Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Molecular mechanisms of human mRNA 3' end formation

Felix Jakob Sandmeir

aus

Gräfelfing, Deutschland

2022

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Frau Prof. Dr. Elena Conti betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 5.10.2022

FELIX SANDMEIR

Dissertation eingereicht am 5.10.2022

1. Gutachterin: Prof. Dr. Elena Conti

2. Gutachter: Prof. Dr. Klaus Förstemann

Mündliche Prüfung am 24.11.2022

Table of contents

Summary

All human protein-coding RNAs require several co-transcriptional processing steps to generate mature mRNA molecules with defined 3' ends ready for export into the cytoplasm. Almost all 3' ends are generated by an essential two-step mechanism, cleavage and polyadenylation. After an endonucleolytic cleavage, positioned by multiple RNA elements within the 3' UTR, the poly(A) tail is added. While the enzymes responsible for both reactions were identified, many central questions regarding the mechanisms remain unanswered. The molecular details of endonuclease activation, including which proteins are required, are not understood. Moreover, although models of many factors or modules are available, a thorough structural characterization of their interplay is missing.

To address these questions, an advanced expression system was leveraged to purify recombinant factors from human cells and characterize both enzymatic activities of 3' end formation biochemically and structurally. In the first part of this study, different nuclease complexes were assembled to reconstitute specific pre-mRNA cleavage. This allowed for the first time to define a 14-subunit complex necessary and sufficient for endonuclease activity *in vitro*; this minimal complex contained CPSF, CstF, CF-II, RBBP6, and poly(A) polymerase (PAPOA). ATP binding but not its hydrolysis stimulates pre-mRNA cleavage. A combination of cryo-EM, AlphaFold2 and biochemical assays revealed that RBBP6 contacts CPSF3 near the active site opening, presumably to license its endonuclease activity.

Next, the second step of 3' end processing, polyadenylation by poly(A) polymerase, was addressed. While nuclear poly(A)-binding protein PABPN1 enhanced PAPOA activity, predominantly cytoplasmic PABPC1 inhibited polyadenylation, contrary to previous reports. Cryo-EM of PABPN1-coated poly(A) RNAs revealed that these RNPs form spherical particles, which are distinct from cytoplasmic poly(A) RNPs and might be involved in terminating processive pre-mRNA polyadenylation.

Overall, this work presented the first reconstitution of specific pre-mRNA cleavage using only purified proteins, thereby providing novel insights and extending the experimental framework available to study both enzymatic activities of 3' end formation. This provides a basis for investigating the three-dimensional architecture of the 3' end processing machinery and its intricate regulatory mechanisms.

1. Introduction

The eukaryotic transcription apparatus produces a tremendous amount of RNA molecules, both in sheer numbers and in diversity of RNA. Hence, nuclear gene transcription is divided among multiple DNA-dependent RNA polymerases, in mammals RNA polymerase (RNA Pol) I, II, and III (Girbig et al., 2022; Khatter et al., 2017; Osman and Cramer, 2020). The vast majority of RNAs is produced by Pol I and Pol III, which generate ribosomal RNAs (rRNA; Pol I) and short, structured RNAs (transfer RNA, tRNA or 5S rRNA; Pol III) (Khatter et al., 2017). Pol II transcribes all protein-coding genes to generate precursor messenger RNAs (pre-mRNAs), which serve as information carriers from DNA to protein, as well as some regulatory non-coding RNAs (Proudfoot, 2016; Tan-Wong et al., 2012; Vorländer et al., 2022; Wyers et al., 2005). Eukaryotic mRNAs, a highly diverse class of transcription units in terms of length and nucleotide composition, share a common architecture as they undergo the same transcription and maturation steps (Girbig et al., 2022; Proudfoot, 2016; Vorländer et al., 2022). The protein-coding sequences are flanked by 5' and 3' untranslated regions (5' UTR and 3' UTR), which have different regulatory functions. Their respective ends are modified with terminal structures, 5' cap and 3' poly(A) tail, which are added during co-transcriptional pre-mRNA processing steps (Mayr, 2017; Proudfoot, 2016).

1.1 Co-transcriptional pre-mRNA processing

Pol II generates pre-mRNAs by cycling through three general phases, transcription initiation, elongation, and termination (Girbig et al., 2022). The coordinated binding and dissociation of numerous auxiliary factors such as general transcription factors regulate the transitions through the different steps of the cycle (Girbig et al., 2022; Osman and Cramer, 2020). Pol II is a highly conserved multi-subunit enzyme; its largest subunit Rpb1 carrying the catalytic activity and a C-terminal domain (CTD), which forms a flexible, tail-like extension from the core enzyme (Cramer et al., 2001; Girbig et al., 2022; Osman and Cramer, 2020; Vannini and Cramer, 2012). The CTD, which is linked to Pol II near the RNA exit tunnel, serves as interaction platform, which plays a critical role in coordinating the processing steps necessary for the maturation of the pre-mRNA (Cramer et al., 2001; Eick and Geyer, 2013; Osman and Cramer,

2020). These co-transcriptional processing steps, 5' capping, splicing, and 3' end formation, are needed to produce mature mRNAs ready for export into the cytoplasm (**Fig. 1)** (Hsin and Manley, 2012; Vorländer et al., 2022). Each maturation step allows several proteins to assemble on the mRNA to form messenger ribonucleoprotein (mRNP) particles. This packaging additionally helps to prevent genomic instability by avoiding the formation of R-loops, harmful DNA:RNA hybrids (Aguilera and García-Muse, 2012; Vorländer et al., 2022).

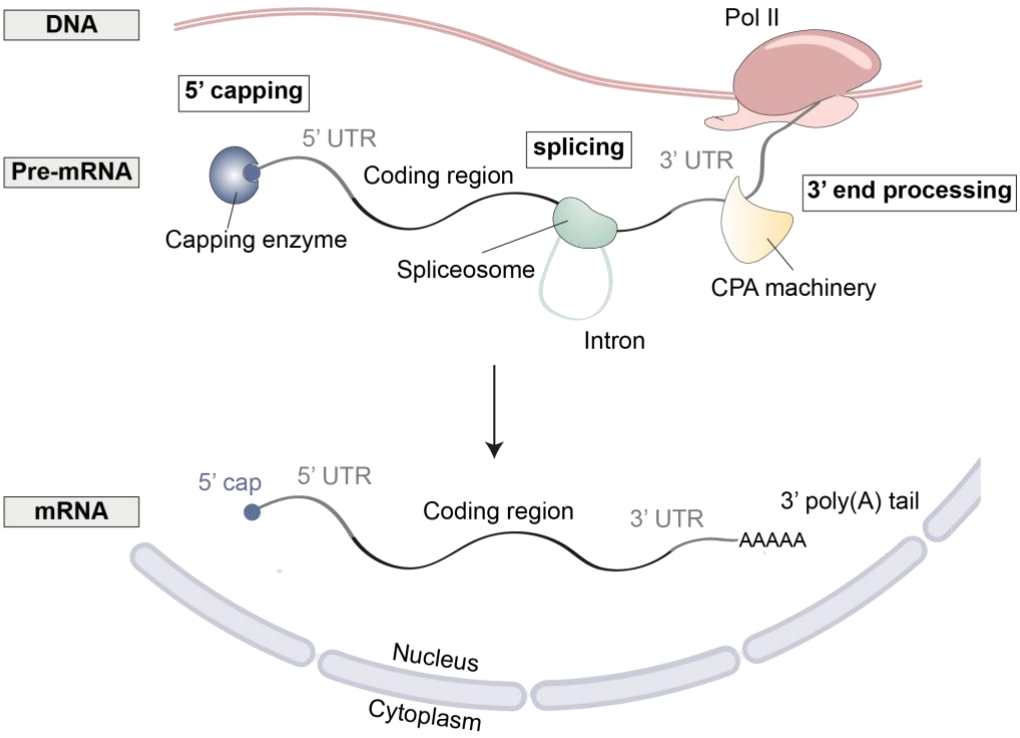

Figure 1: Co-transcriptional pre-mRNA processing. Pre-mRNA molecules are co-transcriptionally processed to generate mature mRNAs, which are subsequently exported into the cytoplasm.

The CTD, an array of heptad repeats with the consensus sequence Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine (YSPTSPS), is a conserved Pol II protrusion with species-dependent variations in repeat number (Ahearn et al., 1987; Eick and Geyer, 2013; Proudfoot, 2016; Stiller and Hall, 2002; Yang et al., 2014). The 52 repeats in mammals can be phosphorylated at five of the seven residues (Y1, S2, T4, S5, S7), with these modifications likely causing an extension of the otherwise compact CTD (Hsin and Manley, 2012; Osman and Cramer, 2020). Depending on the phase of the transcription cycle, the controlled activity of kinases and phosphatases leads to defined phosphorylation patterns, which are specifically recognized by RNA-

processing factors (Hsin and Manley, 2012; Osman and Cramer, 2020). Ser-5 phosphorylation accumulates in promotor-proximal regions and results in binding of 5' capping factors (Cho et al., 1997; Ho and Shuman, 1999; Komarnitsky et al., 2000; McCracken et al., 1997a). Ser-2 phosphorylation predominates in regions more distal from promotors and leads to recruitment of the 3' end processing machinery (Cho et al., 2001; Komarnitsky et al., 2000). Tyr-1 phosphorylation also accumulates in promotor-proximal regions, while Thr-4 and Ser-7 phosphorylation are linked to more specific, nonessential functions (Buratowski, 2009; Descostes et al., 2014; Eick and Geyer, 2013; Meinhart et al., 2005). Moreover, Pol II transcription pausing is linked to the RNA maturation steps, possibly as a regulatory mechanism ensuring efficient and timely pre-mRNA processing (Alexander et al., 2010; Glover-Cutter et al., 2008; Kireeva et al., 2008; Nudler, 2012; Rasmussen and Lis, 1993).

1.1.1 Pre-mRNA 5' capping

Pre-mRNA molecules are modified with a m7G-cap structure attached to the RNA 5′-end through a 5′-5′-triphosphate link shortly after transcription initiation when the newly generated transcript emerges from the Pol II RNA channel (Cho et al., 1997; Coppola et al., 1983; McCracken et al., 1997a; Rasmussen and Lis, 1993). The 5' cap protects the mRNA from exonucleolytic degradation and is required for the recruitment of splicing, 3' end processing, mRNA export, and translation factors (Osman and Cramer, 2020; Ramanathan et al., 2016). Pre-mRNA capping requires three enzymatic activities, triphosphatase, guanylyltransferase, and methyltransferase activity. After cleavage of the γ-phosphate of the RNA 5'-triphosphate, a guanosine monophosphate is attached and subsequently N7-methylated (Furuichi and Shatkin, 2000; Ghosh and Lima, 2010; Martinez-Rucobo et al., 2015; Rasmussen and Lis, 1993). Pol II transcripts are selectively capped as the responsible enzymes are recruited by the CTD and positioned directly at the RNA exit channel, dependent on CTD hyperphosphorylation by Cdk7 (Cho et al., 1997; Ghosh et al., 2011; Ho and Shuman, 1999; Komarnitsky et al., 2000; McCracken et al., 1997b; Moteki and Price, 2002; Noe Gonzalez et al., 2018). Additionally, capping is stimulated by the DRB-sensitivity-inducing factor (DSIF), a general transcription factor containing SPT4 and SPT5, which is recruited after Pol II promotor escape and required for processive RNA

synthesis (Adelman and Lis, 2012; Lidschreiber et al., 2013; NandyMazumdar and Artsimovitch, 2015; Shetty et al., 2017; Wen and Shatkin, 1999; Werner, 2012).

1.1.2 Pre-mRNA splicing

Almost all human genes contain introns - non-protein-coding regions - within the transcript, which have to be removed to generate a translatable open reading frame (ORF) (Wahl et al., 2009). Introns are removed by the spliceosome, a multi-megadalton RNA-protein complex, which forms anew on each intron from five different snRNPs containing U1, U2, U4, U5, or U6 snRNP (Fica et al., 2013; Steitz et al., 1983; Steitz and Steitz, 1993; Wassarman and Steitz, 1991). These snRNPs assemble on consensus sequence elements at the intron 5'-splice site (5'-SS), internal branch site (BS), and the 3'-splice site (3'-SS) before the intron is removed by two transesterification reactions (Kastner et al., 2019; Plaschka et al., 2019; Wilkinson et al., 2020; Yan et al., 2019). First, the 2'-OH of the branch site adenosine is attacking the phosphodiester bond at the 5'-splice site, generating a free 5' exon and an intron lariat-3' exon. Then, the exons are ligated and the intron lariat is excised by an attack of the 5'-SS 3'-OH on the phosphodiester bond at the 3'-splice site (Wahl et al., 2009).

During spliceosome assembly, 5'-splice site and branch site must be selected with high accuracy, which can be challenging since these sites can be degenerate, are part of introns that are up to one million bases long, and alternative splice sites exist for many genes (Piovesan et al., 2019). Hence, splicing is coupled to transcription to enhance its fidelity and allow intricate regulation (Drexler et al., 2020; Neugebauer, 2019; Papasaikas and Valcárcel, 2016; Wahl et al., 2009). Spliceosomes are assembled on the CTD phosphorylated at Ser-2 and Ser-5 and additionally, U1 snRNP can directly bind the Pol II core (Ahn et al., 2004; Davidson et al., 2014; Harlen et al., 2016; Nojima et al., 2015, 2018; S. Zhang et al., 2021). Recruited to the Pol II surface near the RNA channel, the U1 snRNP interacts with the 5'-splice site and positions it near the RNA exit site while the intron is looped out. This might facilitate scanning for the 3'-splice site and prespliceosome assembly, and highlights the extensive coupling of splicing with transcription (S. Zhang et al., 2021).

1.1.3 Pre-mRNA 3' end processing and transcription termination

With the exception of metazoan replication-dependent histone mRNAs, all eukaryotic protein-coding mRNAs are cleaved and polyadenylated at their 3' ends. Replication-dependent histone mRNAs are only cleaved but not polyadenylated, and a different protein machinery is responsible for 3' end formation. Both processes are discussed in more detail below (chapters **1.2** and **1.3**).

3' end processing or cleavage and polyadenylation (CPA) is a critical step in pre-mRNA maturation as it releases the pre-mRNA from chromatin and the DNA template and initiates transcription termination (Proudfoot, 2016; Shi and Manley, 2015). Accordingly, the CPA machinery is recruited co-transcriptionally to Pol II via multiple interactions with the Ser-2 phosphorylated CTD predominating near gene termination windows (Ahn et al., 2004; Davidson et al., 2014; Kim et al., 2004; Komarnitsky et al., 2000; Licatalosi et al., 2002; Lunde et al., 2010).

Transcription termination is tightly coupled to 3' end formation although Pol II genes have distinct termination profiles; while many genes show gradual termination over multiple kilobases, others terminate abruptly after a poly(A) site (Eaton and West, 2020; Porrua and Libri, 2015; Proudfoot, 2016). Although many open questions remain regarding the exact mechanism, a hybrid model was suggested, merging the previously existing allosteric and torpedo models. The allosteric model proposed that the loss of elongation factors or binding of termination factors would induce conformational changes in Pol II that trigger transcript release (Logan et al., 1987; Zhang et al., 2015). The torpedo model suggested that after cleavage at a poly(A) site, the 5'-3' exoribonuclease 2 (Xrn2) degrades the remaining transcript starting from the newly generated 5' end until it catches up with the transcribing Pol II to induce its release (Connelly and Manley, 1988; Kim et al., 2004; Proudfoot, 1989; Skourti-Stathaki et al., 2014; West et al., 2004; Xiang et al., 2009). Pol II is capable of sensing a functional poly(A) site leading to significantly reduced transcription speeds, caused by PP1-mediated SPT5 dephosphorylation and amplified by features in the chromatin structure (Cortazar et al., 2019; Nojima et al., 2015; Proudfoot, 2016; Zhang et al., 2015). Concordantly, this facilitates Xrn2-mediated termination, although it remains unclear how Xrn2 would expel Pol II from the DNA template (Girbig et al., 2022; Proudfoot, 2016).

The extensive coupling of 3' end formation and transcription termination is also demonstrated by multiple interactions between Pol II and components of the CPA complex. The phosphatases Ssu72 and PP1, both subunits of the 3' end machinery, target Pol II causing changes in the CTD phosphorylation pattern and reduced transcription speeds downstream of poly(A) sites, respectively (Cortazar et al., 2019; Shi et al., 2009; Xiang et al., 2010, 2012). Pcf11, a core factor contributing to RNA motif recognition, directly interacts with the Pol II CTD, thereby connecting the transcription apparatus to poly(A) site selection and stimulating both pre-mRNA cleavage and transcription termination events (Barillà et al., 2001; Kamieniarz-Gdula et al., 2019; Licatalosi et al., 2002). Moreover, loss of CPSF3 endonuclease activity leads to genome-wide termination defects and increased readthrough transcription, clearly connecting 3' end formation with efficient Pol II termination (Eaton et al., 2018).

1.2 Canonical 3' end processing

1.2.1 RNA elements involved in 3' end processing

Canonical 3' end processing is controlled by several *cis* elements located upstream and downstream of the cleavage site (**Fig. 2)** (Tian and Graber, 2012; Tian and Manley, 2017). The core signals for poly(A) sites consist of the hexameric polyadenylation signal (PAS) and a U- or GU-rich sequence surrounding the cleavage site (Chan et al., 2011; Gruber et al., 2016; Neve et al., 2017). Additional auxiliary elements were found upstream (upstream sequence element, USE) and downstream (auxiliary downstream sequence element, aux-DSE or DSE) of the cleavage site, which mediate poly(A) site selection and enhance cleavage efficiency (Chan et al., 2011; Neve et al., 2017; Tian and Graber, 2012; Tian and Manley, 2017).

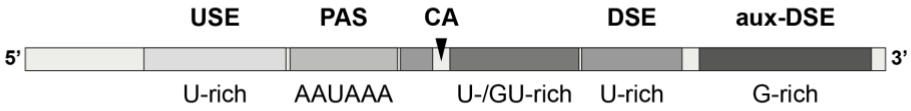

Figure 2: RNA elements in 3' end processing. Several RNA elements define the exact cleavage site (indicated with arrow) in pre-mRNA 3' end formation. (USE) Upstream sequence element, (PAS) Polyadenylation signal, (DSE) Downstream sequence element, (aux-DSE) Auxiliary DSE.

Upstream sequence elements

U-rich auxiliary regions, often UGUA or UAUA motifs, are found in a subset of genes upstream of the cleavage site, providing binding sites for cleavage factor I (CF-I) and thereby enhancing 3' end processing at these sites (Brown and Gilmartin, 2003; Hu et al., 2005; Rüegsegger et al., 1996; Venkataraman et al., 2005; Yang et al., 2010). Although often located ~ 40-50 nucleotides upstream of the PAS, they are found with highly variable positions, copy numbers, and interspacing sequence lengths (Hu et al., 2005; Sun et al., 2020a; Zhu et al., 2018). Many genes are regulated by several of these sequences both up- and downstream of the cleavage site, allowing intricate regulatory mechanisms mediated by cleavage factor I, which can bind two USE motifs simultaneously (Gruber et al., 2016; Hu et al., 2005; Masamha et al., 2014; Yang et al., 2011a, 2011b). Although transcriptome-wide studies identified USE elements as key regulators for poly(A) site selection for a wide range of genes, many details remain unclear (Gruber et al., 2012; Masamha et al., 2014; Neve et al., 2017).

Polyadenylation signal

The critical importance of the polyadenylation signal (PAS) has been underpinned by numerous studies *in vivo* and *in vitro*, as single nucleotide variations or deletion mutations severely impair cleavage efficiency (Fitzgerald and Shenk, 1981; Higgs et al., 1983; Montell et al., 1983; Sheets et al., 1990; Wickens and Stephenson, 1984). The hexameric signal is located within a generally AU-rich area 10-30 nucleotides upstream of the cleavage site, with a dominant peak at positions -21 and -22, and is specifically recognized by the mammalian polyadenylation specificity factor (mPSF) (Chan et al., 2014; Gruber et al., 2016; Legendre and Gautheret, 2003; Schönemann et al., 2014; Sun et al., 2020a; Tian et al., 2005; Wang et al., 2018). Genome-wide studies revealed that the most prevalent PAS variants, AAUAAA and AUUAAA, are present in ~ 60 % and ~ 15 % of genes, respectively (Beaudoing et al., 2000; Gruber et al., 2012; Sun et al., 2020a; Tian et al., 2005). More PAS divergence is achieved by single nucleotide variants, resulting in weaker poly(A) sites as they are bound with reduced affinity by mPSF (Beaudoing et al., 2000; Clerici et al., 2017; Hamilton et al., 2019; Sheets et al., 1990; Tian et al., 2005). Intriguingly, no PAS could be identified for a subset of genes (up to 15 %); thus, despite its importance, the absence of a PAS

can be compensated by other RNA elements (Gruber et al., 2016; Tian et al., 2005; Venkataraman et al., 2005; Zarudnaya et al., 2003).

