buffer and electroblotted on Hybond-ECL nitrocellulose membrane (GE Healthcare) for 1-2 h at 12 V using the semi-dry system and 1x Transfer buffer. To reduce unspecific antibody binding, the membrane was incubated in blocking solution (5% milk in 1xTBS, 0.2% Tween-20) for one hour at RT. Afterwards, the membrane was incubated with the primary antibody (diluted in blocking solution) over night at 4°C. On the following day, the membrane was washed 3x 10 min in TBST (1x TBS + 0.2% Tween 20) followed by incubation with an HRP-conjugated secondary antibody (diluted in blocking solution) for one hour at RT. Afterwards, the membrane was again washed three times with an excess of TBST. The HRP signal was detected with the western lightning plus ECL reagent (PerkinElmer) and visualized on the ImageQuant LAS-4000 (GE Healthcare) detection system.

**List of Antibodies used in this study**

Antibody	Species	Company	Dilution
anti-FLAG	rabbit	Sigma-Aldrich	1:3000
anti-V5	rabbit	QED Bioscience	1:3000
anti-eIF4A3	rabbit	by GenScript with an N-terminal peptide of eIF4A3	1:2000
anti-CWC22	rabbit	Sigma-Aldrich	1:2000
anti-Y14	rabbit	Sigma-Aldrich	1:500
anti-MAGOH	mouse	Abcam	1:2000
anti-CBP80	rabbit	Gift by Elisa Izaurralde	1:1000
anti-eIF4A1	rabbit	Abcam	1:2000
anti-Tubulin	mouse	Sigma-Aldrich	1:10000
anti-rabbit-HRP	goat	Jackson ImmunoResearch	1:10000
Anti-mouse HRP	goat	Jackson ImmunoResearch	1:10000

**6.4.4 Coomassie brilliant blue staining**

After SDS-PAGE, the gel was shortly rinsed in water and then incubated for 15 minutes in coomassie staining solution (2.5g coomassie brilliant blue R-250 in 450 ml methanol, 100 ml acetic acid, 400 ml H<sub>2</sub>O, filtered). Afterwards, the staining solution was removed and the gel incubated in coomassie destaining solution (450 ml methanol, 100 ml acetic acid, 400 ml H<sub>2</sub>O) until the background staining had completely disappeared (after 1-5 hours). The destaining solution was changed several times during the destaining process.

**6.4.5 Bacterial expression of recombinant proteins**

Codon-optimized eIF4A3 (LIFE TECHNOLOGIES) and the respective mutants of eIF4A3, with an N-terminal Strep-tag with and without an N-terminal FLAG-tag were inserted into pET51. Y14 (full length or 1-154) with a C-terminal Strep-tag and MAGOH were cloned into multiple cloning site (MCS) 1 and MCS2 of pETDuet-1, respectively. The

SELOR domain of BTZ (110-372) with a C-terminal Strep-tag with or without an N-terminal FLAG-tag was cloned into pETDuet-1. CWC22(110-409), CWC22(110-665) and CWC22(340-665) and the respective mutants were inserted into pGEX6p3.

Recombinant proteins were expressed in *E.coli* Rosetta II pLysS. Single colonies of transformed bacteria were picked and a 10 ml pre-culture grown over night at 37°C in LB medium supplemented with 100 µg/ml ampicillin. On the following day, 500 ml expression culture was inoculated with the pre-culture. For expression of eIF4A3, the expression culture was grown at 37°C until an OD<sub>600</sub> of 0.6 and then induced by adding 0.2 mM IPTG. The bacterial culture was harvested after 6 hour incubation at 37°C. All other proteins were expressed in auto-induction medium (Studier, 2005). To this end, 470 ml LB medium was supplemented with 10 ml 50x5052 (25% Glycerol, 140 mM Glucose, 292 mM α-D-lactose), 20 ml 25xM (1.25M Na<sub>2</sub>PO<sub>4</sub>, 1.25M KH<sub>2</sub>PO<sub>4</sub>, 2.5M NH<sub>4</sub>CL, 0.25M Na<sub>2</sub>SO<sub>4</sub>), 1M MgSO<sub>4</sub> and 100 µg/ml ampicillin. After inoculation with the pre-culture the auto-induction culture was incubated over night at 28°C. Cells were harvested by centrifugation at 4000xg and the cell pellet was stored at -20°C.

### 6.4.6 Affinity purification of recombinant proteins

#### *Purification of Strep-tagged proteins*

Cell pellets from 500 ml expression culture were lysed in 10 ml bacterial lysis buffer (NP buffer supplemented with 200 µg/ml Lysozyme, 50 µg/ml RNaseA, 50 µg/ml DNase I, 1:250 dilution of protease inhibitor cocktail (Sigma-Aldrich, #P2714-1BTL) and 100 µg/ml PefaBloc) and incubated for 1 hour rotating at RT. Bacterial cells were lysed by sonication with 8x 15 pulses, 30% intensity, using a Branson sonifier 250 and a 3 mm Microtip. The cell lysates were cleared by centrifuging 30 min at 15000xg and the cleared lysates loaded onto a 1 ml StrepTactin Superflow Plus Cartridge (QIAGEN) using the Äkta FPCL (GE Healthcare). The cartridge was washed with 10 column volumes NP buffer and the protein eluted in NP buffer supplemented with 2.5 mM desthiobiotin (Sigma-Aldrich). The StrepTactin cartridge was regenerated by washing with 15 column volumes NP buffer containing 1 mM HABA (2-(4-Hydroxyphenylazo)benzoic acid).

### *Purification of GST-tagged proteins*

Cell pellets were lysed in 10 ml PBS (supplemented with 200 µg/ml Lysozyme, 50 µg/ml RNaseA, 50 µg/ml DNase I, 1:250 dilution of protease inhibitor cocktail, 100 µg/ml PefaBloc and 1 mM DTT). Cell lysates were incubated at RT, sonified and centrifuged as described for the StrepTactin purification above. The cleared lysate was loaded onto a 1 ml GSTrap FF (GE Healthcare) column using the Äkta FPLC. The column was washed with 10 column volumes PBS and the protein was eluted with 10 mM glutathione in 50 mM Tris (pH 8.0).

### Buffers for protein purification

NP buffer      50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0

PBS              Dulbecco's phosphate buffered saline (Gibco)

### *Buffer exchange and protein concentration*

PD-10 desalting columns were used for buffer exchange to rEJC buffer (20 mM HEPES KOH (pH7.9), 125 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 1 mM DTT, 10% Glycerol). The proteins were concentrated by ultracentrifugation using Amicon Ultra-15 Centrifugal filter units (Millipore), snap frozen in liquid nitrogen and stored at -80°C. The protein concentration was determined by measuring the absorption at 280nm (A<sub>280</sub>) using the NanoDrop or by Bradford assay (Bio-Rad protein assay).

### **6.4.7 *In vitro* pulldown experiments**

For *in vitro* interaction studies, recombinant proteins (15 mg each) were mixed in 50 µl rEJC binding buffer (20 mM HEPES-KOH (pH7.9), 125 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 1 mM DTT, 2.5% Glycerol, 0.1% NP-40) and pre-incubated 30 min on ice. 10% of the assembly reaction was set aside as input, the rest was immunoprecipitated with FLAG-M2 magnetic beads in 400 µl rEJC binding buffer. After 2 hours incubation in an overhead shaker at 4°C the beads were washed 3x with rEJC binding buffer and the proteins were eluted in 30 µl 1x SDS loading buffer. Input and IP were analyzed by SDS-PAGE and coomassie staining.

#### **6.4.8 Assembly of recombinant EJCs**

For in vitro EJC assembly, recombinant proteins (15 mg each) were mixed in a total volume of 50  $\mu$ l rEJC binding buffer supplemented with 10 mM ADPNP and 10  $\mu$ M poly(U)<sub>15</sub> ssRNA and incubated overnight at 4°C. FLAG-immunoprecipitations were performed with FLAG-M2 magnetic beads (Sigma) as described above. When indicated, binding was performed in the absence of ADPNP and ssRNA to prevent EJC formation.

#### **6.4.9 Isothermal titration calorimetry (ITC)**

To determine the affinities of S2-eIF4A3 for GST-CWC22(110-409) and GST-CWC22(110-409; KR-330/1-DE), ITC experiments were performed using the MicroCal iTC 200 system (GE Healthcare). Recombinant proteins were dialyzed overnight at 4°C in rEJC binding buffer (20 mM HEPES-KOH (pH 7.9), 125 mM NaCl, 1 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 10% Glycerol, 1 mM  $\beta$ -mercaptoethanol). The proteins were diluted in dialysis buffer to a final concentration of 6  $\mu$ M (S2-eIF4A3) and 60  $\mu$ M (GST-CWC22(110-409) and GST-CWC22(110-409; KR-330/1-DE)). 300  $\mu$ l diluted S2-eIF4A3 was loaded into the sample cell of the isothermal titration calorimeter, 80  $\mu$ l GST-CWC22 (WT or mutant) was loaded into the syringe. ITC measurements were performed at 20°C and 1000 rpm, by injecting aliquots of 2.5  $\mu$ l into the sample cell.

### **6.5 *In vitro* splicing and RNP immunoprecipitation**

#### **6.5.1 Generation of HEK293 whole cell extracts**

To generate HEK293 whole cell extracts,  $1.6 \times 10^6$  HEK293 cells were seeded in 10 cm dishes. 24 hours after seeding, cells were transfected with 8  $\mu$ g of pCI-FLAG expressing the protein of interest and 1.5  $\mu$ g pCI-mVenus as a transfection control. Transfections were performed with BBS/Calciumphosphate as described in chapter 6.2.3. The cells were harvested 48 hours after transfection. To this end, the cells were washed once with 2-3 ml PBS, scraped from the dish in 800  $\mu$ l PBS and transferred to a 1.5 ml reaction tube. After centrifuging 5 min at 200xg and 4°C, the cell pellet was resuspended in 200  $\mu$ l buffer E (20 mM HEPES-KOH (pH 7.9), 100 mM KCl, 0.2 mM

EDTA, 10% Glycerol, 1 mM DTT). The cells were lysed by sonification with 15 pulses, 30% intensity, using a Branson sonifier 250 and a 3 mm Microtip. The cell lysate was cleared by centrifuging for 10 min at 10000xg and stored at -80°C until further use.

### **6.5.2 *In vitro* transcription of radiolabeled mRNA**

The MINX and AdML sequence and their derivatives (MINX-GG, MINX  $\Delta$ intron, AdML-PT60, AdML-PT60/e1(18)) were cloned into pGEM-4Z vector to serve as templates for *in vitro* transcription of radiolabeled mRNA. 6-8  $\mu$ g of the pGEM-4Z plasmid was linearized with BamHI and purified via gel extraction. Capped transcripts were generated by run-off transcription with SP6 polymerase in the presence of m7GpppG cap analog (Promega) and  $\alpha$ -<sup>32</sup>P-GTP. To this end 10  $\mu$ l transcription reactions were assembled, containing 1  $\mu$ l 10x transcription buffer, 1 mM DTT, 0.5 mM ATP, 0.5 mM UTP, 0.5 mM CTP, 0.05 mM GTP, 0.5 mM cap analog (Promega), 20 u SP6 polymerase (NEB), 20 u RNasin, 0.5  $\mu$ g linearized template and 2.5  $\mu$ l  $\alpha$ -<sup>32</sup>P-GTP (800 Ci/mmol) and incubated for 40 min at 40°C. To digest the template DNA 2 u RNase-free Turbo DNase (Life technologies) was added and incubated for 15 min at 37°C. The RNA was extracted by Isol-RNA Lysis Reagent™/BCP extraction as described in chapter 6.3.4. 400  $\mu$ l Isol-RNA Lysis reagent was added to the transcription reaction and phase separation was induced by adding 60  $\mu$ l BCP. After centrifuging for 10 min at 20000xg and 4°C, the upper aqueous phase (~200  $\mu$ l) was transferred to a new 1.5 ml reaction tube. 1.5  $\mu$ l Glycogen was added to the aqueous phase as a precipitation carrier. RNA precipitation was induced by adding 200  $\mu$ l 2-propanol and 30 min centrifuging at 20000xg and 4°C. After the precipitation the liquid was removed completely and the RNA pellet immediately resuspended in 50  $\mu$ l ultra pure H<sub>2</sub>O. 1  $\mu$ l of the transcript was measured in a scintillation counter to determine the efficiency of transcription and  $\alpha$ -<sup>32</sup>P-GTP incorporation. The transcript was diluted in H<sub>2</sub>O to a final working concentration of 10000cpm/  $\mu$ l (for MINX and AdML) or 7000cpm/  $\mu$ l (for MINX-GG and AdML-PT60)

### **6.5.3 *In vitro* splicing reaction**

The *in vitro* splicing reaction was carried out in HeLa nuclear extract (CIL Biotech, CC-01-20-50) supplemented with whole cell extracts of HEK293 cells expressing the FLAG-tagged protein of interest (see chapter 6.5.1). Extracts from HEK293 cells expressing

unfused FLAG served as a negative control. A 25  $\mu$ l splicing reaction was prepared by mixing 5.4 mM MgCl<sub>2</sub>, 0.2 mM ATP, 40 mM creatine phosphate, 20% HeLa nuclear extract (CIL Biotech), 4.8% polyvinyl alcohol (PVA, MW 9000–10000 from Sigma Aldrich, prepare 12% stock solution in H<sub>2</sub>O), 44 mM HEPES-KOH (pH 7.3) and 80 u RNase inhibitor. 20  $\mu$ l of cleared HEK293 whole cell extract (see chapter 6.5.1) and 5  $\mu$ l diluted radiolabeled mRNA (see chapter 6.5.2) were added to the splice mix and the whole reaction incubated for 2 hours at 30°C. After the incubation time, 8  $\mu$ l of the splicing reaction was set aside as input. 400  $\mu$ l Isol-RNA lysis reagent was added to the input and the mix stored at -20°C until further use (see chapter 6.5.5). The rest of the splicing reaction was used for RNP immunoprecipitation with anti-FLAG M2 beads.