Cleavage site

The most prevalent cleavage site appears to be after a CA dinucleotide, although it is heterogenous and often features different variants such as UA or GA (Bogard et al., 2019; Sheets et al., 1990; Wang et al., 2018; Sun et al., 2020a). The exact impact of the nucleotide composition at the cleavage site is not fully understood, caused by partially conflicting data from *in vitro* and *in vivo* reports. While initial *in vitro* studies found only very modest effects after introducing mutations in the cleavage site, a prominent single-nucleotide polymorphism (SNP) in the human population from CG to CA significantly increases cleavage efficiency in the prothrombin pre-mRNA (Gehring et al., 2001; Sheets et al., 1990). More recent data suggest a positional rather than a sequence specificity (Gruber et al., 2016; Tian and Graber, 2012; Wang et al., 2018). However, short-distance heterogeneity, where exact poly(A) sites vary up to 20-30 nucleotides, can be observed in different species, although its functional consequences are not clear (Tian and Graber, 2012).

The cleavage site is often surrounded by an additional U-rich element, which controls poly(A) site selection and stimulates the polyadenylation reaction via recruitment of FIP1 (Hu et al., 2005; Hutchins et al., 2008; Kaufmann et al., 2004; Lackford et al., 2014; Li et al., 2015; Martin et al., 2012).

Downstream sequence elements

The core downstream sequence element is usually found within 40 nucleotides of the cleavage site and despite being mostly U- or GU-rich, it is less conserved and more degenerate than the PAS signal (Gil and Proudfoot, 1987; Hu et al., 2005; Hutchins et al., 2008; Levitt et al., 1989; McDevitt et al., 1986; Salisbury et al., 2006; Zarudnaya et al., 2003; Zhao et al., 1999). It is recognized by the cleavage stimulation factor (CstF), which binds two variably interspaced sequence motifs (MacDonald et al., 1994; Takagaki et al., 1996). For a subset of genes, in particular with degenerated PAS, the DSE becomes the critical core element driving CPA machinery assembly and

mediating cleavage efficiency (Chen et al., 2006; Danckwardt et al., 2004; Kandala et al., 2016; Nunes et al., 2010).

Auxiliary downstream sequence elements, commonly G-rich motifs located more than 30 nucleotides downstream of the cleavage site, can further enhance usage of poly(A) sites (Bagga et al., 1995; Hu et al., 2005; Qian and Wilusz, 1991; Sadofsky et al., 1985; Tian and Graber, 2012; Zarudnaya et al., 2003). As they can exert their influence over distances spanning several hundred bases, the prevalence and influence of these auxiliary DSEs are probably underestimated (Dalziel et al., 2007, 2011; Oberg et al., 2005). These motifs are recognized by cleavage factor II (CF-II) and additional regulatory factors stimulating processing, such as hnRNPH (Arhin et al., 2002; Baejen et al., 2017; Dalziel et al., 2007; Schäfer et al., 2018).

1.2.2 Protein factors involved in 3' end processing

As a central and intricately regulated step in mammalian gene expression, 3' end processing is mediated by a large multi-subunit machinery. More than 85 proteins were proposed to be involved in specific activity, among them over 50 factors mediating crosstalk to other nuclear processes (Shi et al., 2009). The core CPA machinery, containing endonuclease, poly(A) polymerase, and Pol II CTD phosphatase activities, consists of several modules, which are thought to be conformationally dynamic to accommodate the distance variability between RNA motifs. Various single- and multi-protein factors are thought to be required for efficient processing, including the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I and II (CF-I and CF-II), retinoblastoma-binding protein 6 (RBBP6), or poly(A) polymerase α (PAPOA; **Table 1; Fig. 3**) (Kumar et al., 2019; Sun et al., 2020a). However, a minimal set of proteins necessary and sufficient for pre-mRNA cleavage *in vitro* is not defined yet.

Table 1: Mammalian 3' end processing machinery. Enzymes are in bold.

Module		Protein (M. W. in kDa)		Proposed role
CPSF	mPSF	CPSF1/CPSF160	(161)	Scaffold
		WDR33	(146)	Scaffold, RNA binding
		CPSF4/CPSF30	(30)	RNA binding

	mCF	FIP1	(67)	PAPOA recruitment, RNA binding
		Symplekin	(141)	Scaffold
		CPSF2/CPSF100	(88)	(Pseudo-)endonuclease, mPSF binding
		CPSF3/CPSF73	(77)	Endonuclease
		CstF2/CstF64	(61)	RNA binding
CstF		CstF1/CstF50	(48)	Dimerization
		CstF2/CstF64	(61)	RNA binding
		CstF3/CstF77	(83)	Scaffold, dimerization
CF-I		CPSF5/CF-Im25/NUDT21	(26)	RNA binding
		CPSF6/CF-Im68	(59)	RNA binding
		CPSF7/CF-Im59	(52)	RNA binding
CF-II		**Clp1**	(48)	RNA kinase
		Pcf11	(173)	Pol II CTD binding, RNA binding
		RBBP6	(202)	Endonuclease activation
		PAPOA	(83)	Poly(A) polymerase
		PABPN1	(33)	Poly(A) tail binding, PAPOA stimulation
		Ssu72	(23)	Pol II CTD phosphatase
		PP1A	(38)	Pol II CTD phosphatase

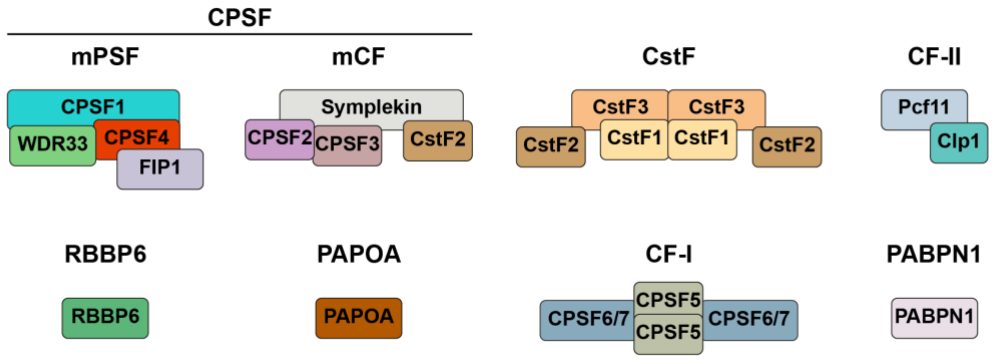

Figure 3: Core factors in 3' end processing. Schematic depiction of core components involved in 3' end processing grouped into distinct modules.

CPSF - mPSF

The cleavage and polyadenylation specificity factor (CPSF) can be further divided into two functional modules, the mammalian polyadenylation specificity factor (mPSF) and the mammalian cleavage factor (mCF). The mPSF consists of CPSF1, WDR33, CPSF4, and FIP1 and provides specificity to the 3' end processing step by recognizing the PAS sequence (**Fig. 4**) (Chan et al., 2014; Schönemann et al., 2014). Acting as central core of the CPA machinery, other cleavage and polyadenylation factors assemble around it (Chan et al., 2014; Schönemann et al., 2014; Zhang et al., 2020).

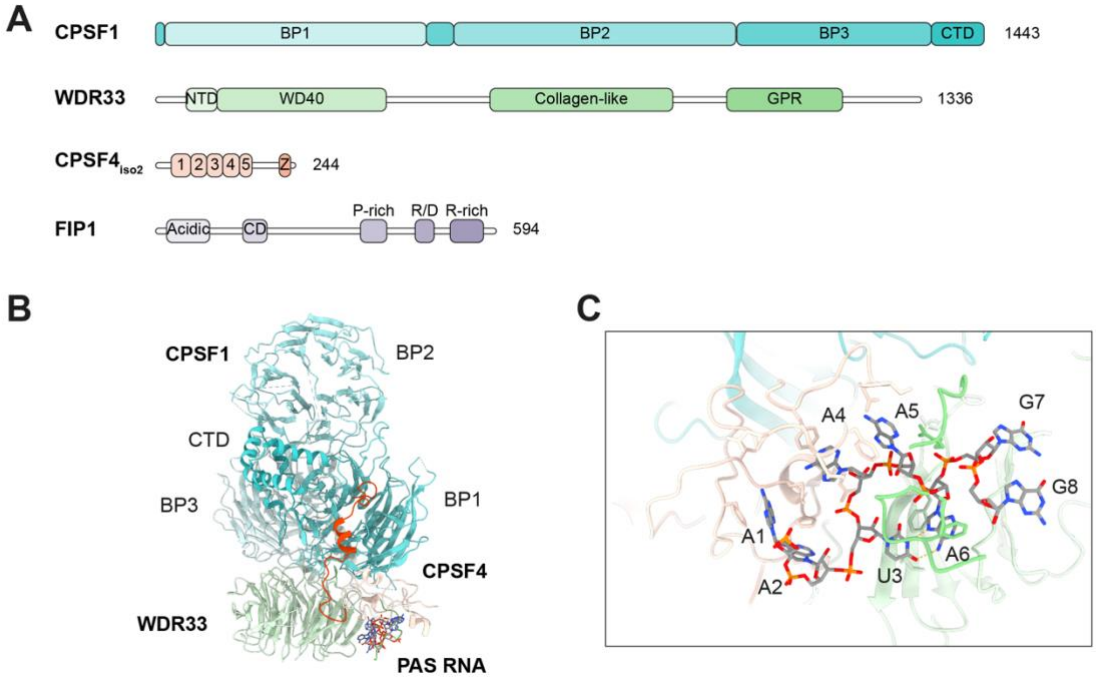

Figure 4: The mammalian polyadenylation specificity factor (mPSF). (A) Domain organization of mPSF subunits. (BP) ß-propeller, (CTD) C-terminal domain, (NTD) N-terminal domain, (GPR) Glycine-Proline-Arginine domain, (1-5) Zinc fingers 1-5, (Z) Zinc knuckle, (CD) Conserved domain, (P-rich) Proline-rich, (R/D) Arginine/Aspartate-rich, (R-rich) Arginine-rich. **(B)** Cryo-EM structure of mPSF bound to a PAS RNA (PDB: 6FBS). **(C)** PAS motif recognition by WDR33 and CPSF4. WDR33 and CPSF4 are shown in cartoon and PAS RNA as sticks. The Hoogsteen base pairing interaction between U3 and A6 is indicated with dotted lines.

CPSF1, which was originally identified as the RNA-binding protein within mPSF, is now established as a scaffold protein, which recruits and pre-organizes WDR33 and CPSF4 for high-affinity RNA-binding (**Fig. 4B**) (Clerici et al., 2017, 2018; Murthy and Manley, 1995; Sun et al., 2018). As a multi-domain protein, it features three WD40 ß-propellers (BP1, BP2, BP3), known protein-protein interaction domains, and a helical C-terminal domain (CTD) (Neer et al., 1994; Sun et al., 2018).

WDR33 stably associates with CPSF1 via its N-terminal (NTD) and WD40 domains, which are additionally involved in RNA-binding. While the NTD protrudes into a cavity between CPSF1 BP1 and BP3, the WD40 domain docks onto their surface (**Fig. 4B, C**) (Clerici et al., 2017, 2018; Sun et al., 2018). The very long, unstructured C-terminus, which is absent in yeast, has no known function and is dispensable for *in vitro* cleavage activity (Boreikaite et al., 2022; Ohnacker et al., 2000; Schmidt et al., 2022).

CPSF4 consists of five CCCH zinc fingers and a C-terminal zinc knuckle of unknown function, which is not present in yeast (Barabino et al., 1997). Cryo-EM analysis

revealed that only its N-terminus including ZF1-3 are stably engaging CPSF1-WDR33 in presence of RNA. The N-terminal segment, which protrudes into the CPSF1 BP3 cavity, and ZF1 are necessary and sufficient for the interaction with CPSF1-WDR33, while ZF2-3 are directly involved in RNA recognition (**Fig. 4B, C**) (Clerici et al., 2017, 2018; Sun et al., 2018). None of the available cryo-EM structures could visualize CPSF4 ZF4-5, indicating that they are highly flexible. ZF4-5 bind to FIP1, which in turn recruits PAPOA, in a CPSF4:FIP1 stoichiometry of 1:2 (Barabino et al., 2000; Hamilton and Tong, 2020). Crystal structures combined with biophysical methods revealed that both zinc fingers can simultaneously interact with one FIP1 molecule each, although the biological relevance remains unclear (Hamilton and Tong, 2020; Muckenfuss et al., 2022).

FIP1 (factor interacting with PAPOA 1) is an intrinsically disordered protein recruiting the poly(A) polymerase to mPSF (Kaufmann et al., 2004). Together with CPSF4 ZF4-5, it generates a long and highly flexible tether, probably allowing PAPOA to stay mPSF-bound despite a growing poly(A) tail (Ezeokonkwo et al., 2011; Kumar et al., 2021). Although both FIP1 copies bound to CPSF4 can engage with PAPOA, only one polymerase is recruited to mPSF at once (Hamilton and Tong, 2020; Muckenfuss et al., 2022). PAPOA is recruited via the N-terminal acidic domain, which also binds CstF3 in a mutually exclusive manner, possibly representing different stages in the cleavage and polyadenylation cycle (Kaufmann et al., 2004; Muckenfuss et al., 2022). The conserved domain (CD) assembles with CPSF4, while the more C-terminal R/D domain mediates interactions to other CPA factors such as CF-I or CPSF1 (Kaufmann et al., 2004; Zhu et al., 2018). The R/D domain and the C-terminal R-rich domain, which binds to U-rich sequences near the cleavage site, are absent in human isoform 4 and in yeast (Kaufmann et al., 2004).

The central PAS sequence is recognized by mPSF subunits WDR33 and CPSF4, via a combination of RNA backbone- and nucleobase-specific interactions (**Fig. 4C**) (Clerici et al., 2018; Sun et al., 2018). The PAS RNA adopts a S-shaped conformation triggered by the intramolecular Hoogsteen base pairing U3-A6, which is further stabilized by π-π stacking interactions between two invariant WDR33 Phenylalanine residues. U3 and A6 are not recognized in base-specific manner, however, they are enclosed by a WDR33 pocket which does not fit other purine-pyrimidine combinations, explaining their PAS sequence conservation. The A1, A2 and A4, A5 bases are bound

by CPSF4 ZF2 and ZF3, respectively. Strictly conserved aromatic CPSF4 residues stack against all four bases, while A1, A4, and A5 are additionally stabilized by two sequence-specific hydrogen bonding interactions with their N1 and N6-amino groups. In contrast, nucleobase A2 only forms a single hydrogen bond to CPSF4, consistent with transcriptome-wide studies showing that the second PAS position is the most variable (Clerici et al., 2018; Derti et al., 2012; Gruber et al., 2016; Hu et al., 2005; Sun et al., 2018).

CPSF - mCF

The mammalian cleavage factor (mCF) containing Symplekin, CPSF2, CPSF3, and CstF2 is the endonuclease complex responsible for cleaving all pre-mRNA substrates; both in canonical and replication-dependent histone pre-mRNA 3' end processing (**Fig. 5A**) (Dominski and Marzluff, 2007; Sun et al., 2020a). It forms a highly flexible, trilobal structure in its inactive state and is specifically recruited to its pre-mRNA substrates by mPSF (Zhang et al., 2020). Its role in histone pre-mRNA 3' end processing is described later (chapter **1.3**).

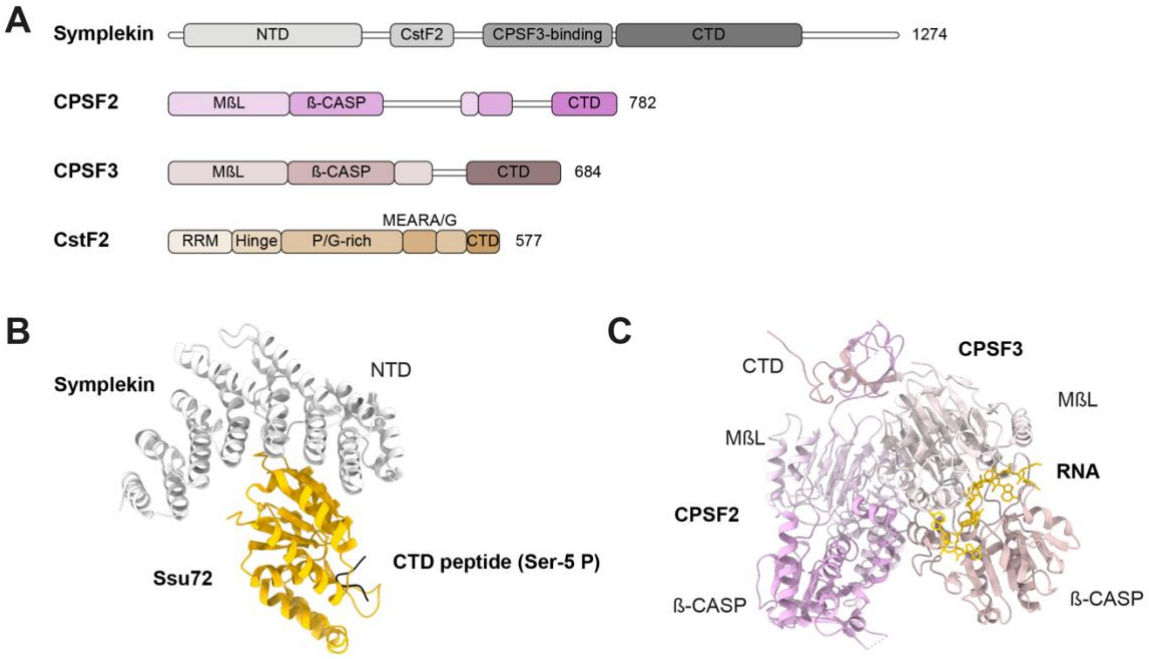

Figure 5: The mammalian cleavage factor (mCF). (A) Domain organization of mCF subunits. (NTD) N-terminal domain, (CTD) C-terminal domain, (MßL) Metallo-ß-lactamase domain, (P/G-rich) Proline/Glycine-rich, (MEARA/G) Methionine-Glutamate-Alanine-Arginine-Alanine-Glycine repeats. **(B)** Crystal structure of the Symplekin NTD-Ssu72 complex bound to a Pol II CTD peptide (PDB: 3O2Q).

(C) Structure of the CPSF2-CPSF3 dimer bound to RNA based on a cryo-EM structure of the active histone pre-mRNA processing complex (PDB: 6V4X).

Symplekin, a mostly α-helical protein, was originally identified as a tight junction protein and only later as the heat-labile factor required for histone pre-mRNA processing (Gick et al., 1987; Kennedy et al., 2009; Keon et al., 1996; Kolev and Steitz, 2005). Subsequent experiments established it as mCF scaffold protein, binding to CstF2 and CPSF3 in its central region and being required for stable dimerization of CPSF2-3 (Dominski et al., 2005; Ghazy et al., 2009; Ruepp et al., 2011b; Takagaki and Manley, 2000). Additionally, the N-terminal HEAT repeats interact with and stimulate *in vitro* activity of Ssu72, an RNA polymerase II CTD phosphatase, thereby linking 3' end processing to transcription (**Fig. 5B**) (Xiang et al., 2010, 2012).

Responsible for mCF catalytic activity, CPSF3 is an endonuclease of the metallo-ß-lactamase (MßL) superfamily of zinc-dependent hydrolases (Aravind, 1999; Callebaut et al., 2002; Dominski et al., 2013). It consists of an N-terminal catalytic MßL domain followed by a ß-CASP domain controlling access to the active site, which is located at their interface (Mandel et al., 2006b). The C-terminal domain (CTD) mediates heterodimerization with CPSF2, which is stabilized by Symplekin (**Fig. 5C**) (Dominski et al., 2005). The two tightly bound zinc atoms in the active center coordinate a hydroxide ion, which acts as nucleophile during catalysis (Mandel et al., 2006b). However, CPSF3 in isolation or within mCF exhibits an inactive conformation by ß-CASP domain movement creating a very narrow active site channel, which cannot accommodate RNA (Mandel et al., 2006b; Zhang et al., 2020). This probably serves as a regulatory feature preventing spurious off-target activity, although the activation mechanism for canonical 3' end processing remains unclear.

CPSF2, a sequence homologue of CPSF3, is likewise a member of the metallo-ß-lactamase superfamily, however, it lacks residues critical for zinc coordination rendering it inactive (Callebaut et al., 2002; Kolev et al., 2008; Mandel et al., 2006b). Altogether, it shares the same domain organization as CPSF3, however, with the addition of a highly hydrophilic and poorly conserved segment (Mandel et al., 2006a). A short and conserved linear motif within this hydrophilic insertion, termed mPSF-interacting motif (PIM), flexibly tethers CPSF2 and mCF to mPSF (Rodríguez-Molina et al., 2022; Zhang et al., 2020). Accordingly, it is thought that CPSF2 evolved as an adaptor protein regulating CPSF3 activity, consistent with their extensive protein-

protein interactions in an active conformation and the absence of detectable CPSF2 RNA binding (Dominski et al., 2013; Ryan et al., 2004; Sun et al., 2020b).

The role of CstF2 in mCF or the equivalent histone cleavage complex (HCC) is currently not clear (Sun et al., 2020b). CstF2 will be discussed in more detail as subunit of the CstF complex.

CstF

The cleavage and stimulation factor (CstF), a dimer of the heterotrimeric CstF1-CstF2-CstF3 complex, specifically recognizes the downstream sequence element (DSE), a U- or G/U-rich RNA element located 3' of the cleavage site (**Fig. 6A**) (Gilmartin and Nevins, 1991; MacDonald et al., 1994; Takagaki and Manley, 1997). It is required for pre-mRNA cleavage as it stabilizes the CPSF-RNA interaction, but is dispensable for polyadenylation (Takagaki et al., 1990; Weiss et al., 1991). While CstF3 and CstF2 are highly conserved among eukaryotes, CstF1 is only present in multicellular eukaryotes without a known yeast homolog (Mitchelson et al., 1993; Takagaki and Manley, 1994).

CstF3, the largest CstF subunit, acts as scaffolding protein, interacting with CstF1, CstF2, and other 3' end processing factors (Takagaki and Manley, 1994, 2000; Murthy and Manley, 1995; Moreno-Morcillo et al., 2011). It comprises a half a tetratricopeptide repeat (HAT) domain and a largely unstructured C-terminus, which includes the binding sites for CstF1 and CstF2 (Moreno-Morcillo et al., 2011; Takagaki and Manley, 2000; Yang et al., 2018). The HAT domain forms a bow-shaped dimer, mediated by the HAT-C subdomains, which also directly engage the mPSF subunits CPSF1-WDR33 in an asymmetric manner (**Fig. 6B**) (Bai et al., 2007; Legrand et al., 2007; Zhang et al., 2020).

CstF2 represents the RNA-binding activity within CstF, its N-terminal RNA recognition module (RRM) binding RNA with a preference for U and G (Pérez Cañadillas and Varani, 2003; Takagaki et al., 1990; Takagaki and Manley, 1997). Incorporation into a fully assembled CstF complex significantly stimulates the RNA-binding affinity, probably by incorporating two closely spaced RRM domains with limited flexibility (Yang et al., 2018). The adjacent Hinge domain interacts with CstF3 or Symplekin in a mutually exclusive manner, thereby creating two distinct CstF2 subpopulations

within the nucleus, bound to CstF or mCF/HCC (Moreno-Morcillo et al., 2011; Ruepp et al., 2011b; Takagaki and Manley, 2000). The role of the C-terminal domains is currently not known.

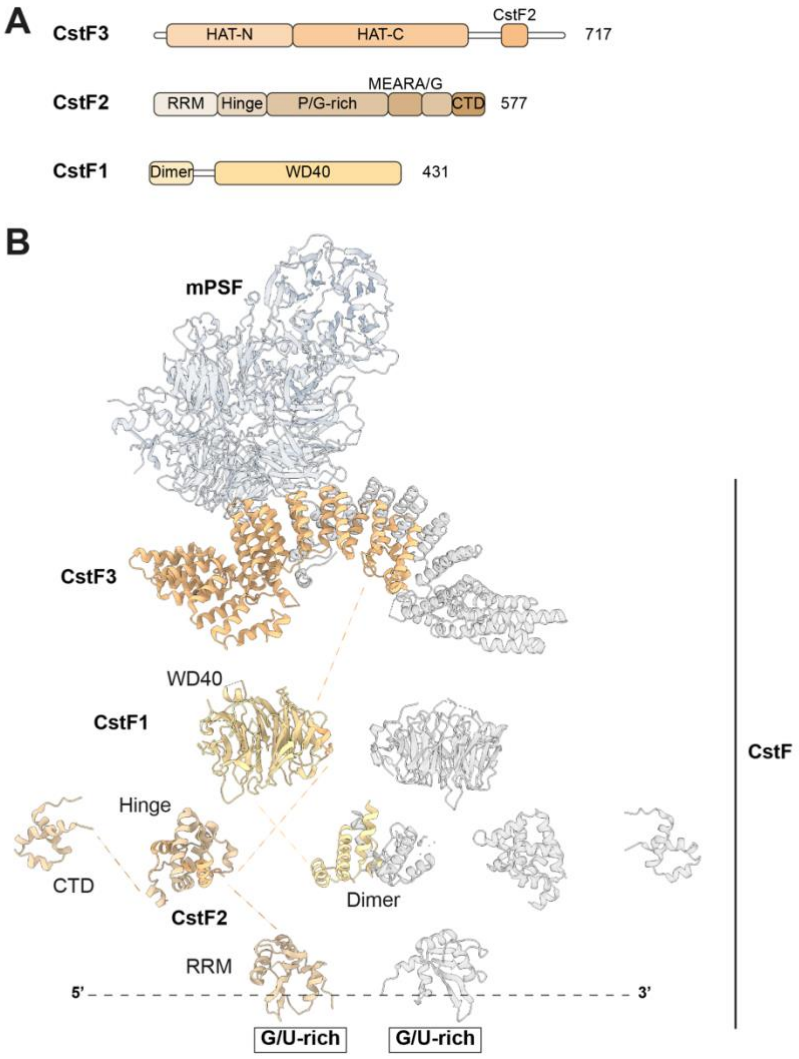

Figure 6: The cleavage stimulation factor (CstF). (A) Domain organization of CstF subunits. (HAT) Half a tetratricopeptide repeat, (P/G-rich) Proline/Glycine-rich, (MEARA/G) Methionine-Glutamate-Alanine-Arginine-Alanine-Glycine repeats, (CTD) C-terminal domain, (Dimer) Dimerization domain. **(B)** Available structures of the CstF complex bound to mPSF. The CstF3 HAT domain interacts with mPSF and connects to the other subunits via flexible linkers, which are indicated with dashed lines. The symmetry copies are shown in grey for clarity (PDBs: 6URO, mPSF-CstF3; 6B3X, CstF1 WD40; 2XZ2, CstF1 Dimer; 2L9B, Rna14-Rna15 (CstF2-3); 2J8P, CstF2 CTD; 1P1T, CstF2 RRM). Updated from (Yang et al., 2018).

Mammals additionally express CstF2τ, a conserved paralog of CstF2 differing mostly in the MEARA/G and C-terminal regions (Wallace et al., 1999). Originally identified as a testis-specific factor regulating alternative polyadenylation, more recent reports describe both variants to be widely expressed in mammalian tissues with varying

levels (Huber et al., 2005; MacDonald, 2019; Wallace et al., 1999; Yao et al., 2013). Having highly similar RNA-binding specificities *in vitro* and *in vivo*, they exert similar but also distinct roles in regulating global alternative polyadenylation profiles (Yao et al., 2013). Although they have mostly overlapping protein interactomes, they exhibit differing binding affinities towards some 3' end processing factors: e.g., CstF2τ does not interact with Symplekin and shows reduced binding to CstF3 (Ruepp et al., 2011b; Yao et al., 2013). Nevertheless, CstF2τ is active in pre-mRNA cleavage, despite slightly lower efficiency, and the functional role of CstF2τ is poorly understood (Schmidt et al., 2022).

CstF1 is the smallest CstF subunit; while its N-terminal domain is mediating homodimerization, the C-terminal seven-bladed WD40 domain is engaged by CstF3 (Moreno-Morcillo et al., 2011; Takagaki and Manley, 2000; Yang et al., 2018). Despite no clear function, it is essential for pre-mRNA cleavage (Schmidt et al., 2022). Using its homodimerization domain, it might limit the overall conformational flexibility of the CstF complex by restricting the unstructured CstF3 C-terminus, thereby contributing to RNA sequence selection (Yang et al., 2018).

CF-I

The U-rich upstream sequence element (USE), although positionally variable often found ~ 40-50 nucleotides upstream of the PAS, is specifically recognized by cleavage factor I (CF-I) (Brown and Gilmartin, 2003; Hu et al., 2005; Rüegsegger et al., 1996; Venkataraman et al., 2005; Yang et al., 2010; Zhu et al., 2018). CF-I is a heterotetrameric complex composed of a dimer of highly conserved CPSF5 along with two molecules of the paralogous proteins CPSF6 or CPSF7 and is considered to be essential for pre-mRNA cleavage (**Fig. 7A**) (Rüegsegger et al., 1996, 1998; Yang et al., 2010; Kim et al., 2010; Ruepp et al., 2011a). As enhancer-dependent activator of 3' end processing, it mediates poly(A) site selection and regulation of alternative polyadenylation, which is also linked to tumor suppression or neurological disorders (Gruber et al., 2012; Kim et al., 2010; Kubo et al., 2006; Li et al., 2015; Martin et al., 2012; Masamha et al., 2014; Masamha, 2022; Tian and Manley, 2017; Zhu et al., 2018; Sartini et al., 2008). It promotes usage of distal poly(A) sites, likely via a direct interaction with mPSF; however, it is currently not fully understood whether poly(A)

site selection is mediated by pre-mRNA looping and subsequent skipping of proximal PAS sites or rather enhanced CF-I-dependent activity at distal PAS sites, where UGUA motifs are often enriched (Li et al., 2015; Rüegsegger et al., 1998; Tian et al., 2005; Venkataraman et al., 2005; Yang et al., 2011a, 2011b; Zhu et al., 2018). Intriguingly, depletion of cleavage factor I proteins CPSF5 and CPSF6, but not CPSF7, leads to widespread shifts to proximal PAS sites and 3' UTR shortening (Gruber et al., 2012; Li et al., 2015; Martin et al., 2012). Overall, as cleavage factor I recognizes two UGUA motifs separated by RNA sequences of highly variable length, it contributes to poly(A) site selection by enhancing or restricting access to RNA elements required for 3' end processing (Tian et al., 2005; Tian and Manley, 2017; Yang et al., 2011a).

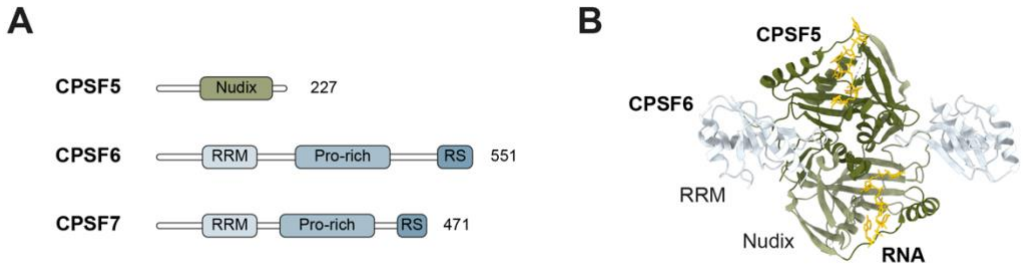

Figure 7: The cleavage factor I (CF-I). (A) Domain organization of CF-I subunits. (Pro-rich) Proline-rich, (RS) Arginine-Serine domain. **(B)** Crystal structure of tetrameric CPSF5-CPSF6 RRM complex bound to UGUA RNA (PDB: 3Q2T).

CPSF5 is a member of the Nudix protein family despite lacking hydrolase activity, which is conserved throughout eukaryotes, yet absent in a subset of protists including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Yang et al., 2010). Albeit having the characteristic Nudix domain fold, it lacks critical Glutamate residues required for the coordination of divalent cations and gained a distinctive α-helix loop motif blocking access to the active site, thus rendering it catalytically inactive (Coseno et al., 2008; McLennan, 2006). Instead, it recognizes RNA in a sequence-specific manner with preference for UGUA motifs (Brown and Gilmartin, 2003; Yang et al., 2010). As CPSF5 dimerizes in an anti-parallel fashion and each monomer recognizes one UGUA motif, the RNA is bound anti-parallel, requiring RNA looping to accommodate its 180˚ turn (**Fig. 7B**) (Brown and Gilmartin, 2003; Yang et al., 2010, 2011a). In addition to its role in alternative polyadenylation, CPSF5 is connected to the splicing machinery as it binds several splicing factors such as U2AF65 or U1 snRNP and was co-purified with spliceosomes, possibly connecting (alternative)

splicing and (alternative) 3' end processing (Awasthi and Alwine, 2003; Millevoi et al., 2006; Yang et al., 2011b; Zhou et al., 2002).

Although CPSF6 and CPSF7 are encoded by different genes, they are highly related in sequence and share a common architecture with an N-terminal RRM domain followed by a central Proline-rich region and a C-terminal RS-like domain (Dettwiler et al., 2004; Martin et al., 2004; Rüegsegger et al., 1998; Ruepp et al., 2011a). The RRM domain directly interacts with CPSF5 and the pre-mRNA, thereby enhancing CF-I RNA binding and facilitating RNA looping (Dettwiler et al., 2004; Yang et al., 2010, 2011a). The C-terminal RS-like domain, a region rich in Arginine-Serine dipeptides commonly found in proteins with important roles in splicing regulation, specifically binds to an RS-like region in the FIP1 R/D domain, thereby mediating the CF-I enhancer-dependent activator function (Nilsen and Graveley, 2010; Venkataraman et al., 2005; Wang et al., 2008; Zhu et al., 2018).

CF-II

Cleavage factor II (CF-II) is an essential 3' end processing factor, which is poorly characterized, probably due to its rather weak or transient interaction with the remaining CPA machinery (**Fig. 8**) (Boreikaite et al., 2022; Schmidt et al., 2022; Shi et al., 2009). The heterodimer of Pcf11 and Clp1 probably recognizes the G-rich auxiliary downstream sequence element and couples 3' end processing with transcription termination by additionally binding to the CTD of RNA polymerase II (Baejen et al., 2017; Barillà et al., 2001; de Vries et al., 2000; Kamieniarz-Gdula et al., 2019; Meinhart and Cramer, 2004; Sadowski et al., 2003; Schäfer et al., 2018).

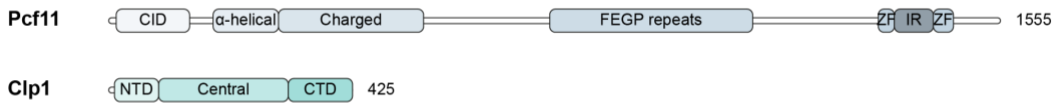

Figure 8: The cleavage factor II (CF-II). Domain organization of CF-II subunits. (CID) CTD-interaction domain, (FEGP repeats) Phenylalanine-Glutamate-Glycine-Proline repeats, (ZF) Zinc finger, (IR) Clp1-interacting region, (NTD) N-terminal domain, (CTD) C-terminal domain.

Pcf11 is a large multi-domain protein with functions in 3' end processing and transcription termination, which can be functionally uncoupled (Sadowski et al., 2003;

Schäfer et al., 2018). The N-terminal Pol II CTD-interaction domain (CID) recognizes Serine-2 phosphorylation, a CTD modification required for efficient 3' end processing *in vivo* (Ahn et al., 2004; Barillà et al., 2001; Meinhart and Cramer, 2004; Ni et al., 2004). The CID is followed by a helical domain of unknown function, a highly charged region, a FEGP repeat region containing 30 repeats of 13 amino acids, and a C-terminal region comprising two zinc fingers, which enclose the Clp1-interacting region (Guéguéniat et al., 2017; Guo et al., 2014; Noble et al., 2007; Schäfer et al., 2018; Xu et al., 2015; Yang et al., 2017). The zinc fingers mediate RNA binding with preference for G-rich sequences and the C-terminal regions encompassing the FEGP repeats and zinc fingers are sufficient for pre-mRNA processing, making the CID dispensable for cleavage *in vitro* (Boreikaite et al., 2022; Schäfer et al., 2018; Schmidt et al., 2022).

Clp1 is an RNA 5' kinase with functions in 3' end processing and tRNA metabolism, as part of the TSEN complex (de Vries et al., 2000; Paushkin et al., 2004; Weitzer and Martinez, 2007). Its kinase activity is dispensable for pre-mRNA cleavage *in vivo* and *in vitro*, consistent with yeast Clp1 lacking catalytic activity (Hanada et al., 2013; Noble et al., 2007; Ramirez et al., 2008; Schäfer et al., 2018). Clp1 was not detected in a mass spectrometry-based approach of an active CPA complex, indicating only weak or transient interaction, and its role in 3' end processing is poorly understood (Shi et al., 2009). However, Clp1 ATP binding is necessary for efficient cleavage, consistent with its ATP-bound state in yeast cells, suggesting a regulatory role (Noble et al., 2007; Schmidt et al., 2022).

RBBP6

Retinoblastoma-binding protein 6 (RBBP6) is a multi-domain and multi-functional protein (**Fig. 9A**). Its N-terminal part shares sequence similarity to yeast Mpe1, while the human protein is considerably longer, gaining additional Rb- and p53-binding sites within its long unstructured C-terminus (Lee and Moore, 2014; Saijo et al., 1995; Sakai et al., 1995; Simons et al., 1997; Vo et al., 2001). It was originally identified as interactor of the tumor suppressor proteins Rb and p53, leading to decreased p53 DNA binding and enhanced p53 ubiquitination and degradation (Simons et al., 1997). The N-terminal fragment, including the domain with no name (DWNN), CCHC zinc knuckle, and Ring finger domains, is found in all eukaryotic genomes including yeast, where Mpe1 is required for mRNA processing (**Fig. 9B**) (Lee and Moore, 2014; Pugh

et al., 2006; Vo et al., 2001). The DWNN represents an unusual ubiquitin-like domain, while the RING E3 ubiquitin ligase domain was shown to ubiquitinate substrates, including the RNA-binding protein YB-1 or transcriptional repressor ZBTB38, leading to their proteasomal degradation (Chibi et al., 2008; Miotto et al., 2014; Pugh et al., 2006).

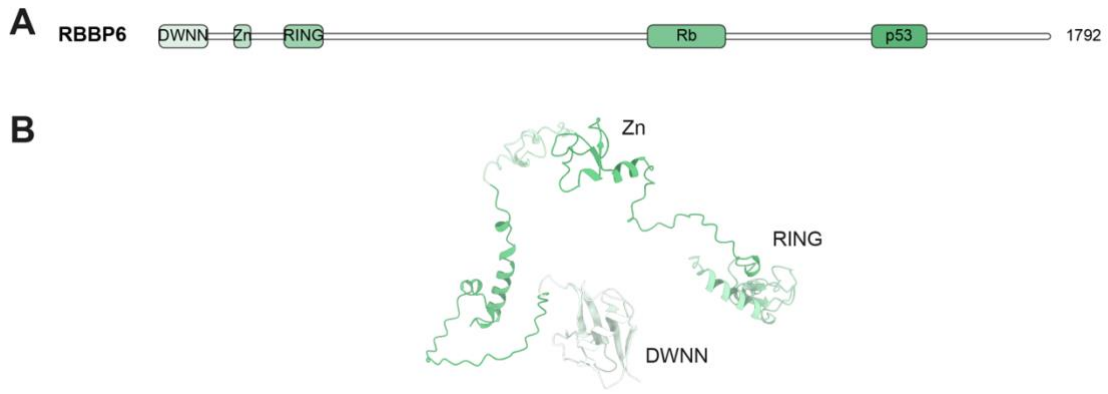

Figure 9: Retinoblastoma-binding protein 6 (RBBP6). (A) Domain organization of RBBP6. (DWNN) Domain with no name, (Zn) Zinc knuckle, (Rb) Retinoblastoma-binding region, (p53) p53-binding region. **(B)** Structural model of the N-terminal truncation construct (residues 1-335) predicted with AlphaFold2 (Jumper et al., 2021).

Additionally, RBBP6 was implicated in pre-mRNA cleavage but is dispensable for polyadenylation, however, its conserved N-terminal region is sufficient for the activity (Di Giammartino et al., 2014; Shi et al., 2009). Knockdown of RBBP6 results in decreased transcript levels, especially for mRNAs with AU-rich 3' UTRs, and increased usage of distal poly(A) sites, suggesting a role in alternative polyadenylation (Di Giammartino et al., 2014). In human cells, a single-domain isoform 3, which solely expresses the DWNN domain and a very short C-terminus, inhibits 3' end processing by competing with the long isoform 1 for binding to core factors, especially CstF2 (Di Giammartino et al., 2014). The importance of RBBP6 in gene expression is further underlined by its absence leading to embryonic lethality in mice, which can be partially rescued by a p53-null mutation, and its downregulation in some cancers resulting in altered 3' end processing patterns (Li et al., 2007; Mbita et al., 2012).

PAPOA

Poly(A) polymerase α (PAPOA), the enzyme responsible for the synthesis of the poly(A) tail, is a template-independent nucleotide transferase of the DNA polymerase ß family (Edmonds, 1990; Wahle, 1991a). Its structured portion contains the N-terminal catalytic domain, a central domain, and a globular RNA-binding domain and is conserved throughout eukaryotes (**Fig. 10A**) (Martin and Keller, 1996; Martin et al., 1999, 2000; Zhelkovsky et al., 1995). In the active center, which is located at the bottom of a large cleft between the folded domains, three conserved Aspartate residues coordinate magnesium ions required for catalysis (**Fig. 10B**) (Martin and Keller, 1996; Martin et al., 2000). These allow a nucleophilic attack of the RNA 3'-OH on the ATP α-phosphate resulting in AMP incorporation and pyrophosphate release, thereby showing very high nucleotide specificity towards ATP both *in vitro* and *in vivo* (Legnini et al., 2019; Martin et al., 2000; Wahle, 1991a). In vertebrates, the protein exhibits an extended C-terminus rich in Serine and Threonine residues, which has important regulatory roles mediated by post-translational modifications, including phosphorylation, acetylation, or sumoylation (Colgan et al., 1996; Martin and Keller, 1996; Mizrahi and Moore, 2000; Shimazu et al., 2007; Vethantham et al., 2007; Wahle and Rüegsegger, 1999).