### **6.5.4 RNP immunoprecipitation**

The RNP immunoprecipitation was performed with red anti-FLAG-M2 affinity gel (Sigma-Aldrich) in EJC buffer (20 mM HEPES-KOH (pH 7.9), 200 mM NaCl, 2 mM MgCl<sub>2</sub>, 0.2% Triton X-100, 0.1% Nonidet P40, 0.05% Na-deoxycholic acid). 11  $\mu$ l anti-FLAG-M2 affinity gel and 200  $\mu$ l EJC buffer were added to the splicing reacting and incubated 2 hours in an overhead shaker at 4°C. After the incubation time, the beads were washed 3 times with 600  $\mu$ l EJC buffer. To this end the beads were centrifuged at 30xg for 30 sec, the supernatant was removed and replaced by new EJC buffer. After the last washing step, the beads were resuspended in 400  $\mu$ l EJC buffer and loaded onto an empty spin column (receiver column from Macherey-Nagel). The column was allowed to empty by gravity flow and the beads were washed twice with 400  $\mu$ l EJC buffer (by gravity flow). Afterwards the column was dried through short centrifugation (15 sec at 500xg). 400  $\mu$ l Isol-RNA lysis reagent was added to the column in order to elute the RNA from the beads. The eluate was collected in a new 1.5 ml reaction tube.

### **6.5.5 RNA extraction**

RNA was extracted from the splicing input (see chapter 6.5.3) and the RNP IP (chapter 6.5.4) in parallel using Isol-RNA lysis reagent and BCP. 60  $\mu$ l BCP was added to induce phase separation and the mix was centrifuged for 15 min at 20000xg. The aqueous phase (200  $\mu$ l) was transferred to a new 1.5 ml reaction tube. 2  $\mu$ l of a 1:1 mix of glycogen and blue Dextran (Sigma-Aldrich, #D5751) was used as precipitation carrier to

better visualize the small RNA pellets. 200  $\mu$ l 2-propanol was added to the aqueous phase, the mix was vortexed vigorously and incubated over night at -20°C. On the following day the RNA was precipitated by centrifugation for 30 min at 20000xg and 4°C. After precipitation, the RNA pellets were resuspended in 8  $\mu$ l (IP) or 16  $\mu$ l (input) RNA loading dye (80% formaldehyde, 2 mM EDTA, traces of bromophenol blue and xylene cyanol) and resolved through shaking for 10 min at 1100rpm and 55°C.

### 6.5.6 Denaturing TBE-gel electrophoresis

A 10% denaturing TBE-polyacrylamide gel mix was prepared from 10% acrylamide (19:1), 42% (w/v) urea, 1 TBE. Polymerization of the gel was induced by adding 50  $\mu$ l fresh APS (10%) and 10  $\mu$ l TEMED to 10 ml gel mix. Gels were cast with the standard mini-gel system from Bio-Rad using 1 mm combs with 10 wells. The gels were polymerized for at least 1 hour at RT or stored over night at 4°C. Before loading of the samples the gel was pre-run for 15–30 min at 15 mA in 1x TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) running buffer. Immediately before loading, the wells were flushed with 1x TBE to remove urea that has diffused into the gel. The RNA samples were boiled for 3 minutes at 90°C and then directly loaded onto the gel. 8  $\mu$ l input (50%) and 8  $\mu$ l IP (100%) were loaded onto the gel, so that the input represented ~10% of the total splicing reaction. Unspliced mRNA transcript was diluted in RNA loading dye (~1000cpm in 5  $\mu$ l loading dye) and loaded next to the splicing samples. The gel was run for approximately 1 hour at 15 mA in 1x TBE until the Xylene cyanol running front had reached the bottom quarter of the gel. The gel was rinsed 5 minutes in water, followed by 10 minutes incubation in 5% Glycerol (diluted in water). Afterwards, the gel was placed onto a wet filter paper (Whatman) and dried for 1.5 hours at 80°C, using a vacuum-dryer (Model 583 gel dryer from Bio-rad). The radioactive signal was detected using a storage phosphor screen (exposed for 4 hours to overnight) and the Typhoon phosphor imaging system (GE life sciences). Band intensities were quantified using the ImageQuant TL software (GE life sciences).

### 6.5.7 In vitro splicing and oligonucleotide-directed RNase H digestion

To analyze the assembly of protein complexes at position -24 of spliced mRNA, the *in vitro* splicing assay was combined with an oligonucleotide-directed RNase H digestion.

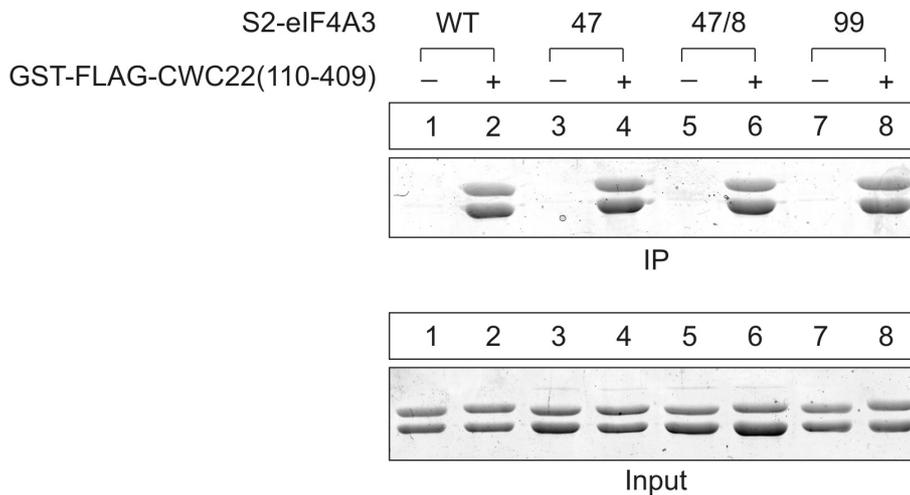
To this end, an *in vitro* splicing reaction was assembled as described before. Instead of supplementing the reaction with 20  $\mu$ l HEK 293 whole cell extract, the splicing was carried out in 25  $\mu$ l HeLa nuclear extract. The splicing reaction was mixed with diluted radiolabeled mRNA substrate and increasing concentrations of recombinant GST-tagged CWC22 (5  $\mu$ g, 15  $\mu$ g, 30  $\mu$ g). After splicing for 2 hours at 30°C, 1  $\mu$ l DNA oligo (100  $\mu$ M) complementary to the mRNA at position -24nt was added to the reaction and the mix was incubated for another 20 min at 30°C. During this time, the RNase H present in HeLa nuclear extract digests DNA:RNA duplexes that form when the oligo anneals to the mRNA. The reaction was stopped by adding 400  $\mu$ l Isol-RNA lysis reagent and the RNA was extracted as described before. After RNA-precipitation, the pellet was resuspended in 15  $\mu$ l RNA loading dye and 5  $\mu$ l of the sample (1/3) was loaded onto a 10% polyacrylamide-urea gel and analyzed as described before. Splicing and digestion of the intronless mRNA substrate MINX  $\Delta$ intron served as a positive control.