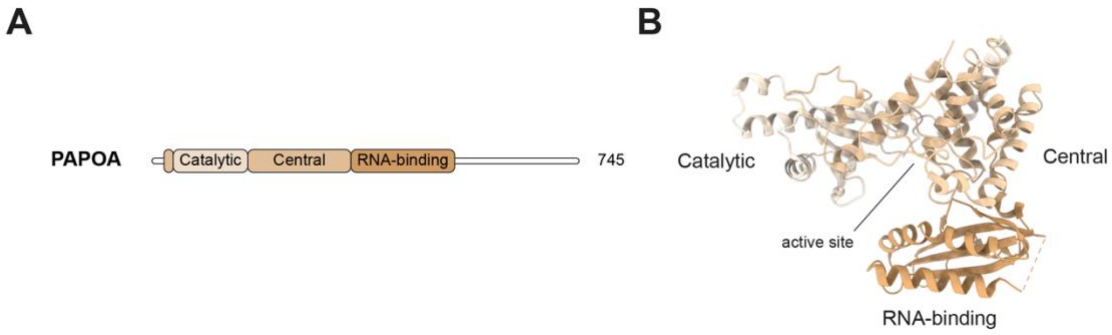

Figure 10: The poly(A) polymerase α (PAPOA). (A) Domain organization of PAPOA. **(B)** Crystal structure of the PAPOA catalytic domain (PDB: 1Q79). The active site is indicated.

In isolation, PAPOA features only slow and distributive polyadenylation activity, due to its weak and non-specific RNA-binding properties (Wahle, 1991a). The enzyme, which shows slight preference for A at the substrate 3' end, gains fast and processive activity through combined stimulation by mPSF and PABPN1, which tether PAPOA stably to pre-mRNAs (Bienroth et al., 1993; Kerwitz et al., 2003; Takagaki et al., 1988, 1989;

Wahle, 1991b; Wahle and Keller, 1992; Gilmartin and Nevins, 1989). After addition of ~ 250 A nucleotides, disruption of the interaction of PAPOA with mPSF likely stops processive activity and polyadenylation terminates (Bienroth et al., 1993; Eckmann et al., 2011). Controlled termination is thought to rely on a counting mechanism, as the PABPN1-coated poly(A) tail collapsing into spherical structures allows maintaining the contact between mPSF and PAPOA. However, addition of ~ 250 nucleotides onto the growing poly(A) tail leads to disruption of this contact and rapid, processive elongation stops (Eckmann et al., 2011; Kühn et al., 2009). PAPOA contains multiple binding sites within mPSF, directly interacting with CPSF1 and FIP1, although the two FIP1 copies present in mPSF only recruit one polymerase (Kaufmann et al., 2004; Muckenfuss et al., 2022; Murthy and Manley, 1995). Nevertheless, PAPOA is not a stable core component of mPSF in cells, different than in yeast (Casañal et al., 2017; Chan et al., 2014; Kaufmann et al., 2004).

PABPN1

The poly(A)-binding protein nuclear 1 (PABPN1) is a mainly nuclear poly(A)-binding protein, which can shuttle to the cytoplasm (Calado et al., 2000). It is mainly involved in pre-mRNA polyadenylation, but a plethora of different functions were described in recent years (Kühn and Wahle, 2004; Wigington et al., 2014). PABPN1 is conserved through eukaryotes, however absent in *Saccharomyces cerevisiae* (Banerjee et al., 2013; Winstall et al., 2000). It consists of a central coiled-coil domain followed by an RNA-binding domain (RRM), which separate a Glutamate-rich N-terminus from the Arginine-rich C-terminus (**Fig. 11**) (Kühn and Wahle, 2004). The RRM and C-terminal regions are necessary for high-affinity poly(A) RNA binding with a footprint of ten or eleven nucleotides, which are recognized in a base-specific manner (Kühn et al., 2003; Kühn and Wahle, 2004; Meyer et al., 2002; Nemeth et al., 1995; Smith et al., 1999). PABPN1 stimulates the activity of poly(A) polymerase by tethering the enzyme to its RNA substrate via a protein-protein interaction mediated by the coiled-coil domain (Bienroth et al., 1993; Kerwitz et al., 2003; Wahle, 1991b). The charged termini are involved in oligomerization, resulting in formation of either fibrils (without RNA) or spherical particles (with RNA) *in vitro* (Keller et al., 2000; Kühn et al., 2003; Meyer et al., 2002). These spherical structures were proposed to be crucial for poly(A) tail length control and a regulated termination of processive polyadenylation, however the details

are poorly understood (Kühn and Wahle, 2004; Kühn et al., 2009; Keller et al., 2000). Additionally, PABPN1 was implicated in a diverse set of nuclear RNA processing steps, including poly(A) site selection in alternative polyadenylation, quality control pathways as the poly(A) tail exosome targeting connection (PAXT), or mRNA export (Apponi et al., 2010; Bresson and Conrad, 2013; Dower et al., 2004; Jenal et al., 2012; Meola et al., 2016; Poon et al., 2000).

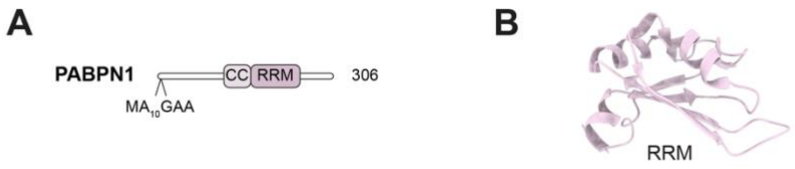

Figure 11: The poly(A)-binding protein nuclear 1 (PABPN1) (A) Domain organization of PABPN1. The N-terminal oligo-Alanine stretch (A10) is indicated. (CC) Coiled-coil domain. **(B)** Crystal structure of PABPN1 RRM domain (PDB: 3UCG).

The initiating Methionine is immediately followed by an oligo-Alanine stretch (A10), which is extended to at least 12 Alanine in oculopharyngeal muscular dystrophy (OPMD), a late-onset human genetic disease leading to muscle weakness mostly in eye lids and pharynx (Banerjee et al., 2013; Brais et al., 1998). The repeat expansion disorder causes intranuclear PABPN1 aggregation in form of regular filaments, sequestering monomeric PABPN1 and potentially other proteins, which is limited to skeletal muscle cells despite its ubiquitous expression (Calado et al., 2000; Hino et al., 2004). Although OPMD patient cells do not show severe polyadenylation defects, a mouse model expressing an extended PABPN1 variant (A17) results in genome-wide shifts to proximal poly(A) sites (Calado et al., 2000; de Klerk et al., 2012). Overall, it remains unclear how the modest Alanine expansion leads to a tissue-specific disease, and whether PABPN1's role in nuclear pre-mRNA polyadenylation is contributing to the disease phenotype.

1.2.3 Alternative polyadenylation

Alternative polyadenylation (APA) allows the expression of distinct transcript isoforms from a single gene dependent on cellular conditions, thereby substantially increasing transcriptome diversification (Gruber and Zavolan, 2019; Reyes and Huber, 2018). This process describes the regulated selection of a defined poly(A) site for one

transcript when multiple poly(A) sites are present, leading to different mRNA isoforms (Edwalds-Gilbert et al., 1997). Dependent on the poly(A) site location, alternative polyadenylation results in different 3' UTRs or terminal exons, thereby possibly affecting both mRNA and protein diversification (**Fig. 12**) (Ren et al., 2020; Tian and Manley, 2013; Y. Zhang et al., 2021). The primary protein structure is affected if a poly(A) site upstream of the terminal exon in coding or intronic regions is selected, a process called coding-region APA (CR-APA) and intronic APA (IPA), respectively (Mohanan et al., 2022; Pereira-Castro and Moreira, 2021). UTR-APA is defined as poly(A) site selection in the terminal exon downstream of the translation stop codon, changing 3' UTR end and length. This can have a plethora of different functional consequences as 3' UTRs affect mRNA stability, translation efficiency, mRNA subcellular or tissue localization, and protein interaction and function (Berkovits and Mayr, 2015; Di Giammartino et al., 2011; Elkon et al., 2013; Gruber and Zavolan, 2019; Mayr, 2017; Neve et al., 2017; Reyes and Huber, 2018; Tian and Manley, 2017).

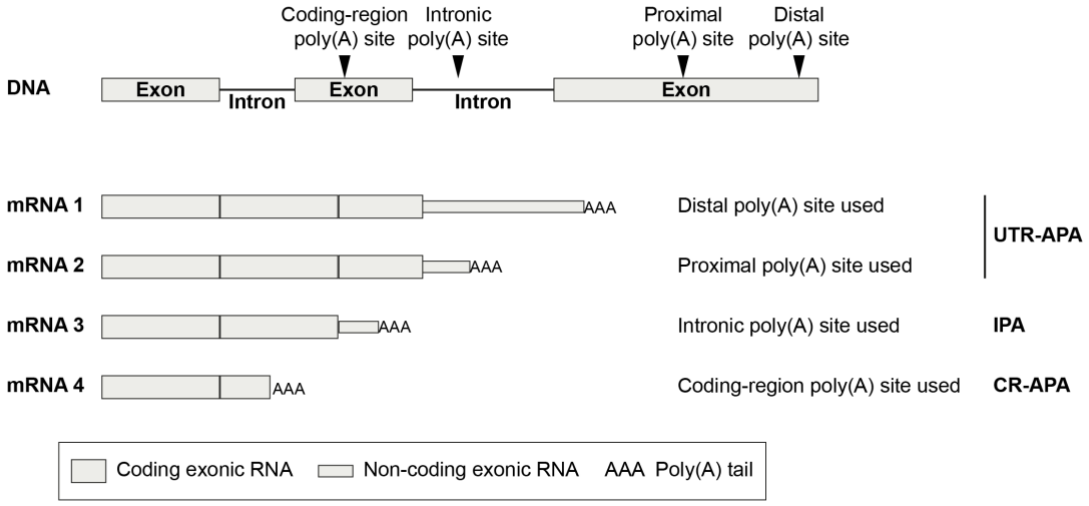

Figure 12: Alternative polyadenylation sites in a gene. A gene transcript can have multiple poly(A) sites; for simplicity, only two poly(A) sites are shown in the terminal exon. Processing at proximal or distal poly(A) sites in the terminal exon leads to different 3' UTR lengths, while use of coding-region or intronic poly(A) sites upstream of the last exon leads to mRNA isoforms with altered protein sequence.

In higher eukaryotes, alternative polyadenylation is highly prevalent, in humans at least 70 % of genes contain more than one poly(A) site (Ozsolak et al., 2010; Shepard et al., 2011; Tian et al., 2005; Tian and Manley, 2017; Wu et al., 2011; Yan and Marr, 2005). Different poly(A) sites within one transcript are recognized with varying efficiencies, accordingly termed weak or strong poly(A) sites. The 3' most poly(A)

sequences are typically strong and contain more frequently an AAUAAA PAS motif, presumably to ensure proper transcription termination (Tian and Graber, 2012; Tian and Manley, 2017). 3' UTRs have expanded during animal evolution, suggesting a corresponding increase in the complexity of post-transcriptional regulation (Mayr, 2016). Consistently, alternative polyadenylation was linked to the establishment of cell identity and additionally controls cell fate, cancer progression and many pathological conditions (Brumbaugh et al., 2018; Gruber et al., 2018; Lianoglou et al., 2013; Xia et al., 2014).

Poly(A) site choice is regulated by core 3' end processing factors or specialized RNA-binding proteins, particularly splicing factors allow intricate cross-regulation between polyadenylation and splicing (Gruber and Zavolan, 2019; Tian et al., 2005, 2007). They mediate poly(A) site selection either in gene-specific or transcriptome-wide manner, exemplified by an average trend towards more proximal sites during cell proliferation and more distal sites during cell differentiation (Gruber and Zavolan, 2019). Core 3' end formation factors promoting proximal poly(A) sites include Pcf11, RBBP6, CstF2, or FIP1, while CPSF5, CPSF6, PABPN1, or U1 snRNP lead to increased use of distal sites (Gruber and Zavolan, 2019; Pereira-Castro and Moreira, 2021; Tian and Manley, 2017). The exact molecular mechanism favoring particular poly(A) sites is often unclear but includes modulation of the expression levels of core factors or changing the accessibility of RNA sequence motifs (Gruber and Zavolan, 2019; Tian and Manley, 2017). One better characterized factor is U1 snRNP, which suppresses premature CPA at intronic, cryptic poly(A) sites, a process called telescripting, which is required for full-length gene transcription (Kaida et al., 2010; Ran et al., 2021; So et al., 2019). Accordingly, U1 knock-down or inhibition using antisense oligos leads to an increased use of poly(A) sites in upstream intronic regions and premature transcription termination (Kaida et al., 2010; Oh et al., 2017; So et al., 2019). Collectively, U1 snRNP has separate functions in pre-mRNA splicing and 3' end formation, additionally providing an interconnection between these processes (Ran et al., 2021).

1.3 Replication-dependent histone pre-mRNA processing

Metazoan replication-dependent histone mRNAs are the only known eukaryotic mRNAs which do not carry a poly(A) tail, instead they end with a highly conserved

stem-loop structure (Dominski and Marzluff, 2007). During the S phase of the cell cycle, synthesis of all five histone proteins increases drastically reflecting the necessity to rapidly package the newly replicated DNA into chromatin (Marzluff, 2005; Osley, 1991). This is achieved by a combination of transcriptional activation with co- and post-transcriptional mechanisms, leading to a ~ 40-fold increase of mature histone mRNA in the cytoplasm (Harris et al., 1991; Stein et al., 1996). Transcriptional activation is facilitated by gene clusters containing ~ 50 genes for the five histone proteins while mRNA maturation is greatly accelerated by the absence of introns; hence, a single cleavage reaction is sufficient to create mature histone mRNAs that are subsequently exported and translated (Dominski and Marzluff, 2007; Marzluff and Duronio, 2002; Romeo and Schümperli, 2016). Intriguingly, several replication-dependent histone genes contain at least one canonical poly(A) site downstream of the stem-loop structure, which is used in small amounts outside of S phase, probably to produce low numbers of histone proteins throughout the cell cycle (Romeo et al., 2014).

Histone pre-mRNA cleavage is defined by two critical sequence elements, the conserved stem-loop and the histone downstream element (HDE) (Birchmeier et al., 1984; Dominski et al., 1999; Mowry and Steitz, 1987; Mowry et al., 1989; Vasserot et al., 1989). The stem-loop recruits the stem-loop binding protein (SLBP) while the HDE forms an RNA duplex with U7 snRNA, thereby recruiting the Lsm-bound U7 snRNP (Galli et al., 1983; Scharl and Steitz, 1994; Strub et al., 1984; Sun et al., 2020b). After assembly of the full processing machinery, the pre-mRNA is cleaved between both critical structures, typically five nucleotides downstream of the stem-loop after an adenosine (**Fig. 13**) (Dominski and Marzluff, 2007; Furger et al., 1998; Scharl and Steitz, 1994; Sun et al., 2020b). The responsible enzyme is CPSF3, embedded in the histone cleavage complex (HCC), the same complex as mCF in canonical pre-mRNA cleavage (Dominski et al., 2005; Dominski and Marzluff, 2007; Romeo and Schümperli, 2016; Sun et al., 2020b).

The U7 snRNP is formed by the short U7 snRNA bound by an unusual, heptameric Sm ring containing the unique subunits Lsm10 and Lsm11, which replace SmD1 and SmD2 of the spliceosomal Sm ring (Dominski and Marzluff, 2007; Pillai et al., 2001, 2003). Lsm11 exhibits an unusually long N-terminal extension that is required to bind the FLICE-associated huge protein (FLASH), which in turn is needed to recruit the HCC to the U7 snRNP-FLASH complex and connects the HCC component Symplekin

to the SLBP-bound 3' stem-loop (Aik et al., 2017; Sabath et al., 2018; Skrajna et al., 2018; Sun et al., 2020b; Yang et al., 2009).

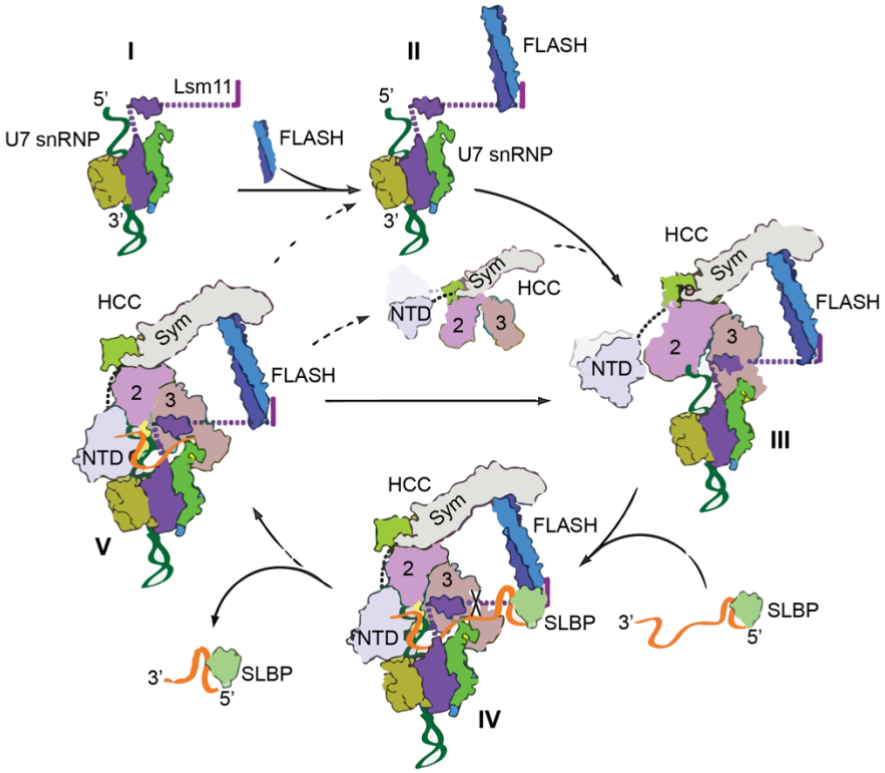

Figure 13: Histone pre-mRNA processing cycle. Putative model for histone pre-mRNA processing. U7 snRNP **(I)** recruits FLASH **(II)** and HCC **(III)**. After substrate binding, CPSF3 endonuclease activity is induced and the pre-mRNA is cleaved **(IV)**. After substrate release, the downstream product will be degraded **(V)** and the processing machinery is either re-used or disassembled. Adapted from (Sun et al., 2020b).

As revealed by a recent cryo-EM structure, the HDE-U7 duplex found in the center of the complex forms 12 consecutive Watson-Crick base pairs, which are surrounded by the HCC (**Fig. 14**) (Sun et al., 2020b). The CPSF3 MßL, CPSF2 ß-CASP, and the Symplekin NTD domains are enclosing the duplex from three side without contacting the bases, consistent with earlier data showing that base pairing rather than the exact sequence is important (Aik et al., 2017; Dominski and Marzluff, 2007; Romeo and Schümperli, 2016; Sun et al., 2020b). The pre-mRNA is bound in the CPSF3 active center; the scissile phosphate coordinated by two zinc ions and the target adenosine oriented by hydrogen-bonding interactions to its N1 and N6 atoms, consistent with the nucleotide preference at the cleavage site (Dominski and Marzluff, 2007; Romeo and Schümperli, 2016; Sun et al., 2020b). A large structural rearrangement of the CPSF3 ß-CASP relative to the MßL domain is necessary to form an open, active CPSF3

conformation with a narrow canyon large enough to accommodate single-stranded RNA (Sun et al., 2020b; Zhang et al., 2020).

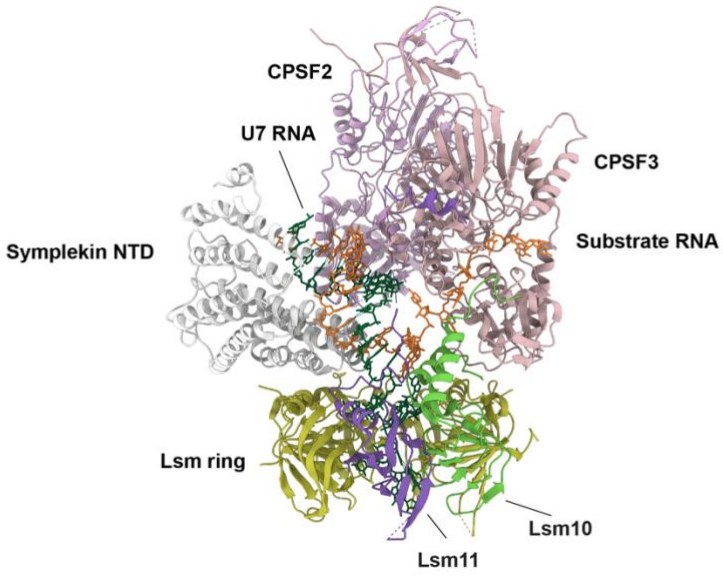

Figure 14: Histone pre-mRNA processing complex. Cryo-EM structure of the histone pre-mRNA processing complex core, colored as in **Fig. 13**. The Symplekin NTD, CPSF2, and CPSF3 enclose the U7 snRNP-substrate RNA duplex (PDB: 6V4X).

While the Symplekin N-terminal domain is essential for activity, its possible binding partner Ssu72 has an inhibitory effect, in contrast to its stimulatory activity observed in canonical 3' end processing (He et al., 2003; Wani et al., 2014). Structural analysis revealed that Ssu72 would clash with both the RNA duplex and CPSF3, explaining the inhibitory mechanism and suggesting that Ssu72 might allow targeted endonuclease regulation dependent on the type of pre-mRNA substrate (Sun et al., 2020b).

1.4 Diseases connected to 3' end processing defects

As 3' end formation is a central processing step affecting almost all mRNA molecules, its malfunction or inhibition is connected to a plethora of human diseases (Gruber and Zavolan, 2019; Mohanan et al., 2022). Some influenza virus strains were shown to shut-down host antiviral responses by targeting CPA factors including CPSF4 or PABPN1 to prevent their interaction with RNA via a NS1 protein-mediated direct competition mechanism (Chen et al., 1999; Das et al., 2008; Twu et al., 2006, 2007). Moreover, many immunological, hematological, or neurological diseases, and cancers

are characterized by altered poly(A) site usage, caused by either single nucleotide polymorphisms (SNPs) changing individual poly(A) sites or genome-wide perturbations of alternative polyadenylation patterns (Chang et al., 2017; Curinha et al., 2014; Gruber and Zavolan, 2019; Mohanan et al., 2022). SNPs in CPA sequence elements can result in either loss-of-function, as a PAS A to G mutation in α- and β-thalassemia patients causing low mRNA levels, or gain-of-function mutations, exemplified by systemic lupus erythematosus (SLE), where a new proximal poly(A) site in the interferon regulatory factor 5 (IRF5) transcript leads to 3' UTR shortening and increased protein levels causing the disease (Graham et al., 2007; Hellquist et al., 2007; Higgs et al., 1983; Orkin et al., 1985). Alternatively, mutations can change mRNA expression levels by elevating the pre-mRNA processing efficiencies while maintaining 3' UTR composition. Some thrombophilia patients carry a CG to CA mutation at the cleavage site of the prothrombin gene transcript causing more efficient 3' end formation, higher mRNA expression and protein levels (Ceelie et al., 2004; Ferraresi et al., 1997; Gehring et al., 2001; Poort et al., 1996; Ridker et al., 1999).

Many human cancers demonstrate widespread changes in alternative polyadenylation, leading to a vast set of different effects dependent on specific transcript isoforms (Gruber et al., 2018; Mohanan et al., 2022; Singh et al., 2009; Xia et al., 2014). Most cancers feature global 3' UTR shortening, consistent with its association with higher proliferative states (Gruber et al., 2014, 2018; Xia et al., 2014; Xue et al., 2018; Sandberg et al., 2008; Elkon et al., 2012). This transcriptome-wide shift is mediated by a combination of globally upregulated 3' end formation, specific activities of core processing factors influencing APA including CstF2 or CPSF5, and the extensive use of intronic poly(A) sites causing truncated protein isoforms, especially for tumor suppressor proteins (Elkon et al., 2012; Lee et al., 2018; Masamha et al., 2014; Mohanan et al., 2022; Xia et al., 2014). Accordingly, JTE-607, a small molecule compound binding to CPSF3 and restricting access to its mRNA substrate channel, acts as an anti-tumor drug targeting a subset of acute myeloid leukemia (AML) and Ewing's sarcoma cell lines (Ross et al., 2020).

1.5 Aim of this thesis

The 3' ends of almost all mammalian mRNAs are generated by an essential two-step mechanism, cleavage and polyadenylation, which releases the pre-mRNA from its DNA template and is crucial for transcription termination. While biochemical and structural studies have identified the respective enzymes and uncovered general features of their activity, many details remain unanswered. How the endonuclease is activated after substrate recognition it not understood. Insights into the architecture of active enzyme complexes bound to pre-mRNA or regulating factors are missing.

Previous studies have revealed that CPSF3 is the pre-mRNA endonuclease, which is positioned by various RNA elements. However, while many sequence motifs have been characterized, it remains unclear which set of factors is needed to license CPSF3 cleavage and how this process is mediated on a molecular or atomic level. Reconstituting pre-mRNA cleavage *in vitro* and understanding CPSF3 activity on a biochemical and structural level will shed light on a key event in a crucial pre-mRNA processing pathway.

Nuclear pre-mRNA polyadenylation has been reconstituted and extensively studied, yet many relevant features are poorly understood. Characterization of regulated polyadenylation termination on a molecular and structural level will improve our understanding how the poly(A) tail is formed and whether it adopts a defined three-dimensional RNP architecture recognized by downstream factors.

Answering the outlined questions will enhance our understanding of mammalian pre-mRNA 3' end formation and point into new directions of investigation: no structural data of enzyme-substrate complexes are available and the question of their regulation remains an area of active research.

2. Methods

2.1 Protein expression

HEK293T stable expression cell lines were established using the *piggyBac* transposon system by initially transfecting the cells using polyethyleneimine (Li et al., 2013; Yusa et al., 2011). Pools of cells were generated that stably express either mCF, CPSF, PM, CstF, CF-I, CF-II, RBBP6, or PABPN1. Construct details and purification tags are listed in **Table 2**.

Additionally, proteins were transiently expressed using polyethyleneimine transfection.

For protein expression, cultures were adjusted to a density of $1x10^6$ cells per mL in FreeStyle 293 expression medium (Gibco, Thermo Fisher). The cells, either 5 mL for small scale expression or 400 mL for large scale expression, were induced with final 1 µg/mL doxycycline and harvested 48 h post induction.

Table 2: HEK293T stable expression cell lines

Stable cell lines	Protein factors
mCF	TwinStrep-3C-*Hs*Symplekin, CPSF2, CPSF3, CstF2
CPSF	TwinStrep-3C-*Hs*Symplekin, CPSF2, CPSF3, CstF2, CPSF1, WDR33 (residues 1-413), $CPSF4_{iso2}$
PM	TwinStrep-3C-*Hs*PAPOA (wildtype or catalytic mutant D113A D115A), FIP1 (full-length or residues 1-393), CPSF1, WDR33 (full-length, residues 1-572 or 1-413), $CPSF4_{iso2}$
CstF	TwinStrep-3C-*Hs*CstF1, CstF2, CstF3
CF-I	TwinStrep-3C-*Hs*CPSF5, CPSF6, CPSF7
CF-II	TwinStrep-3C-*Hs*Clp1, Pcf11 (full-length or residues 753-1555)
RBBP6	TwinStrep-3C-*Hs*RBBP6 (residues 1-335)
PABPN1	TwinStrep-3C-*Hs*PABPN1 (wildtype, A13 or A18 mutants)

2.2 Protein purification

For all purifications, lysis buffers were supplemented with DNase I, benzonase and EDTA-free complete Protease Inhibitor Cocktail (Roche). Cells were lysed with a glass

dounce homogenizer and cleared by centrifugation (30 min at 25,000 rpm). All purification steps were performed at 4 ˚C and monitored using SDS-PAGE followed by Coomassie staining.

Mammalian cleavage factor (mCF)

The four-subunit mCF was lysed in 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$ and purified using a StrepTrap HP column (Cytiva). After washing with lysis buffer and a high salt buffer containing 300 mM NaCl, bound proteins were eluted with 5 mM desthiobiotin in lysis buffer, concentrated and further purified over a Superose 6i 10/300 gel filtration column (Cytiva). Fractions containing pure protein were pooled, concentrated and flash frozen in liquid nitrogen.

Cleavage and polyadenylation specificity factor (CPSF)

The seven-subunit CPSF was purified in 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$, 2 mM $MgCl_2$ using a StrepTrap HP column (Cytiva). After extensive washing, bound proteins were eluted with 5 mM desthiobiotin in wash buffer. Most protein preparations were concentrated and flash frozen in liquid nitrogen. Protein preparations used for pull-down experiments were applied to a HiTrap Heparin column (Cytiva) for further purification. After washing, bound proteins were eluted using a gradient increasing the salt concentration to 1000 mM NaCl. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$, 2 mM $MgCl_2$, concentrated and flash frozen in liquid nitrogen.

Polymerase module (PM)

The five-subunit PM was lysed in 1x DPBS supplemented with 1 mM TCEP, 10 µM $ZnCl_2$, 2 mM $MgCl_2$ and purified using a StrepTrap HP column (Cytiva). After washing with lysis buffer and a high salt buffer containing 300 mM NaCl, bound proteins were eluted with 5 mM desthiobiotin in lysis buffer, concentrated and further purified over a Superose 6i 10/300 gel filtration column (Cytiva). Fractions containing pure protein were pooled, concentrated and flash frozen in liquid nitrogen.

Cleavage stimulation factor (CstF)

The three-subunit CstF was purified in 1x DPBS supplemented with 2 mM DTT using a StrepTrap HP column (Cytiva). After extensive washing, bound proteins were eluted with 5 mM desthiobiotin in wash buffer and applied to a HiTrap Heparin column (Cytiva) for further purification. After washing, bound proteins were eluted using a gradient increasing the salt concentration to 1000 mM NaCl. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, concentrated and flash frozen in liquid nitrogen.

Cleavage factor I (CF-I)

The three-subunit CF-I was lysed in 1x DPBS supplemented with 2 mM DTT and purified using a StrepTrap HP column (Cytiva). After washing with lysis buffer and a high salt buffer containing 300 mM NaCl, bound proteins were eluted with 5 mM desthiobiotin in lysis buffer and further purified over a HiTrap Heparin column (Cytiva). After washing, bound proteins were eluted using a gradient increasing the salt concentration to 1000 mM NaCl. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, concentrated and flash frozen in liquid nitrogen.

Cleavage factor II (CF-II)

The two-subunit CF-II was purified in 50 mM Tris-HCl pH 8.6, 200 mM KCl, 0.2 mM EDTA, 10 % glycerol, 2 mM DTT using a StrepTrap HP column (Cytiva). After extensive washing, bound proteins were eluted with 5 mM desthiobiotin in wash buffer. The overall salt concentration was reduced to < 100 mM KCl by dilution with 50 mM Tris-HCl pH 8.6, 0.1 mM EDTA, 10 % glycerol, 2 mM DTT and the proteins were applied to a MonoQ 5/50 GL column (Cytiva). After washing, bound proteins were eluted using a gradient from 100 mM to 1000 mM KCl in 50 mM Tris-HCl pH 8.6, 0.1 mM EDTA, 10 % glycerol, 2 mM DTT. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$, 2 mM $MgCl_2$, concentrated and flash frozen in liquid nitrogen.

Retinoblastoma-binding protein 6 (RBBP6)

RBBP6 variants with and without C-terminal FLAG tag were purified in 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$ using a StrepTrap HP column (Cytiva). After extensive washing, bound proteins were eluted using 5 mM desthiobiotin in wash buffer. The N-terminal TwinStrep tag was removed by addition of His-3C protease (in-house) and the target protein was further purified over a MonoQ 5/50 GL column (Cytiva). After washing, bound protein was eluted using a gradient increasing the salt concentration to 1000 mM NaCl. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$, 2 mM $MgCl_2$, concentrated and flash frozen in liquid nitrogen.

Poly(A) polymerase α (PAPOA)

PAPOA (TwinStrep-3C-*Hs*PAPOA) was purified in 1x DPBS supplemented with 2 mM DTT, 2 mM $MgCl_2$ using a StrepTrap HP column (Cytiva). After extensive washing, bound proteins were eluted with 5 mM desthiobiotin. Fractions containing pure protein were pooled, concentrated and flash frozen in liquid nitrogen.

Ssu72

Ssu72 (TwinStrep-3C-*Hs*Ssu72, wildtype or catalytic mutant C12S) was purified in 1x DPBS supplemented with 2 mM DTT using a StrepTrap HP column (Cytiva). After extensive washing, bound proteins were eluted with 5 mM desthiobiotin in wash buffer, concentrated and further purified over a Superdex 75 10/300 gel filtration column (Cytiva). Fractions containing pure protein were pooled, concentrated and flash frozen in liquid nitrogen.

HnRNPU

HnRNPU (TwinStrep-3C-*Hs*hnRNPU) was lysed in 1x DPBS supplemented with 2 mM DTT and purified using a StrepTrap HP column (Cytiva). After washing with lysis buffer and a high salt buffer containing 1000 mM NaCl, bound proteins were eluted with 5 mM desthiobiotin in lysis buffer and further purified over a HiTrap Heparin column (Cytiva). After washing, bound proteins were eluted using a gradient increasing the

salt concentration to 1000 mM NaCl. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, concentrated and flash frozen in liquid nitrogen.

Poly(A)-binding protein nuclear 1 (PABPN1)

PABPN1 was lysed in 1x DPBS supplemented with 2 mM DTT and purified using a StrepTactin XT column (IBA). After washing with lysis buffer and a high salt buffer containing 2000 mM NaCl, bound RNA was removed by partial on-column urea denaturing. Bound proteins were partially unfolded using a gradient from lysis to urea buffer (1x DPBS supplemented with 2 mM DTT, 1000 mM NaCl, 1 M urea) over four column volumes (CV; flow rate of 2 mL/min), followed by washing with urea buffer (6 CV; 2 mL/min) and on-column refolding with a gradient to lysis buffer over 8 CV (0.5 mL/min). Then, proteins were eluted with 1x BXT buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 50 mM biotin; IBA) diluted in 1x DPBS and 2 mM DTT. The N-terminal TwinStrep tag was removed by addition of His-3C protease (in-house) and the target protein was further purified over a Blue Sepharose column (Cytiva). After washing, bound proteins were eluted using a gradient to wash buffer containing 1 M Arginine at pH 7.4. Fractions containing pure protein were dialyzed overnight into 1x DPBS supplemented with 2 mM DTT, concentrated and flash frozen in liquid nitrogen.

ZC3H14

ZC3H14 (TwinStrep-3C-*Hs*ZC3H14) was purified essentially as described for PABPN1, with the sole difference that all buffers were additionally supplemented with 10 µM $ZnCl_2$.

Poly(A)-binding protein cytoplasmic 1 (PABPC1)

PABPC1 (His-GST-TEV-*Hs*PABPC1) was lysed in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM $MgCl_2$, 250 mM urea, 5 mM ß-Mercaptoethanol (ß-ME) and purified using GST agarose beads (Cytiva). After extensive washing with lysis buffer and wash buffer containing 20 mM Tris-HCl pH 8.0, 2000 mM NaCl, 5 mM ß-ME, bound proteins were eluted by overnight digestion with TEV protease (in-house) in 20 mM Tris-HCl pH 8.0,

500 mM NaCl, 0.1 mM $MgCl_2$, 5 mM ß-ME. After removal of His-GST and non-cleaved proteins using GST beads, the elution was further purified via Blue Sepharose beads (Cytiva). After washing with two high salt buffers (20 mM Tris-HCl pH 8.0, 2000 mM NaCl, 5 mM ß-ME followed by 50 mM Tris-HCl pH 8.0, 4000 mM NaCl, 150 mM KCl, 5 mM ß-ME), bound proteins were eluted with elution buffer containing 50 mM Tris-HCl pH 8.0, 150 mM KCl, 1000 mM Arginine, 5 mM ß-ME. The elution fractions were concentrated and further purified over a Superdex 200i 10/300 gel filtration column (Cytiva) in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM $MgCl_2$, 4 mM DTT. Fractions containing pure protein were pooled, concentrated and flash frozen in liquid nitrogen. PABPC1 was purified by Steffen Schüssler.

2.3 Analytical size exclusion chromatography (SEC) assays

For SEC assays, Superdex 200i 3.2/300 or Superose 6i 3.2/300 columns (Cytiva) were pre-equilibrated in SEC running buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 2 mM DTT) at 4 ˚C. Equimolar amounts of proteins or RNA (100 pmol; final concentration at 4 µM) were mixed in SEC buffer and incubated 30 min at 4 ˚C prior to injection. The absorption at $UV_{280\ nm}$ and $UV_{260\ nm}$ was recorded and peak fractions were analyzed using SDS-PAGE followed by Coomassie staining.

2.4 Pull-down experiments

2.4.1 Co-precipitation experiments

For co-precipitation assays using HEK cell lysates, $5x10^6$ cells per condition were transiently transfected as described before. Flash-frozen cells were thawed and lysed in wash buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 0.1 % NP-40 substitute), and EDTA-free Complete Protease Inhibitor Cocktail (Roche) was added. Automated pull-down experiments were performed using magnetic Streptactin-coupled Dynabeads M270 (Thermo Fisher) and a KingFisher pull-down system (Thermo Fisher) operated at room temperature. After binding for 30 min, beads were washed four times and bound proteins eluted in SDS-containing sample buffer.

2.4.2 Pull-down experiments using α-FLAG beads

For FLAG pull-down assays with purified proteins, α-FLAG M2 antibody (Sigma F3165) was added to magnetic Protein G Dynabeads (Invitrogen 10004D) equilibrated in binding buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 0.01 % NP-40 substitute). After rotation at 4 ˚C for 30 min, beads were washed three times and used immediately. 25 pmol FLAG-tagged RBBP6 was mixed with an equal amount of putative binding partner and RNA if applicable (final concentration 1 µM each) and incubated for 60 min at 4 ˚C in binding buffer. Equilibrated beads were added and rotated for 60 min at 4 ˚C. The beads were washed three times with 20 volumes of binding buffer and bound proteins eluted with 0.2 mg/mL 3x FLAG peptide in binding buffer.

2.4.3 RNA pull-down experiments using Streptavidin-beads

FLAG pull-down elutions were used as input samples for RNA Streptavidin pull-down experiments. FLAG pull-downs were performed as described above, with the sole difference that 5 µL His-3C (in-house) was added to remove TwinStrep tags. Streptavidin beads were pre-equilibrated in binding buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 0.01 % NP-40 substitute) and combined with equal volumes of FLAG elution. After rotation at room temperature for 60 min, beads were washed three times with 10 volumes of binding buffer and bound proteins were eluted with 50 mM biotin diluted in binding buffer.

2.5 RNA processing assays

2.5.1 *In vitro* transcription

RNAs were generated by *in vitro* transcription using T7 RNA polymerase (in-house) with PCR-generated and gel-purified template DNA. The reaction (**Table 3**) was scaled up as needed and incubated four hours at 37 ˚C. Then, 1 µL DNase I per reaction was added and the reaction was incubated for further 30 min. The RNA was purified using phenol/chloroform/isoamyl alcohol (Invitrogen) extraction followed by Ethanol precipitation. After resuspension in RNase-free H_2O, the RNA was 5' dephosphorylated using Quick CIP (NEB) for 60 min at 37 ˚C in CutSmart buffer

(NEB). The reaction was stopped by phenol/chloroform/isoamyl alcohol extraction and Ethanol precipitation. The RNA was resuspended in 70 % OFE (RNA-loading buffer containing Orange G) and further purified over a 10 % (w/v) and 7 M urea denaturing PAGE. After visualization by $UV_{254\ nm}$ shadowing, the bands of interest were cut with a sterile scalpel and the RNA was eluted from the gel by incubation overnight on a rotating wheel in 60 mM NaOAc pH 5.2, 300 mM NaCl, 0.2 % SDS. The RNA was finally purified using phenol/chloroform/isoamyl alcohol extraction followed by Ethanol precipitation, resuspended in RNase-free H_2O and stored at -20 ˚C.

Table 3: 1x transcription mix

	x1
NTPs (25 mM each)	1 µL
10x transcription buffer	1 µL
T7 RNA polymerase	1 µL
DNA template	1 pmol
H_2O	
Per 1x reaction	10 µL

The composition of the 10x transcription buffer is described in **Table 4**.

Table 4: 10x transcription buffer

	final
Tris-HCl pH 8.6 (4 ˚C)	400 mM
$MgCl_2$	280 mM
Triton X-100	0.1 %
DTT	50 mM
Spermidine	10 mM
H_2O	

2.5.2 Radioactive RNA 5' labeling using T4 polynucleotide kinase

In vitro transcribed and dephosphorylated RNAs were radioactively 5' labeled using T4 polynucleotide kinase (T4 PK; NEB) and γ^{32}P-ATP (10 µCi/µL; Perkin-Elmer). 10 pmol RNA were mixed with 1 µL 10x T4 PK buffer, 0.5 µL γ^{32}P-ATP and 1 µL T4 PK in a total of 10 µL and incubated 30 min at 37 ˚C. After phenol/chloroform/isoamyl alcohol (Invitrogen) extraction followed by Ethanol precipitation, the RNA was

resuspended in RNase-free H_2O and stored at -20 ˚C. Loss-less purification was assumed and successful labeling was verified by separation on a 7 % (w/v) and 7 M urea denaturing PAGE followed by phosphorimaging using a Typhoon RGB (Cytiva).

2.5.3 Pre-mRNA cleavage assays

Pre-mRNA cleavage assays using fluorescently labeled minimal substrates

Synthetic minimal model substrates contained an AAUAAA PAS sequence and were 5' fluorescently labeled (IDT). 2 pmol RNA was mixed with either HeLa nuclear extract (Ipracell) or equimolar amounts of purified proteins in a final buffer of 20 mM HEPES-KOH pH 7.9, 80 mM KCl, 20 mM creatine phosphate, 1 mM $MgCl_2$, 3.5 % (w/v) PEG-6000, 2 mM DTT. After incubation for 75 min at 30 ˚C in presence of 3'-dATP, reactions were stopped by addition of stopping buffer (100 mM EDTA, 0.2 % SDS) and proteins were removed by proteinase K (NEB) digestion for 10 min at 37 ˚C. 5x RNA loading buffer (95 % formamide, 0.05 % SDS) was added before RNAs were separated on 15-20 % (w/v) and 7 M urea denaturing PAGE, and visualized by fluorescence imaging using a Typhoon RGB (Cytiva).