### **6.6 Bioinformatics and statistics**

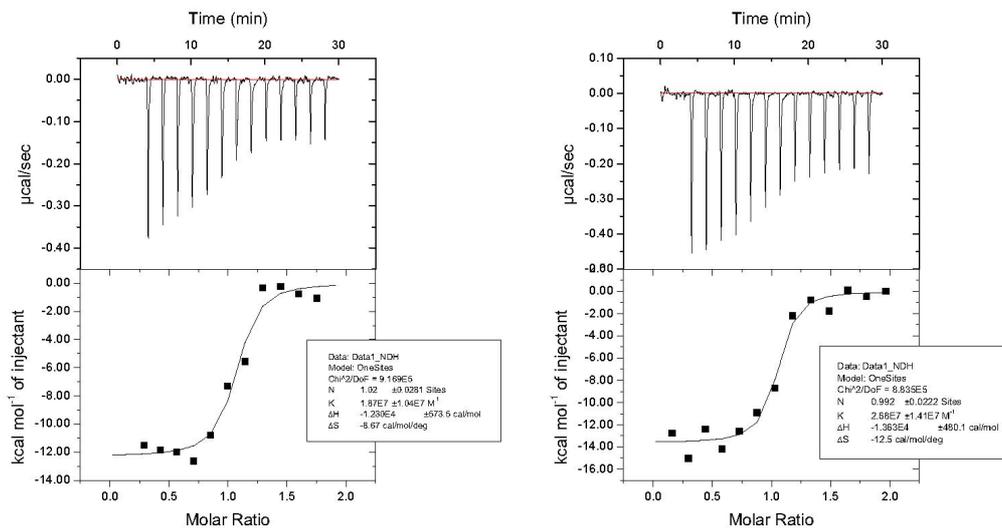
Homology searches were performed using BLAST (NCBI). Domain structures were predicted with SMART (Schultz et al., 1998) or Phyre2 (Kelley and Sternberg, 2009). Structural superimpositions were generated with DaliLite (Holm and Park, 2000) and structural images were rendered with PyMOL (DeLano, 2010).

Statistical significances were calculated with student's t test using the program GraphPad PRISM.

## 7 Supplementary Figures

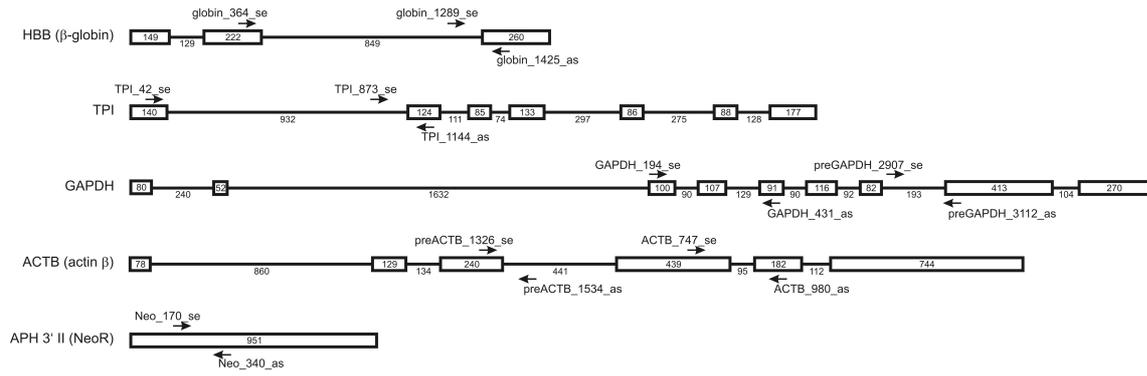


**Supplementary Figure 1: Mutations in the RecA1 domain of eIF4A3 do not affect the binding to CWC22.** Interaction between recombinant proteins *in vitro*. Co-precipitation of recombinant S2-eIF4A3 and the indicated mutants with immobilized GST-FLAG-CWC22(110-665). The protein complexes were resolved by SDS-PAGE and visualized with coomassie. 10% of the assembly reaction was loaded as input.

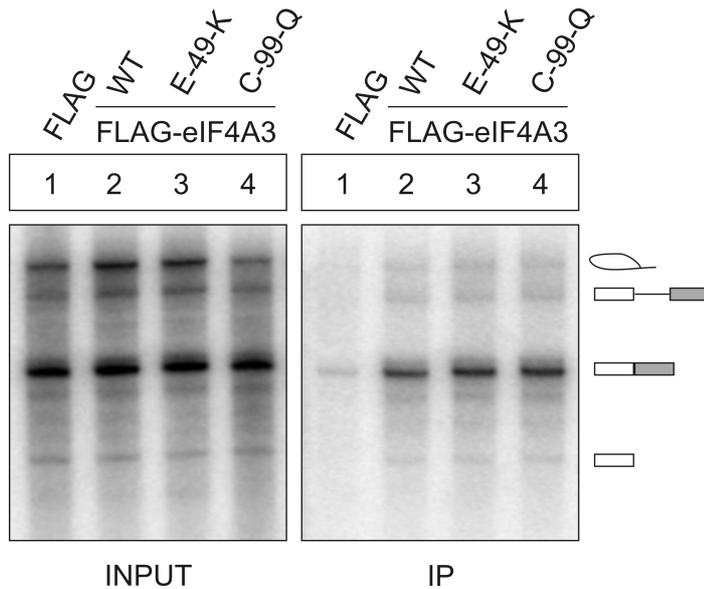


**Supplementary Figure 2: Mutations in the C-terminal part of GST-CWC22(110-409) do not affect the affinity for eIF4A3.** ITC experiments with recombinant S2-eIF4A3 and GST-CWC22(110-409) (left side) or GST-CWC22(110-409; KR-330/1-DE) (right side). The determined binding affinities of S2-eIF4A3 are 53nM for GST-CWC22(110-409) and 37nM for GST-CWC22(110-409; KR-330/1-DE).