Pre-mRNA cleavage assays using radioactively labeled substrates

Cleavage substrates were generated by *in vitro* transcription followed by radioactive 5' labeling. 50 fmol RNA was mixed with an excess of proteins (0.5 - 5 pmol) in final buffer of 20 mM HEPES-KOH pH 7.9, 55 mM NaCl, 20 mM creatine phosphate, 1 mM $MgCl_2$, 2 mM DTT, in absence or presence of 0.8 mM ATP (analogue). After 1 h incubation at 30 ˚C, reactions were stopped by addition of stopping buffer (100 mM EDTA, 0.2 % SDS) and proteins were removed by proteinase K (NEB) digestion for 10 min at 30 ˚C. The RNA was extracted with phenol/chloroform/isoamyl alcohol (Invitrogen), precipitated with Ethanol, separated on 8 % (w/v) and 7 M urea denaturing PAGE, and visualized by phosphorimaging using a Typhoon RGB (Cytiva).

2.5.4 Polyadenylation assays

For polyadenylation assays, 2 pmol 5' fluorescently labeled RNA containing an AAUAAA PAS sequence (IDT) was mixed with equimolar amounts of PM and up to 60

pmol PABPN1 (30x excess). Reactions were assembled in a final buffer of 1x DPBS supplemented with 5 mM $MgCl_2$, 10 µM $ZnCl_2$ and time-course experiments were started by addition of 2 mM ATP (analogue) and transfer to 37 ˚C. At indicated time points, reactions were quenched with stopping buffer (100 mM EDTA, 0.2 % SDS) and proteins were removed by proteinase K (NEB) digestion for 10 min at 37 ˚C. 5x RNA loading buffer (95 % formamide, 0.05 % SDS) was added before RNAs were separated on 6-10 % (w/v) and 7 M urea denaturing PAGE, and visualized by fluorescence imaging using a Typhoon RGB (Cytiva).

2.5.5 Splint ligation

Splint ligation for RNA 5' labeling with desthiobiotin was performed essentially as described before (Moore and Query, 2000). In brief, equimolar amounts of 5' desthiobiotinylated upstream RNA (IDT), 5' phosphorylated downstream RNA (IDT) and complimentary cDNA oligo (Sigma) were mixed in T4 DNA ligase buffer (Invitrogen). RNA and DNA secondary structures were denatured by incubation at 75 ˚C for two minutes followed by five minutes at room temperature. Ligation was started by addition of final 1 mM ATP and 0.5 U/µL T4 DNA ligase (Invitrogen). After incubation for two hours at 30 ˚C, the cDNA oligo was removed with DNase I (Roche) for 30 min at 30 ˚C. The ligated RNA was purified using denaturing urea PAGE followed by phenol/chloroform/isoamyl alcohol (Invitrogen) extraction and Ethanol precipitation, as described for the *in vitro* transcription (chapter **2.5.1**).

2.5.6 Thin-layer chromatography (TLC)

For TLC experiments, pre-mRNA cleavage assays were assembled essentially as described before (chapter **2.5.3**), but using unlabeled RNA and α^{32}P-ATP (10 µCi/µL; Perkin-Elmer). After incubation for 60 min at 30 ˚C, proteins were digested with proteinase K (NEB) for 10 min at 30 ˚C. Then, 0.5 µL reaction mixture was spotted onto a PEI-cellulose TLC plate (Merck). After migration with 0.3 M potassium phosphate pH 7.6 as running buffer, TLC plates were dried and visualized by phosphorimaging using a Typhoon RGB (Cytiva).

2.5.7 RNase protection

Body-labeled RNA was generated by *in vitro* transcription in presence of α^{32}P-UTP (10 µCi/µL; Perkin-Elmer) followed by denaturing urea PAGE purification. 5 pmol RNA was mixed with equal amounts of protein in a final buffer containing 20 mM HEPES-KOH pH 7.9, 55 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 20 mM creatine phosphate, 2 mM DTT, 0.1 % NP-40. After 1 h incubation at 30 ˚C (with or without 3'-dATP), reactions were treated with 0.5 µL benzonase (in-house) for 25 min at 25 ˚C. Then, protected RNA fragments were extracted twice with phenol/chloroform/isoamyl alcohol (Invitrogen), precipitated with Ethanol, separated on 22 % (w/v) and 5 M urea denaturing PAGE, and visualized by phosphorimaging using a Typhoon RGB (Cytiva).

2.5.8 Electrophoretic mobility shift assay (EMSA)

PABPN1 was mixed with 5' fluorescently labeled RNA (15A or 35A; IDT) in a final buffer of 1x DPBS supplemented with 5 mM $MgCl_2$. After incubation for 15 min at 22 ˚C, final 7.5 % glycerol in reaction buffer was added. The mixture was separated at 4 ˚C on a 7.5 % (w/v) native PAGE followed by fluorescence imaging using a Typhoon RGB (Cytiva).

2.6 Electron microscopy

2.6.1 Negative stain electron microscopy

Glow-discharged Quantifoil Cu200 Carbon support grids were used for negative staining EM. Grids were prepared by sample incubation (60 s), followed by three times washing with H_2O or SEC buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 2 mM DTT) and three staining steps with uranyl acetate (2 or 3 %; 5 s, 10 s, 20 s). After each step, excess liquid was manually blotted using Cellulose filter papers (Whatman). Air-dried grids were transferred into a FEI Titan Halo microscope equipped with a Falcon 3 camera. Data were collected using a nominal magnification of 45,000x, corresponding to a pixel size of 2.37 Å at the specimen level. Using EPU, the sample was imaged with a total exposure of approximately 100 $e^-/Å^2$ evenly spread over 4 s and 10 frames. Relion 3.1 was used to pre-process movies and extract picked candidate particles (Zivanov et al., 2018). Then, particles were

imported into CryoSPARC v3.1 for reference-free 2D classification (Punjani et al., 2017).

2.6.2 Streptavidin grids

Streptavidin-coated grids were prepared by Dr. Christian Benda as described before (Han et al., 2016). Grids were rehydrated with H_2O directly before use. Negative staining was performed essentially as described above, but with 5 min sample incubation and four washing steps (three times with buffer, once with H_2O).

2.6.3 Single-particle cryo-electron microscopy

Structure visualization, analysis and rigid-body model docking was carried out using UCSF ChimeraX v1.2.5 and PyMOL v2.3.2 (Pettersen et al., 2021).

Four-subunit mCF complex

Prior to grid preparation, mCF was cross-linked with Bis(sulfosuccinimidyl)suberate (BS3), quenched with Tris-HCl pH 7.5 and run over a Superose 6i 3.2/300 column (Cytiva) equilibrated in 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$, 0.5 % (w/v) trehalose. The peak fractions were pooled and 0.04 % (v/v) n-octyl-ß-D-glucoside (ß-OG) was added immediately before 4 µL of sample were applied onto glow-discharged Quantifoil R2.1, Cu 300 mesh grids. Grids were blotted for 3.5 s and plunge-frozen in a liquid ethane/propane mix using a Vitrobot Mark IV (Thermo Fisher) operated at 4 ˚C and 100 % humidity.

Cryo-EM data were collected on a FEI Titan Krios microscope (Thermo Fisher) equipped with a post-column GIF (energy width 20 eV) and a Gatan K2 camera used in counting mode. The nominal magnification during data collection was 130,000x, corresponding to a pixel size of 1.060 Å at the specimen level. Using SerialEM, the sample was imaged with a total exposure of 31 $e^-/Å^2$ evenly spread over 4 s and 20 frames (Schorb et al., 2019). The target defocus ranged between -1.2 and -3.3 µm. All processing steps, including movie pre-processing, reference-free 2D classification, and 3D classification, were performed in Relion 3.0 (Zivanov et al., 2018).

Seven-subunit CPSF complex

Prior to grid preparation, CPSF was incubated for 30 min at 4 ˚C with equal amounts of RNA (final concentration at 1 µM) in a final buffer of 1x DPBS supplemented with 2 mM $MgCl_2$, 10 µM $ZnCl_2$. 0.04 % (v/v) ß-OG was added immediately before 4 µL of sample were applied onto glow-discharged Quantifoil R1.2/1.3, Cu 200 mesh grids. Grids were blotted for 3.5 s and plunge-frozen in a liquid ethane/propane mix using a Vitrobot Mark IV (Thermo Fisher) operated at 4 ˚C and 100 % humidity.

Cryo-EM data were collected on a FEI Titan Krios microscope (Thermo Fisher) equipped with a post-column GIF (energy width 20 eV) and a Gatan K3 camera used in counting mode. The nominal magnification during data collection was 81,000x, corresponding to a pixel size of 1.094 Å at the specimen level. Using SerialEM and a beam-tilt based acquisition scheme, the sample was imaged with a total exposure of 77 $e^-/Å^2$ evenly spread over 4 s and 80 frames (Schorb et al., 2019). The target defocus ranged between -0.8 and -2.6 µm.

Movies were pre-processed on-the-fly using Focus while automatically discarding images of poor quality (Biyani et al., 2017). Picked candidate particles were extracted in Relion 3.1 (Zivanov et al., 2018). After several rounds of reference-free 2D classification, particles were imported into CryoSPARC v3.1 for further processing in 3D (Punjani et al., 2017).

Eight-subunit CPSF-RBBP6 complex

Prior to grid preparation, RBBP6 was run over a Superdex 200i 3.2/300 column (Cytiva) in SEC buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$). The peak fraction was incubated for 40 min at 4 ˚C with equal amounts of CPSF (final concentration at 1.5 µM) in SEC buffer. 0.04 % (v/v) ß-OG was added immediately before 4 µL of sample were applied onto glow-discharged Quantifoil R1.2/1.3, Cu 200 mesh grids. Grids were blotted for 3.5 s and plunge-frozen in a liquid ethane/propane mix using a Vitrobot Mark IV (Thermo Fisher) operated at 4 ˚C and 100 % humidity.

Cryo-EM data were collected on a FEI Titan Krios microscope (Thermo Fisher) equipped with a post-column GIF (energy width 10 eV) and a Gatan K3 camera used in counting mode. The nominal magnification during data collection was 81,000x,

corresponding to a pixel size of 1.094 Å at the specimen level. Using SerialEM and a beam-tilt based acquisition scheme, the sample was imaged with a total exposure of 63 $e^-/Å^2$ evenly spread over 9.5 s and 63 frames (Schorb et al., 2019). The target defocus ranged between -0.7 and -2.8 µm.

Movies were pre-processed on-the-fly using Focus while automatically discarding images of poor quality (Biyani et al., 2017). Picked candidate particles were extracted in Relion 3.1 (Zivanov et al., 2018). After several rounds of reference-free 2D classification, particles were imported into CryoSPARC v3.1 for further processing in 3D (Punjani et al., 2017).

Five-subunit polymerase module

Prior to grid preparation, polymerase module (including wildtype PAPOA) was incubated for 20 min at 4 ˚C with equal amounts of RNA and 10x excess of 3'-dATP in final buffer of 1x DPBS supplemented with 1 mM TCEP, 2 mM $MgCl_2$, 10 µM $ZnCl_2$. Then, the complex was cross-linked in batch for 30 min at room temperature with a mixture of 0.5 mM BS3, 20 µM $BM(PEG)_2$, and 20 µM $PM(PEG)_3$ before quenching with final 25 mM Tris-HCl pH 7.5 and 25 mM DTT. 0.04 % (v/v) ß-OG was added immediately before 4 µL of sample were applied onto glow-discharged Quantifoil R1.2/1.3, Cu 200 mesh grids. Grids were blotted for 3.5 s and plunge-frozen in a liquid ethane/propane mix using a Vitrobot Mark IV (Thermo Fisher) operated at 4 ˚C and 100 % humidity.

Cryo-EM data were collected on a FEI Talos Arctica microscope equipped with a Falcon 3 camera. The nominal magnification during data collection was 73,000x, corresponding to a pixel size of 1.997 Å at the specimen level. Using EPU, the sample was imaged with a total exposure of 44 $e^-/Å^2$ evenly spread over 3 s and 30 frames. The target defocus ranged between -2.0 and -3.5 µm.

PABPN1-poly(A) RNPs

Prior to grid preparation, poly(A) RNA was mixed with excess of PABPN1 in final buffer of 1x DPBS supplemented with 50 mM HEPES-NaOH pH 7.5, 5 mM $MgCl_2$, 10 µM $ZnCl_2$ and incubated for 30 min at 4 ˚C. A 10.5x and 21x excess of PABPN1 was used for 90A and 200A RNA, respectively. 4 µL of sample were applied onto glow-

discharged Quantifoil R2.1, Cu 200 mesh grids. Grids were blotted for 3.5 s and plunge-frozen in a liquid ethane/propane mix using a Vitrobot Mark IV (Thermo Fisher) operated at 4 ˚C and 100 % humidity.

Cryo-EM data were collected on a FEI Talos Arctica microscope equipped with a Falcon 3 camera. The nominal magnification during data collection was 73,000x, corresponding to a pixel size of 1.997 Å at the specimen level. Using EPU, the samples were imaged with a total exposure of 60 (90A) or 45 (200A) $e^-/Å^2$ evenly spread over 4 s (90A) or 3 s (200A) and 40 frames. The target defocus ranged between -2.0 and -3.5 µm.

2.7 Western blot

After SDS-PAGE separation, proteins were transferred onto a PVDF membrane (pore size 0.2 µm; Millipore) activated with 100 % Methanol. Transfer was performed on ice for 90 min at 200 mA using cold transfer buffer (20 mM Trizma base, 150 mM glycine, 0.01 % SDS, 5 % Methanol). Then, the membrane was blocked with 5 % (w/v) non-fat dry milk in PBS-T (1x DPBS and 0.1 % Tween-20) overnight at 4 ˚C. Blocking solution was replaced by primary antibody solution, followed by incubation for 1 h at 4 ˚C. After washing four times with PBS-T, secondary antibody solution was added for 30 min at 4 ˚C. The membrane was washed three times with PBS-T, once with 1x DPBS, rinsed with H_2O and finally imaged using ECL-prime reagent (GE Healthcare) and a LAS 4000 imager (Cytiva).

All antibodies used here (**Table 5**) were diluted in 5 % (w/v) non-fat dry milk in PBS-T directly before use.

Table 5: Antibodies

	Antigen	AB species	AB type	Dilution	Source
primary	CPSF2	rabbit	polyclonal	1:533	Thermo Fisher (PA5-55023)
	CPSF3	rabbit	polyclonal	1:533	Thermo Fisher (PA5-30631)
secondary	rabbit	goat	polyclonal	1:10,000	Bio-Rad (172-1019)

2.8 Cross-linking mass spectrometry (XL-MS)

For cross-linking mass spectrometry (XL-MS), 1.0 µM CPSF and RBBP6 were mixed with an RNA substrate in a buffer containing 20 mM HEPES-KOH pH 7.9, 55 mM NaCl and 1 mM $MgCl_2$. The sample was incubated 45 min at 4 ˚C before 0.5 mM BS3 was added. After 20 min incubation at 4 ˚C, the reaction was quenched by adding ~ 40 mM Tris-HCl pH 7.5 and incubating 15 min at 4 ˚C. The sample was spun for 10 min at 18,000 g. For denaturation of the crosslinked proteins, 4 M Urea and 50 mM Tris was added to the supernatant and the samples were sonicated using a Bioruptor Plus sonication system (Diogenode) for 10x 30 s at high intensity. For reduction and alkylation of the proteins, 40 mM 2-chloroacetamide (CAA, Sigma-Aldrich) and 10 mM TCEP were added. After incubation for 20 min at 37 °C, the samples were diluted 1:2 with MS grade water (VWR). Proteins were digested overnight at 37 °C by addition of 1 µg of trypsin (Promega). Thereafter, the solution was acidified with trifluoroacetic acid (TFA; Merck) to a final concentration of 1 %, followed by desalting of the peptides using Sep-Pak C18 1cc vacuum cartridges (Waters). The elution was vacuum dried and the desalted peptides were further pre-fractionated into eight fractions using a high pH reversed-phased nano-fractionation system (Kulak et al., 2017).

Fractionated peptides were loaded onto a 30 cm analytical column (inner diameter: 75 microns; packed in-house with ReproSil-Pur C18-AQ 1.9-micron beads, Dr. Maisch GmbH) by the Thermo Easy-nLC 1000 (Thermo Fisher) with buffer A (0.1 % (v/v) Formic acid) at 400 nl/min. The analytical column was heated to 60 °C. Using the nanoelectrospray interface, eluting peptides were sprayed into the benchtop Orbitrap Q Exactive HF (Thermo Fisher) (Hosp et al., 2015). As gradient, the following steps were programmed with increasing addition of buffer B (80 % Acetonitrile, 0.1 % Formic acid): linear increase from 8 to 30 % over 60 min, followed by a linear increase to 60 % over 5 min, a linear increase to 95 % over the next 5 min, and finally maintenance at 95 % for another 5 min. The mass spectrometer was operated in data-dependent mode with survey scans from m/z 300 to 1650 Th (resolution of 60k at m/z = 200 Th), and up to 15 of the most abundant precursors were selected and fragmented using stepped Higher-energy C-trap Dissociation (HCD with a normalized collision energy of value of 19, 27, 35). The MS2 spectra were recorded with dynamic m/z range (resolution of 30k at m/z = 200 Th). AGC target for MS1 and MS2 scans were set to 3x 106 and 105, respectively, within a maximum injection time of 100 and 60 ms for

the MS1 and MS2 scans, respectively. Charge state 2 was excluded from fragmentation to enrich the fragmentation scans for cross-linked peptide precursors.

The acquired raw data were processed using Proteome Discoverer (version 2.5.0.400) with the XlinkX/PD nodes integrated (Klykov et al., 2018). To identify the cross-linked peptide pairs, a database search was performed against a FASTA containing the sequences of the proteins under investigation. DSS was set as a cross-linker. Cysteine carbamidomethylation was set as fixed modification and methionine oxidation and protein N-term acetylation were set as dynamic modifications. Trypsin/P was specified as protease and up to two missed cleavages were allowed. Furthermore, identifications were only accepted with a minimal score of 40 and a minimal delta score of 4. Otherwise, standard settings were applied. Filtering at 1 % false discovery rate (FDR) at peptide level was applied through the XlinkX Validator node with setting simple.

3. Results

3.1 Reconstitution of an inactive pre-mRNA cleavage complex

3.1.1 Reconstitution of the endonuclease complex mCF

Mammalian 3' end processing consists of two consecutive enzymatic activities, first is the endonucleolytic cleavage of the pre-mRNA by CPSF3. In order to reconstitute this activity *in vitro*, the four-subunit mammalian cleavage factor mCF (Symplekin, CPSF2, CPSF3, CstF2) was overexpressed in HEK293T suspension cells and purified to homogeneity (**Figure 15A-C**). The nuclease activity was assessed using a fluorescently labeled RNA, with HeLa nuclear extract (NXT) serving as positive control (Moore and Sharp, 1985). This minimal RNA reporter contained an AAUAAA PAS sequence and a similar RNA was sufficient for specific activity using recombinant yeast proteins (Hill et al., 2019). However, while nuclear extract yielded a defined cleavage pattern, mCF did not cleave the RNA (**Fig. 15D**, left panel). Moreover, adding mCF to nuclear extract reduced its activity, clearly indicating that mCF in isolation is an inactive endonuclease complex (**Fig. 15D**, right panel).

To structurally characterize the inactive complex, the sample was subjected to single-particle cryo-electron microscopy (cryo-EM) analysis. Despite several attempts using different conditions (including with and without RNA or chemical cross-linking), only the lobes of the homologous proteins CPSF2 and CPSF3 could be visualized (**Fig. 15E, F**). They form a trilobal, cloverleaf-like structure dimerizing with their respective C-terminal domains, while the globular, larger lobes correspond to the MßL- and ß-CASP domains (Dominski et al., 2005; Mandel et al., 2006b). However, although the 2D class averages showed structural details, high degrees of conformational flexibility drastically reduced the overall resolution of the 3D reconstruction, making it impossible to assign the different lobes to one of the proteins with certainty.

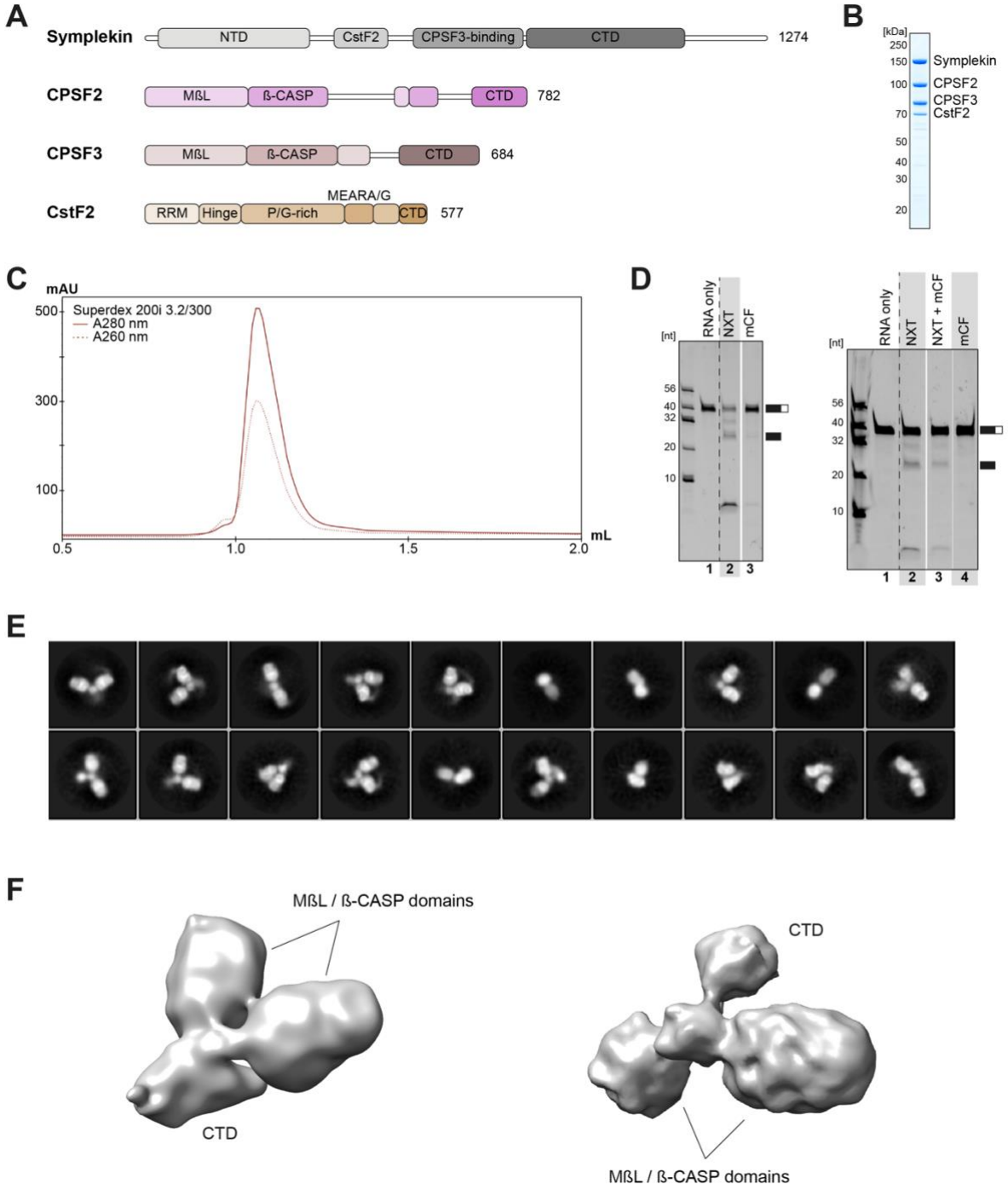

Figure 15: Characterization of mCF. (A) Domain organization of mCF subunits. **(B)** Coomassie-stained SDS-PAGE analysis of purified mCF. **(C)** Analytical SEC profile of mCF showing a monodisperse peak. **(D)** Fluorescent pre-mRNA cleavage assay showing that only nuclear extract (NXT) but not mCF cleaved the RNA (left panel). Addition of mCF into nuclear extract reduced the cleavage activity of NXT (right panel). **(E)** Representative cryo-EM 2D class averages of picked particles. **(F)** Two representative cryo-EM 3D class averages showing the conformational flexibility of the lobes of CPSF2 and CPSF3.

3.1.2 Reconstitution of the endonuclease complex CPSF

The mCF is thought to be recruited to pre-mRNA substrates by mPSF, which recognizes the AAUAAA PAS signal (Schönemann et al., 2014). Accordingly, TwinStrep-tagged mCF and the mPSF subunits CPSF1-WDR33 were separately overexpressed in HEK293T cells to test whether these factors interact directly using pull-down experiments (**Fig. 16A, B**). After co-lysis of both modules, mCF indeed efficiently pulled-down full-length CPSF1 and WDR33 (**Fig. 16B**, lane 3). As CPSF4 is crucial for mPSF RNA recognition, a seven-subunit cleavage and polyadenylation specificity factor (CPSF; mPSF + mCF) was purified subsequently (**Fig. 16A, C**). However, despite the addition of the main RNA-binding factor, recombinant CPSF could still not process the minimal pre-mRNA substrate in these conditions (**Fig. 16D**).

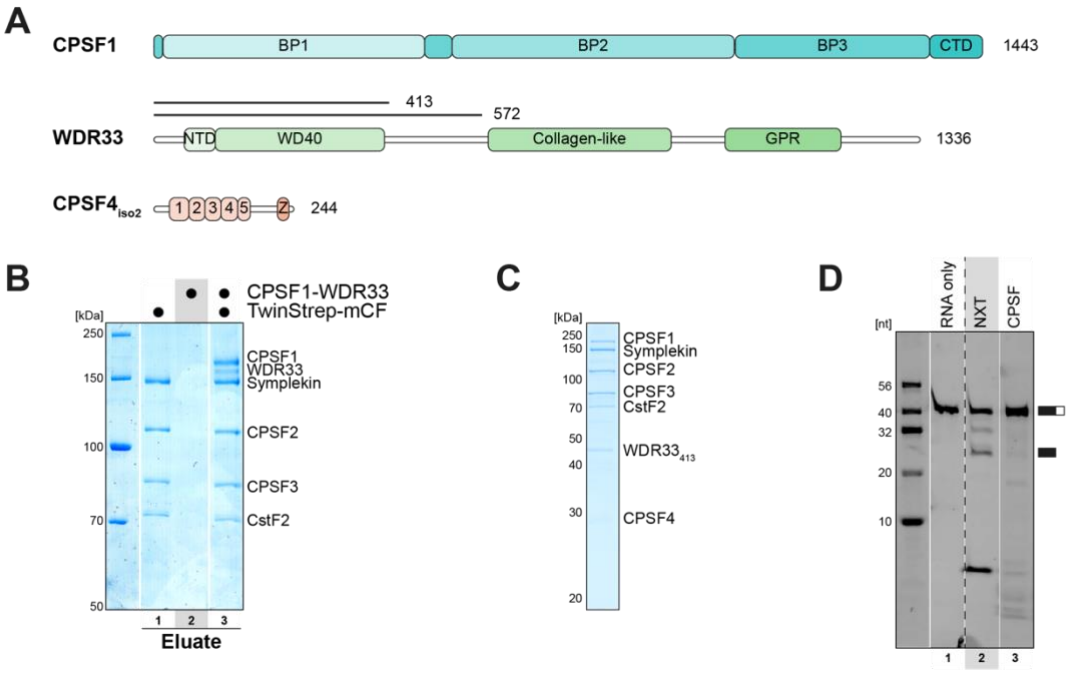

Figure 16: mCF recruits mPSF to form CPSF. (A) Domain organization of mPSF subunits. Boundaries of truncated protein constructs are indicated above as lines. **(B)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that mCF directly interacts with CPSF1-WDR33. **(C)** Coomassie-stained SDS-PAGE analysis of purified CPSF. **(D)** Fluorescent pre-mRNA cleavage assay showing that only nuclear extract (NXT) but not CPSF cleaved the RNA.

Next, CPSF was reconstituted with an unlabeled RNA substrate for cryo-EM analysis. Although the 2D class averages clearly revealed secondary structure features and allowed identification of both modules, preferred orientation and structural flexibility limited the overall resolution of the 3D reconstruction (**Fig. 17A, B**). The assembly

contained a cloverleaf structure adjacent to a more rigid shape, which is consistent with mPSF. The cloverleaf structure, which is highly similar to mCF in isolation, consists of two globular lobes (corresponding to CPSF2 and CPSF3) and an extended lobe (**Fig. 15, 17**). Only one of the globular lobes is directly contacting mPSF and therefore named proximal lobe (**Fig. 17B**). However, the limited resolution combined with the high similarity between CPSF2 and CPSF3 prevented an unambiguous assignment of the respective lobes.

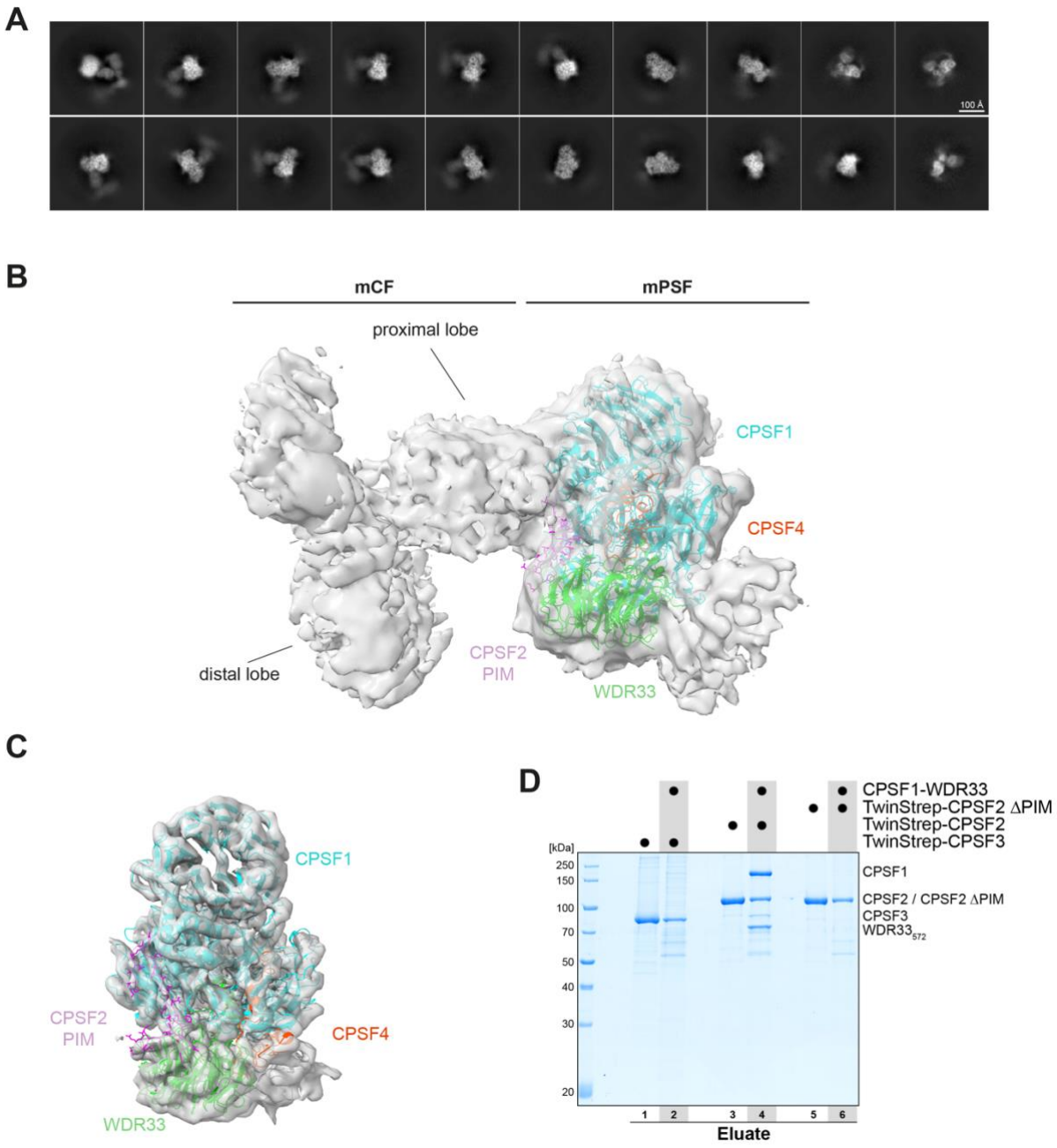

Figure 17: Structural characterization of CPSF. (A) Representative cryo-EM 2D class averages of picked particles. Scale bar ~ 100 Å. **(B)** Cryo-EM 3D reconstruction of CPSF with fitted structural model

of mPSF-CPSF2 PIM (PDB: 6URG) showing a rigid mPSF next to the flexible mCF cloverleaf structure. The proximal and distal lobes of mCF are indicated. **(C)** Cryo-EM 3D reconstruction of CPSF with fitted structural model of mPSF-CPSF2 (PDB: 6URG) showing the CPSF2 PIM peptide bound to mPSF. **(D)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that the CPSF2 PIM is necessary to recruit mPSF to mCF.

More detailed analysis revealed the presence of an additional, peptide-like density contacting the mPSF subunits CPSF1 and WDR33 (**Fig. 17C**). Shortly before obtaining the 3D reconstruction, Liang Tong's lab published highly similar structures and identified this peptide as belonging to CPSF2 (mPSF-interaction motif, PIM) (Zhang et al., 2020). Their structural model of the mPSF-CPSF2 PIM complex could be fitted with high confidence into the obtained 3D reconstruction, also explaining the peptide-like density (**Fig. 17C**). As reported by Liang Tong's lab, deletion of the PIM sequence completely abolished the interaction with CPSF1-WDR33, showing that the CPSF2 PIM is tethering mCF to mPSF (**Fig. 17D**) (Zhang et al., 2020). Interestingly, the binding surface of the proximal lobe with mPSF alone is not sufficient for a stable recruitment as neither CPSF3 nor CPSF2 ΔPIM are interacting with CPSF1-WDR33 in pull-down experiments (**Fig. 17B, D**).

3.1.3 Identification of binding partners as potential nuclease-activating factors

Reconstitution of yeast pre-mRNA cleavage using recombinant proteins revealed that the endonuclease Ysh1 is incorporated into an eight-subunit complex with low but specific enzymatic activity (Hill et al., 2019). Consequently, many different conditions, including buffers, RNA substrates or additives, were screened to assess whether the similar CPSF complex features likewise residual activity. Comparison of varying CPSF preparations revealed that particular batches possessed endonuclease activity, which was reproducible over several purification attempts. While single-step Strep column-purified CPSF cleaved the minimal RNA substrate, inclusion of an additional Heparin column resulted in almost completely abolished activity (**Fig. 18A**). This suggested that an additional factor was co-purified, which either cleaved the RNA unspecifically or activated the CPSF endonuclease activity. Triplicates of both CPSF purifications were submitted for mass spectrometry analysis to identify potential binding partners. Label-free quantitation (LFQ) mass spectrometry uncovered a group of proteins, which were enriched in the nuclease-active samples (**Fig. 18A-C**).

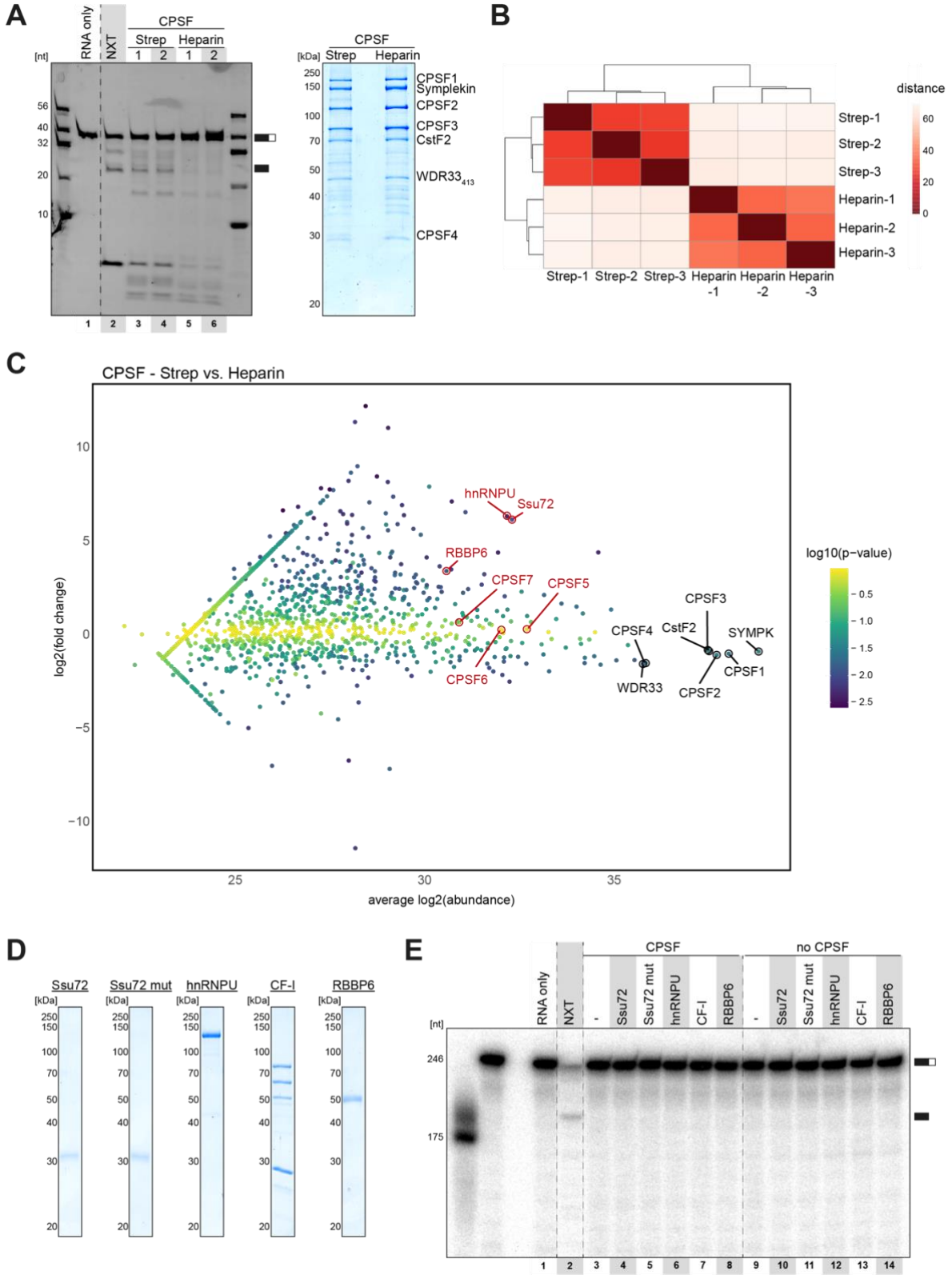

Figure 18: CPSF activity is not activated by a single factor. (A) Fluorescent pre-mRNA cleavage assay showing that NXT and Strep-purified CPSF cleaved the RNA, but not Heparin-purified CPSF (left panel). Coomassie-stained SDS-PAGE analysis of purified CPSF (right panel). **(B)** LFQ mass spectrometry heat map of CPSF samples. **(C)** LFQ mass spectrometry enrichment plot of CPSF samples. The CPSF subunits are labeled in black while potential binding partners are labeled in red.

(D) Coomassie-stained SDS-PAGE analysis of purified factors. **(E)** Radioactive pre-mRNA cleavage assay showing that potential binding partners do not stimulate CPSF endonuclease activity.

While most proteins were clearly contaminants caused by the minimal purification protocol (e.g., histones, translation factors or mitochondrially-localized proteins), some factors were previously implicated in diverse aspects of RNA processing. As they potentially enabled CPSF3 endonuclease activity, a subset of these factors was expressed and purified for pre-mRNA cleavage assays (**Fig. 18D**). To avoid potential artefacts of a minimal reporter RNA which might lack sequence elements required for cleavage *in vitro*, two radioactively labeled viral 3' UTR substrates were used for subsequent experiments. Both pre-mRNAs, adenovirus-derived L3 or simian virus 40 (SV40) RNAs, contained different canonical motifs of a single poly(A) site, thus allowing endonuclease activity at one defined sequence position (Moore and Sharp, 1985; Sheets et al., 1987). However, none of the purified factors showed any detectable nuclease activity, neither in presence or in absence of CPSF (**Fig. 18E**).

3.1.4 Characterization of the CPA interaction network

Before pre-mRNA cleavage, specific RNA motifs have to be recognized to allow CPSF3 activity; hence, different CPA modules might directly interact with mCF. In order to test these potential interactions, CF-II and CstF complexes, two modules previously implicated in recognition of the downstream RNA elements, were expressed and purified (**Fig. 19A, B**) (Chan et al., 2011; Yang and Doublié, 2011).

First, mCF and CstF complexes were mixed in absence of an RNA substrate and analyzed using analytical size exclusion chromatography (SEC). The mixture eluted essentially with the same retention volume as both subcomplexes individually, indicating that they are not interacting under these conditions (**Fig. 19C**). This was consistent with earlier reports that CstF2, a subunit present in both modules, can interact with either Symplekin in mCF or CstF3 in a mutually exclusive manner (Ruepp et al., 2011b).