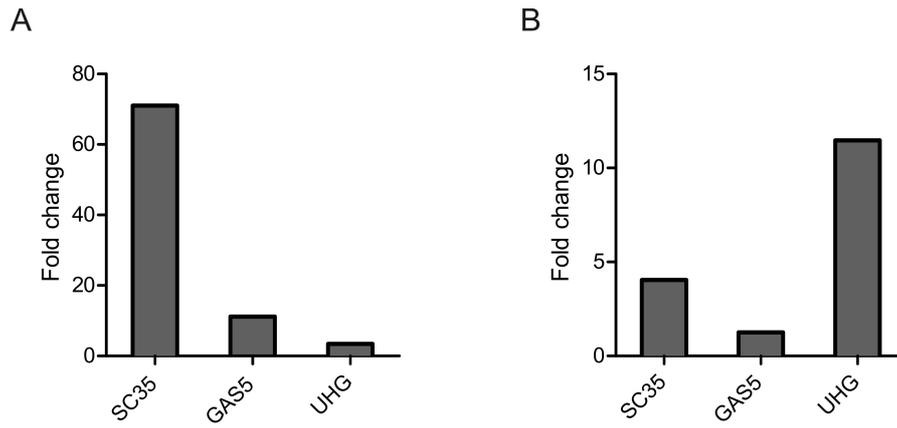
## Appendix



**Supplementary Figure 3: Exon-intron structure of genes analyzed by qPCR.** Arrows denote the primers used for the amplification of mRNA and pre-mRNA.



**Supplementary Figure 4: Mutations in the RecA1 domain of eIF4A3 do not affect the splicing-dependent EJC assembly.** <sup>32</sup>P-body-labeled MINX mRNA was spliced in the presence of FLAG-eIF4A3, the respective mutants of FLAG-eIF4A3 or unfused FLAG as a negative control. After FLAG-immunoprecipitation of mRNPs, the RNA was extracted, resolved on a denaturing urea-gel and visualized by phosphor imaging autoradiography. 10% of the splicing reaction was loaded as input. Schemes on the right site of the panels depict the splicing products.



**Supplementary Figure 5: CWC22-depletion and CWC22(NK-171/2-DE) over-expression differentially affect endogenous NMD targets.** (A) (A) Flp-In 293 cells were induced with doxycycline and the change in SC35 (1.7kb), GAS5 and UHG mRNA levels compared to un-induced control cells were quantified by qRT-PCR. (B) Quantification of the change in SC35 (1.7kb), GAS5 and UHG mRNA levels upon depletion of CWC22.



## 8 Bibliography

- Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., and Luhrmann, R. (2011). Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* *31*, 2667-2682.
- Aguilera, A., and Klein, H.L. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the *Saccharomyces cerevisiae* TOP1 gene. *Mol Cell Biol* *10*, 1439-1451.
- Alexandrov, A., Colognori, D., Shu, M.D., and Steitz, J.A. (2012). Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc Natl Acad Sci U S A* *109*, 21313-21318.
- Alexandrov, A., Colognori, D., and Steitz, J.A. (2011). Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. *Genes Dev* *25*, 1078-1090.
- Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* *432*, 112-118.
- Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for *C.elegans* nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. *EMBO J* *22*, 641-650.
- Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Seraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* *313*, 1968-1972.
- Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W., and Bork, P. (2001). Comparison of ARM and HEAT protein repeats. *J Mol Biol* *309*, 1-18.
- Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* *280*, 7131-7134.
- Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrancois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., *et al.* (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in *Drosophila*. *Cell* *143*, 251-262.
- Bachi, A., Braun, I.C., Rodrigues, J.P., Pante, N., Ribbeck, K., von Kobbe, C., Kutay, U., Wilm, M., Gorlich, D., Carmo-Fonseca, M., *et al.* (2000). The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA* *6*, 136-158.
- 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.
- Baltz, A.G., Munschauer, M., Schwanhauser, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M., *et al.* (2012). The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* *46*, 674-690.
- Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* *19*, 983-990.

## Bibliography

---

- 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. *EMBO J* **21**, 2724-2735.
- Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* **17**, 251-256.
- Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Luhrmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16**, 2384-2403.
- Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846-850.
- Bhaskar, P.T., and Hay, N. (2007). The two TORCs and Akt. *Dev Cell* **12**, 487-502.
- Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. *Cell* **126**, 37-47.
- Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713-725.
- Bono, F., Ebert, J., Unterholzner, L., Guttler, T., Izaurralde, E., and Conti, E. (2004). Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep* **5**, 304-310.
- Braddock, M., Muckenthaler, M., White, M.R., Thorburn, A.M., Sommerville, J., Kingsman, A.J., and Kingsman, S.M. (1994). Intron-less RNA injected into the nucleus of *Xenopus* oocytes accesses a regulated translation control pathway. *Nucleic Acids Res* **22**, 5255-5264.
- Buchwald, G., Schussler, S., Basquin, C., Le Hir, H., and Conti, E. (2013). Crystal structure of the human eIF4AIII-CWC22 complex shows how a DEAD-box protein is inhibited by a MIF4G domain. *Proc Natl Acad Sci U S A* **110**, E4611-4618.
- Budiman, M.E., Bubenik, J.L., and Driscoll, D.M. (2011). Identification of a signature motif for the eIF4a3-SECIS interaction. *Nucleic Acids Res* **39**, 7730-7739.
- Budiman, M.E., Bubenik, J.L., Miniard, A.C., Middleton, L.M., Gerber, C.A., Cash, A., and Driscoll, D.M. (2009). Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. *Mol Cell* **35**, 479-489.
- Callis, J., Fromm, M., and Walbot, V. (1987). Introns increase gene expression in cultured maize cells. *Genes Dev* **1**, 1183-1200.
- Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. *Curr Opin Struct Biol* **12**, 123-133.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406.
- Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat Struct Mol Biol* **15**, 85-93.
- Chan, S.P., and Cheng, S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J Biol Chem* **280**, 31190-31199.
- Chazal, P.E., Daguene, E., Wendling, C., Ulryck, N., Tomasetto, C., Sargueil, B., and Le Hir, H. (2013). EJC core component MLN51 interacts with eIF3 and activates translation. *Proc Natl Acad Sci U S A* **110**, 5903-5908.

## Bibliography

---

- 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.
- Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* 349, 487-493.
- Conti, E., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* 17, 316-325.
- Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene* 367, 17-37.
- 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. *EMBO J* 23, 2478-2487.
- Craig, A.W., Haghighat, A., Yu, A.T., and Sonenberg, N. (1998). Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. *Nature* 392, 520-523.
- Craig, N.L. (1988). The mechanism of conservative site-specific recombination. *Annu Rev Genet* 22, 77-105.
- Cui, Y., Hagan, K.W., Zhang, S., and Peltz, S.W. (1995). Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. *Genes Dev* 9, 423-436.
- Czaplinski, K., Ruiz-Echevarria, M.J., Paushkin, S.V., Han, X., Weng, Y., Perlick, H.A., Dietz, H.C., Ter-Avanesyan, M.D., and Peltz, S.W. (1998). The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs. *Genes Dev* 12, 1665-1677.
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* 26, 5528-5543.
- 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.
- DeLano, W.L. (2010). The PyMOL Molecular Graphics System, Version 1.3r1 (New York: Schrodinger, LLC).
- Diem, M.D., Chan, C.C., Younis, I., and Dreyfuss, G. (2007). PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol* 14, 1173-1179.
- Dossani, Z.Y., Weirich, C.S., Erzberger, J.P., Berger, J.M., and Weis, K. (2009). Structure of the C-terminus of the mRNA export factor Dbp5 reveals the interaction surface for the ATPase activator Gle1. *Proc Natl Acad Sci U S A* 106, 16251-16256.
- Dostie, J., and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr Biol* 12, 1060-1067.
- Eberle, A.B., Lykke-Andersen, S., Muhlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* 16, 49-55.
- Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends Genet* 24, 280-288.
- Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* 14, 496-506.

## Bibliography

---

- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol Cell* *36*, 593-608.
- Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* *20*, 313-324.
- Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., Bertola, D., Felix, T.M., de Souza, J., Raskin, S., Twigg, S.R., Weiner, A.M., Armas, P., *et al.* (2014). A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. *Am J Hum Genet* *94*, 120-128.
- 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.
- Fourmann, J.B., Schmitzova, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio, P., and Luhrmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev* *27*, 413-428.
- Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci U S A* *102*, 16176-16181.
- Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003). A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol* *10*, 433-439.
- 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.
- Giorgi, C., and Moore, M.J. (2007). The nuclear nurture and cytoplasmic nature of localized mRNPs. *Semin Cell Dev Biol* *18*, 186-193.
- Giorgi, C., Yeo, G.W., Stone, M.E., Katz, D.B., Burge, C., Turrigiano, G., and Moore, M.J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell* *130*, 179-191.
- Glanzer, J., Miyashiro, K.Y., Sul, J.Y., Barrett, L., Belt, B., Haydon, P., and Eberwine, J. (2005). RNA splicing capability of live neuronal dendrites. *Proc Natl Acad Sci U S A* *102*, 16859-16864.
- Gonzalez, C.I., Bhattacharya, A., Wang, W., and Peltz, S.W. (2001). Nonsense-mediated mRNA decay in *Saccharomyces cerevisiae*. *Gene* *274*, 15-25.
- 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.
- Haremake, T., and Weinstein, D.C. (2012). Eif4a3 is required for accurate splicing of the *Xenopus laevis* ryanodine receptor pre-mRNA. *Dev Biol* *372*, 103-110.
- He, F., Brown, A.H., and Jacobson, A. (1997). Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. *Mol Cell Biol* *17*, 1580-1594.
- Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Luhrmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* *45*, 567-580.
- Hillen, W., and Berens, C. (1994). Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annu Rev Microbiol* *48*, 345-369.

## Bibliography

---

- Ho, C.K., and Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* **3**, 405-411.
- Hodge, C.A., Tran, E.J., Noble, K.N., Alcazar-Roman, A.R., Ben-Yishay, R., Scarcelli, J.J., Folkmann, A.W., Shav-Tal, Y., Wenthe, S.R., and Cole, C.N. (2011). The Dbp5 cycle at the nuclear pore complex during mRNA export I: dbp5 mutants with defects in RNA binding and ATP hydrolysis define key steps for Nup159 and Gle1. *Genes Dev* **25**, 1052-1064.
- Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode *Caenorhabditis elegans*. *Genetics* **123**, 301-313.
- Holm, L., and Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566-567.
- Huertas, P., and Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* **12**, 711-721.
- Huntzinger, E., Kashima, I., Fauser, M., Sauliere, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14**, 2609-2617.
- Ideue, T., Sasaki, Y.T., Hagiwara, M., and Hirose, T. (2007). Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev* **21**, 1993-1998.
- Jankowsky, E., and Fairman, M.E. (2007). RNA helicases--one fold for many functions. *Curr Opin Struct Biol* **17**, 316-324.
- Kahvejian, A., Roy, G., and Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb Symp Quant Biol* **66**, 293-300.
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* **20**, 355-367.
- Kataoka, N., Diem, M.D., Kim, V.N., Yong, J., and Dreyfuss, G. (2001). MAGOH, a human homolog of *Drosophila mago nashi* protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J* **20**, 6424-6433.
- Kataoka, N., Yong, J., Kim, V.N., Velazquez, F., Perkinson, R.A., Wang, F., and Dreyfuss, G. (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* **6**, 673-682.
- Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371.
- Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* **35**, 125-131.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**, 2452-2460.
- Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**, 909-915.
- Kressler, D., de la Cruz, J., Rojo, M., and Linder, P. (1997). Fal1p is an essential DEAD-box protein involved in 40S-ribosomal-subunit biogenesis in *Saccharomyces cerevisiae*. *Mol Cell Biol* **17**, 7283-7294.
- Kuhn, U., and Wahle, E. (2004). Structure and function of poly(A) binding proteins. *Biochim Biophys Acta* **1678**, 67-84.

## Bibliography

---

- Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* *446*, 926-929.
- Lau, C.K., Diem, M.D., Dreyfuss, G., and Van Duynne, G.D. (2003). Structure of the Y14-MAGOH core of the exon junction complex. *Curr Biol* *13*, 933-941.
- Le Hir, H., and Andersen, G.R. (2008). Structural insights into the exon junction complex. *Curr Opin Struct Biol* *18*, 112-119.
- Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* *20*, 4987-4997.
- Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. (2000a). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* *19*, 6860-6869.
- Le Hir, H., Moore, M.J., and Maquat, L.E. (2000b). Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev* *14*, 1098-1108.
- Leeds, P., Peltz, S.W., Jacobson, A., and Culbertson, M.R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* *5*, 2303-2314.
- Lei, J., Mesters, J.R., von Brunn, A., and Hilgenfeld, R. (2011). Crystal structure of the middle domain of human poly(A)-binding protein-interacting protein 1. *Biochem Biophys Res Commun* *408*, 680-685.
- 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.
- Li, Q., Imataka, H., Morino, S., Rogers, G.W., Jr., Richter-Cook, N.J., Merrick, W.C., and Sonenberg, N. (1999). Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol Cell Biol* *19*, 7336-7346.
- Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* *122*, 365-378.
- Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* *12*, 505-516.
- Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J., and Slonimski, P.P. (1989). Birth of the D-E-A-D box. *Nature* *337*, 121-122.
- Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* *27*, 2125-2138.
- Loh, P.G., Yang, H.S., Walsh, M.A., Wang, Q., Wang, X., Cheng, Z., Liu, D., and Song, H. (2009). Structural basis for translational inhibition by the tumour suppressor Pdc4. *EMBO J* *28*, 274-285.
- Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A* *96*, 14937-14942.
- Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* *413*, 644-647.
- Lykke-Andersen, J. (2002). Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay. *Mol Cell Biol* *22*, 8114-8121.

## Bibliography

---

- Ma, X.M., Yoon, S.O., Richardson, C.J., Julich, K., and Blenis, J. (2008). SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* 133, 303-313.
- Macchi, P., Kroening, S., Palacios, I.M., Baldassa, S., Grunewald, B., Ambrosino, C., Goetze, B., Lupas, A., St Johnston, D., and Kiebler, M. (2003). Barentsz, a new component of the Staufen-containing ribonucleoprotein particles in mammalian cells, interacts with Staufen in an RNA-dependent manner. *J Neurosci* 23, 5778-5788.
- Marintchev, A., Edmonds, K.A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C., Herdy, B., Sonenberg, N., and Wagner, G. (2009). Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* 136, 447-460.
- Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* 19, 1512-1517.
- Matsumoto, K., Wassarman, K.M., and Wolffe, A.P. (1998). Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J* 17, 2107-2121.
- Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Krainer, A.R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *EMBO J* 18, 4560-4570.
- McGuire, A.M., Pearson, M.D., Neafsey, D.E., and Galagan, J.E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol* 9, R50.
- Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* 36, 1073-1078.
- 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.
- Michelle, L., Cloutier, A., Toutant, J., Shkreta, L., Thibault, P., Durand, M., Garneau, D., Gendron, D., Lapointe, E., Couture, S., *et al.* (2012). Proteins associated with the exon junction complex also control the alternative splicing of apoptotic regulators. *Mol Cell Biol* 32, 954-967.
- Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* 41, 21-32.
- Mishler, D.M., Christ, A.B., and Steitz, J.A. (2008). Flexibility in the site of exon junction complex deposition revealed by functional group and RNA secondary structure alterations in the splicing substrate. *RNA* 14, 2657-2670.
- 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 Dev* 15, 2886-2899.
- Monshausen, M., Gehring, N.H., and Kosik, K.S. (2004). The mammalian RNA-binding protein Staufen2 links nuclear and cytoplasmic RNA processing pathways in neurons. *Neuromolecular Med* 6, 127-144.
- Montpetit, B., Thomsen, N.D., Helmke, K.J., Seeliger, M.A., Berger, J.M., and Weis, K. (2011). A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export. *Nature* 472, 238-242.
- Moore, M.J. (2005). From birth to death: the complex lives of eukaryotic mRNAs. *Science* 309, 1514-1518.
- Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136, 688-700.

- Moriarty, P.M., Reddy, C.C., and Maquat, L.E. (1998). Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA. *Mol Cell Biol* *18*, 2932-2939.
- Mourtada-Maarabouni, M., and Williams, G.T. (2013). Growth arrest on inhibition of nonsense-mediated decay is mediated by noncoding RNA GAS5. *Biomed Res Int* *2013*, 358015.
- Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat Rev Genet* *14*, 275-287.
- Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* *23*, 198-199.
- Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* *21*, 708-718.
- Nielsen, K.H., Chamieh, H., Andersen, C.B., Fredslund, F., Hamborg, K., Le Hir, H., and Andersen, G.R. (2009). Mechanism of ATP turnover inhibition in the EJC. *RNA* *15*, 67-75.
- Noble, K.N., Tran, E.J., Alcazar-Roman, A.R., Hodge, C.A., Cole, C.N., and Wenthe, S.R. (2011). The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. *Genes Dev* *25*, 1065-1077.
- 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.
- Ohi, M.D., and Gould, K.L. (2002). Characterization of interactions among the Cef1p-Prp19p-associated splicing complex. *RNA* *8*, 798-815.
- Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K.R., Grimson, A., Hachiya, T., Hentze, M.W., Anderson, P., and Ohno, S. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol Cell* *12*, 1187-1200.
- Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2012). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res* *40*, 1251-1266.
- 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.
- Patel, S.S., and Picha, K.M. (2000). Structure and function of hexameric helicases. *Annu Rev Biochem* *69*, 651-697.
- Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* *16*, 272-278.
- Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. *Genes Dev* *25*, 1770-1782.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* *12*, 1231-1245.
- Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol* *15*, 326-331.
- 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 Dev* *16*, 2778-2791.

## Bibliography

---

- Rodrigues, J.P., Rode, M., Gatfield, D., Blencowe, B.J., Carmo-Fonseca, M., and Izaurralde, E. (2001). REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc Natl Acad Sci U S A* *98*, 1030-1035.
- Roignant, J.Y., and Treisman, J.E. (2010). Exon junction complex subunits are required to splice *Drosophila* MAP kinase, a large heterochromatic gene. *Cell* *143*, 238-250.
- Sauer, B. (1994). Site-specific recombination: developments and applications. *Curr Opin Biotechnol* *5*, 521-527.
- Sauliere, J., Haque, N., Harms, S., Barbosa, I., Blanchette, M., and Le Hir, H. (2010). The exon junction complex differentially marks spliced junctions. *Nat Struct Mol Biol* *17*, 1269-1271.
- Sauliere, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonqueze, O., Audic, Y., Paillard, L., Roest Crollius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat Struct Mol Biol* *19*, 1124-1131.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* *95*, 5857-5864.
- Schutz, P., Bumann, M., Oberholzer, A.E., Bieniossek, C., Trachsel, H., Altmann, M., and Baumann, U. (2008). Crystal structure of the yeast eIF4A-eIF4G complex: an RNA-helicase controlled by protein-protein interactions. *Proc Natl Acad Sci U S A* *105*, 9564-9569.
- Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol Cell* *30*, 743-754.
- Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., Kidd, V.J., Manley, J.L., Lahti, J.M., and Reinberg, D. (2003). ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol Cell Biol* *23*, 2981-2990.
- Seeher, S., Mahdi, Y., and Schweizer, U. (2012). Post-transcriptional control of selenoprotein biosynthesis. *Curr Protein Pept Sci* *13*, 337-346.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* *125*, 287-300.
- Shatkin, A.J. (1976). Capping of eucaryotic mRNAs. *Cell* *9*, 645-653.
- Shi, H., and Xu, R.M. (2003). Crystal structure of the *Drosophila* Mago nashi-Y14 complex. *Genes Dev* *17*, 971-976.
- 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. *Nat Struct Mol Biol* *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* *12*, 360-374.
- Shuman, S. (2001). Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog Nucleic Acid Res Mol Biol* *66*, 1-40.
- Silverman, E., Edwalds-Gilbert, G., and Lin, R.J. (2003). DExD/H-box proteins and their partners: helping RNA helicases unwind. *Gene* *312*, 1-16.
- Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* *151*, 750-764.
- Singh, K.K., Erkelenz, S., Rattay, S., Dehof, A.K., Hildebrandt, A., Schulze-Osthoff, K., Schaal, H., and Schwert, C. (2010). Human SAP18 mediates assembly of a splicing regulatory multiprotein complex via its ubiquitin-like fold. *RNA* *16*, 2442-2454.

## Bibliography

---

- Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745.
- Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep* **2**, 454-461.
- Story, R.M., and Steitz, T.A. (1992). Structure of the recA protein-ADP complex. *Nature* **355**, 374-376.
- Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234.
- Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P., and Izaurralde, E. (2000). REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**, 638-650.
- Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* **20**, 1785-1796.
- Svitkin, Y.V., Pause, A., Haghghat, A., Pyronnet, S., Witherell, G., Belsham, G.J., and Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7**, 382-394.
- Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA helicase activity of the ribosome. *Cell* **120**, 49-58.
- 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.
- Tanner, N.K., Cordin, O., Banroques, J., Doere, M., and Linder, P. (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* **11**, 127-138.
- Thermann, R., Neu-Yilik, G., Deters, A., Frede, U., Wehr, K., Hagemeyer, C., Hentze, M.W., and Kulozik, A.E. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* **17**, 3484-3494.
- Tsukamoto, T., Shibagaki, Y., Niihara, Y., and Mizumoto, K. (1998). Cloning and characterization of three human cDNAs encoding mRNA (guanine-7)-methyltransferase, an mRNA cap methylase. *Biochem Biophys Res Commun* **251**, 27-34.
- Tuteja, N., and Tuteja, R. (2004a). Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery. *Eur J Biochem* **271**, 1835-1848.
- Tuteja, N., and Tuteja, R. (2004b). Unraveling DNA helicases. Motif, structure, mechanism and function. *Eur J Biochem* **271**, 1849-1863.
- van Eeden, F.J., Palacios, I.M., Petronczki, M., Weston, M.J., and St Johnston, D. (2001). Barentsz is essential for the posterior localization of oskar mRNA and colocalizes with it to the posterior pole. *J Cell Biol* **154**, 511-523.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701-718.
- Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945-951.
- 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.
- Yamada-Okabe, T., Doi, R., Shimmi, O., Arisawa, M., and Yamada-Okabe, H. (1998). Isolation and characterization of a human cDNA for mRNA 5'-capping enzyme. *Nucleic Acids Res* **26**, 1700-1706.

## Bibliography

---

- Yamashita, A. (2013). Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. *Genes Cells* *18*, 161-175.
- Yang, H.S., Cho, M.H., Zakowicz, H., Hegamyer, G., Sonenberg, N., and Colburn, N.H. (2004). A novel function of the MA-3 domains in transformation and translation suppressor Pcd4 is essential for its binding to eukaryotic translation initiation factor 4A. *Mol Cell Biol* *24*, 3894-3906.
- Yang, Q., and Jankowsky, E. (2006). The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* *13*, 981-986.
- Yeh, T.C., Liu, H.L., Chung, C.S., Wu, N.Y., Liu, Y.C., and Cheng, S.C. (2011). Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome to escape from a futile pathway. *Mol Cell Biol* *31*, 43-53.
- 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.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* *419*, 182-185.
- 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.
- Zillmann, M., Zapp, M.L., and Berget, S.M. (1988). Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol Cell Biol* *8*, 814-821.



## Danke!

Diese Arbeit wäre nicht möglich gewesen ohne tatkräftige Unterstützung. Aus diesem Grund möchte ich den Menschen, die mich in den vergangenen Jahren unterstützt und in vielerlei Hinsicht zum Gelingen dieser Arbeit beigetragen haben, meinen ganz persönlichen Dank aussprechen.

Mein besonderer Dank gilt dem Betreuer meiner Doktorarbeit, Niels Gehring. Danke für die Möglichkeit, dieses spannende Projekt zu bearbeiten und für die großartige Unterstützung hierbei. Ich danke dir auch für die unzähligen Diskussionen, die Inspiration und die vielen guten Ideen. Ganz besonders möchte ich dir aber dafür danken, dass du mich nicht nur bei der wissenschaftlichen Arbeit unterstützt, sondern auch meine persönliche Entwicklung als Wissenschaftler und als Mensch gefördert hast. Du warst ein großartiger Mentor und ich habe sehr viel von dir gelernt. Vielen Dank dafür!

Des Weiteren danke ich Frau Prof. Karin Schnetz dafür, dass sie sich die Zeit nimmt, meine Arbeit zu lesen und zu bewerten. Herrn Prof. Baumann danke ich für die Übernahme des Prüfungsvorsitzes und Martina Rembold für das Protokollieren der Prüfung.

Ein ganz besonderer Dank geht an die komplette Arbeitsgruppe Gehring für eine wunderbare Zeit im Labor und in Köln! Danke an Aga, Tobi, Franzi, Simo, Volker, Kusum und Marie für die tolle Laboratmosphäre und gute Zusammenarbeit - aber noch viel mehr für die Kaffeepausen, die Horror-movie-nights und ganz besonders für die Mauerbierchen!!

*People of Gehringia, you are awesome!! Thanks a lot for the great time!*

Heidi and Juliane möchte ich für die technische Unterstützung danken und dafür, dass sie mit Bestellungen und Organisation unser Labor am Laufen halten. Und danke für die Kuchen, Heidi!

Des Weiteren danke ich der Graduate school IGSDHD für die finanzielle und organisatorische Unterstützung der Doktorarbeit und insbesondere Isabell Witt, die sich immer persönlich um uns Doktoranden gekümmert hat. Ganz besonders zu Beginn meiner Doktorandenzeit hast du mir sehr geholfen – danke dafür!

Und schließlich möchte ich all den Menschen danken, die mich im Leben begleiten und immer unterstützt haben. Danke an meine ‚alten‘ Freunden aus Freiburg und die ‚neuen‘ Freunden aus Köln und Bonn! Ihr sei großartig, ich bin sehr froh, euch in meinem Leben zu haben!

Ganz besonders danke ich auch meiner Familie, insbesondere meinen Eltern, für ihre Liebe und ihre Unterstützung.

Und schließlich gilt mein ganz besonderer Dank Ben. Danke für eine großartige Zeit und dafür, dass du immer für mich da bist, mich unterstützt und an mich glaubst! Und natürlich danke für das Essen - ohne dich wäre ich bestimmt schon an einer Überdosis TK-Pizza zugrunde gegangen...

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