In contrast, mixing of mCF with CF-II resulted in a pronounced and reproducible peak shift compared to the migration pattern of both individual modules (**Fig. 19D**). Analysis of the peak fraction with SDS-PAGE revealed the formation of a stochiometric mCF-

CF-II complex (**Fig. 19D**, right panel ***1***). To further characterize this novel interaction, the peak fraction was used to prepare negative stain EM grids. The resulting 2D class

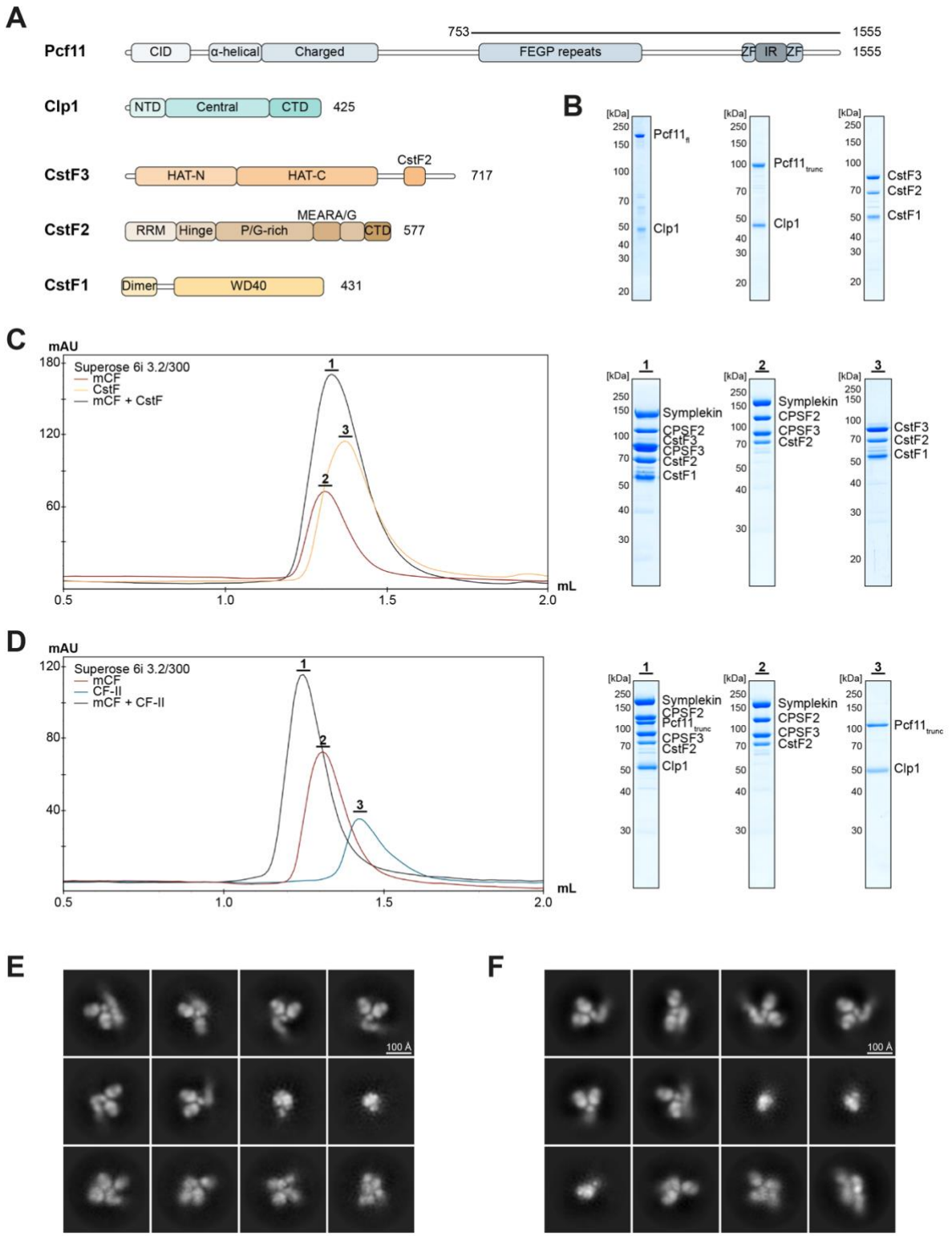

Figure 19: mCF interacts with CF-II. (A) Domain organization of CF-II and CstF subunits. Boundaries of truncated protein constructs are indicated above as lines. **(B)** Coomassie-stained SDS-PAGE analysis of purified CF-II and CstF. **(C)** Analytical SEC assay showing that mCF and CstF do not directly

interact. Overlayed SEC $UV_{280\ nm}$ traces (left panel) and Coomassie-stained SDS-PAGE analysis of peak fractions (right panel) are shown. **(D)** Analytical SEC assay showing that mCF and CF-II directly interact. Overlayed SEC $UV_{280\ nm}$ traces (left panel) and Coomassie-stained SDS-PAGE analysis of peak fractions (right panel) are shown. **(E, F)** Representative negative stain 2D class averages of picked particles of mCF-CF-II without (E) and with (F) BS3 cross-linking. Scale bars ~ 100 Å.

averages featured three distinct groups of particles: the cloverleaf structure of mCF in isolation, a globular shape, which might be part of the CF-II alone, and a multi-lobal assembly (**Fig. 19E**). While these latter class averages might represent the mCF-CF-II complex, only a small subset of particles showed these features, indicating that it is prone to disassemble. Thus, the complex was chemically cross-linked with BS3 before injection onto the SEC column. The same peak fraction as for the non-cross-linked sample was used to prepare negative stain grids. However, BS3 treatment did not significantly increase the number of particles of the intact complex (**Fig. 19F**).

As chemical cross-linking was not successful, mPSF subunits were included to potentially stabilize the complex and limit its flexibility. CPSF and CF-II were mixed and analyzed via analytic SEC assays (**Fig. 20A**).

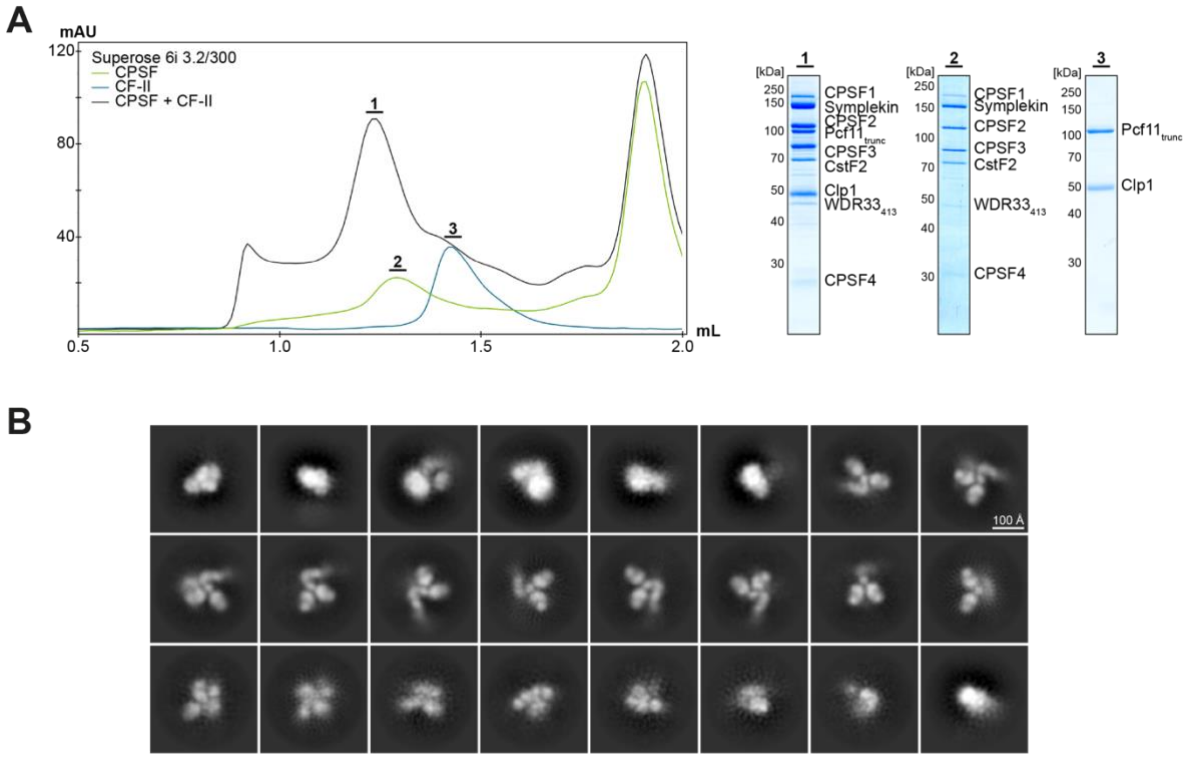

Figure 20: CPSF interacts with CF-II. (A) Analytical SEC assay showing that CPSF and CF-II directly interact. Overlayed SEC $UV_{280\ nm}$ traces (left panel) and Coomassie-stained SDS-PAGE analysis of peak fractions (right panel) are shown. **(B)** Representative negative stain 2D class averages of picked particles of CPSF-CF-II. Scale bar ~ 100 Å.

A CPSF-CF-II complex was formed, although with slightly sub-stochiometric mPSF subunits, reminiscent of the input CPSF preparation (**Fig. 20**, compare right panel ***1*** and ***2***). Again, the peak fraction was used to prepare negative stain grids. However, as revealed by the 2D class averages, this complex also seemed to be disassembling into its stable modules mPSF, mCF, CPSF, and CF-II (**Fig. 20B**).

3.2 Reconstitution of an active pre-mRNA cleavage complex

3.2.1 Defining a minimal pre-mRNA cleavage complex

Although different CPA complexes were formed, all previously tested reconstitutions lacked specific and efficient pre-mRNA cleavage activity, indicating that one or several required factors might be missing. In order to assemble an active pre-mRNA cleavage complex from purified proteins, we collaborated with the lab of Elmar Wahle at the University Halle-Wittenberg, which has extensive expertise studying 3' end formation *in vitro*. As the minimal complex necessary and sufficient for endonuclease activity was not known, different combinations of purified proteins were tested in pre-mRNA cleavage assays. These included the previously characterized factors CPSF, CstF, CF-I, CF-II, Ssu72, and RBBP6 as well as FIP1 or PAPOA in different stoichiometries and arrangements (**Fig. 21A, B**). As before, nuclear extract served as positive control, which cleaved (in presence of chain-terminating 3'-dATP) and polyadenylated (in presence of ATP) a radioactively labeled L3 pre-mRNA substrate (**Fig. 21C**). Comparable activity was observed using the purified proteins, although with differences in the polyadenylation pattern (**Fig. 21C**). Thus, specific cleavage of a pre-mRNA substrate using recombinant factors allowed for the first time to test which exact set of proteins is required for endonucleolytic cleavage *in vitro*. Omitting individual components enabled screening for a minimal complex. While mCF, PM (mPSF + PAPOA), CstF, and CF-II were necessary for cleavage activity, CF-I or Ssu72 could readily be omitted without decreased nuclease efficiency (**Fig. 21D**). Omission of RBBP6 drastically reduced the processing activity, suggesting either significantly impaired nuclease activation in its absence or co-purification of trace amounts of RBBP6 with CPSF, as seen in the LFQ mass spectrometry data (**Fig. 18, 21D**). Overall, a 14-subunit cleavage complex (CC) was necessary and sufficient for specific and efficient pre-mRNA cleavage activity *in vitro* (**Fig. 21E**).

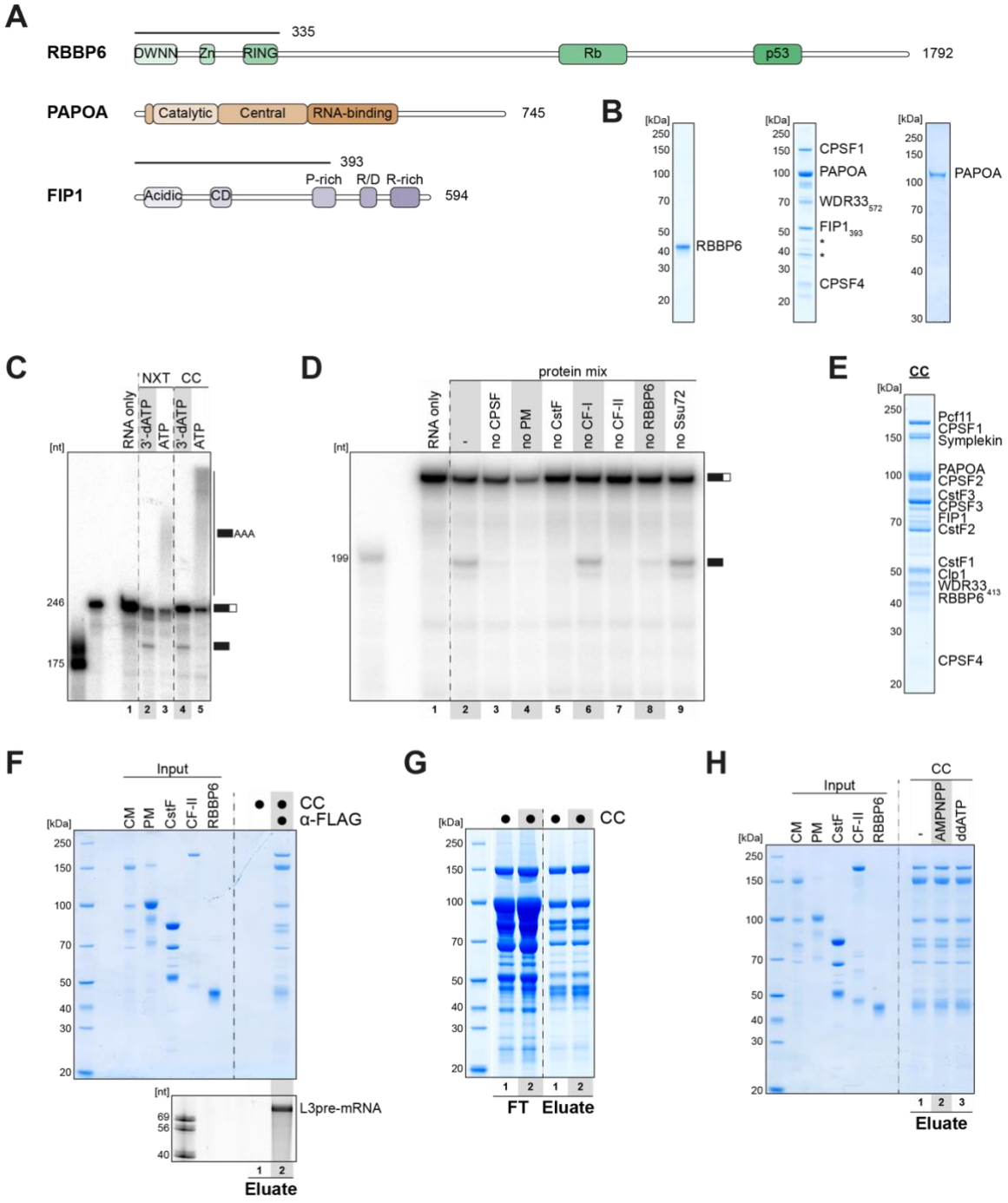

Figure 21: Reconstitution of pre-mRNA cleavage. (A) Domain organization of RBBP6, PAPOA, and FIP1. Boundaries of truncated protein constructs are indicated above as lines. **(B)** Coomassie-stained SDS-PAGE analysis of purified RBBP6, PM, and PAPOA. Asterisks mark WDR33 degradation products. **(C)** Radioactive pre-mRNA cleavage assay showing that nuclear extract and the recombinant cleavage complex (CC) cleaved the RNA. **(D)** Radioactive pre-mRNA cleavage assay omitting single factors to define a minimal active pre-mRNA cleavage complex. **(E)** Coomassie-stained SDS-PAGE analysis of a minimal active pre-mRNA cleavage complex (CC) containing 14 subunits. **(F)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that the RNA-bound cleavage complex (CC) can be purified specifically. **(G)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that only a subset of input proteins is co-purified (*Eluate*) while most remain unbound (*FT*). **(H)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that nucleotides do not influence co-purification efficiency.

Next, pull-down experiments were used to assess whether this minimal machinery would form a stable assembly. Using FLAG-tagged RBBP6 as bait, the protein components as well as a shortened L3 pre-mRNA could be specifically co-purified (**Fig. 21F**). Whether PAPOA and FIP1 actually stably and stoichiometrically interact with the other subunits is still unclear due to their overlapping SDS-PAGE migration patterns, although mass spectrometry analysis revealed that both factors are co-purified at least sub-stoichiometrically (data not shown). Moreover, comparing unbound and bound fractions of the pull-down experiment revealed that only a subset of the input proteins could be co-purified, suggesting the formation of a transient or low-affinity complex (**Fig. 21G**, compare *FT* and *Eluate*). Interestingly, complex formation under these conditions was not influenced by the presence of nucleotides (**Fig. 21H**).

3.2.2 Biochemical characterization of pre-mRNA cleavage

Previous *in vitro* experiments using nuclear extracts had generated conflicting results regarding the requirement of nucleotides or creatine phosphate as essential cofactor for pre-mRNA cleavage (Hirose and Manley, 1997; Moore and Sharp, 1985; Zhang and Cole, 1987). Using the reconstitution with purified proteins allowed addressing these and other questions in a highly defined and controlled setup. Moreover, using a PAPOA catalytic mutant or the chain-terminating ATP analogue 3'-dATP allowed an uncoupling of the pre-mRNA cleavage and polyadenylation reactions.

Addition of ATP or 3'-dATP greatly stimulated the endonuclease activity, even though it did not seem to be strictly required (**Fig. 22A**). In contrast to previous reports, creatine phosphate was not necessary for the cleavage reaction as omission essentially did not influence processing efficiencies (data not shown). Interestingly, the SV40 substrate was cleaved more efficiently than L3, rendering cleavage of L3 without any nucleotide barely detectable (**Fig. 22A**). The observed ATP stimulation raised the question whether ATP binding or hydrolysis was required for the enhancing effect, especially as the minimal pre-mRNA cleavage complex contained at least two subunits capable of ATP binding and hydrolysis. The poly(A) polymerase hydrolyzes the α-β-phosphate bond to add AMP onto the growing poly(A) tail, while Clp1 is an RNA 5' kinase, thus able to break the β-γ-phosphate link. Thin-layer chromatography (TLC)

experiments using α^{32}P-ATP revealed that ATP binding but not hydrolysis was necessary for elevated endonuclease activity (**Fig. 22B**). The CPA machinery containing wildtype PAPOA generated a radioactively labeled poly(A) tail, while simultaneously releasing AMP. In contrast, the complex containing catalytically inactive PAPOA did not hydrolyze any ATP. As both variants supported pre-mRNA cleavage, ATP binding but not hydrolysis was required for the endonuclease activity. UPF1, an RNA-dependent helicase, served as migration control as it generates ADP (**Fig. 22B**).

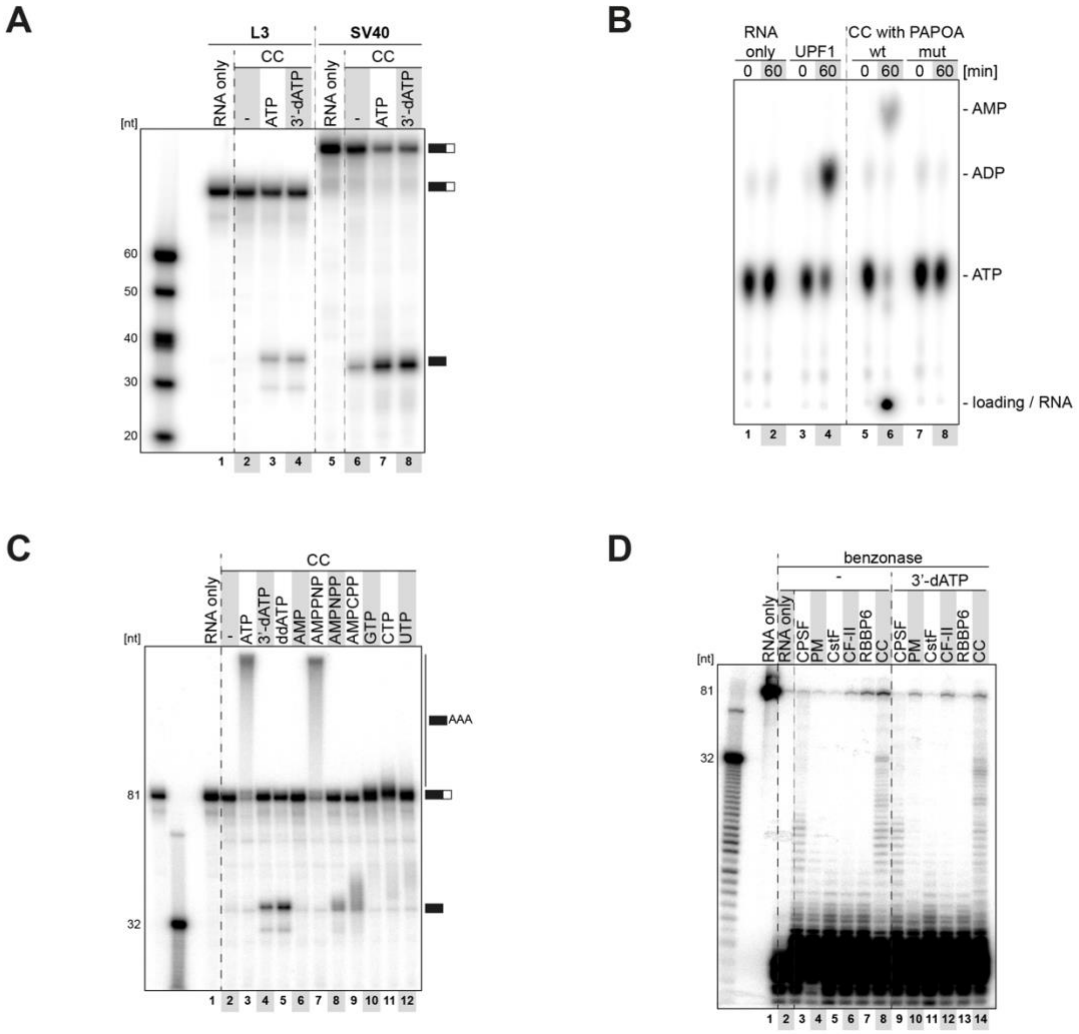

Figure 22: Biochemical characterization of pre-mRNA cleavage. (A) Radioactive pre-mRNA cleavage assay using two different substrate RNAs showing that ATP (analogues) stimulate cleavage efficiency. **(B)** Thin-layer chromatography experiment with α^{32}P-ATP showing that only a CPA machinery with wildtype PAPOA hydrolyzes ATP, either for RNA polyadenylation or release of AMP. **(C)** Radioactive pre-mRNA cleavage assay using different nucleotides showing that only adenosine nucleotides with a (modified) triphosphate group stimulated cleavage activity. **(D)** Radioactive RNase protection experiment with body-labeled RNA showing protected fragments after benzonase treatment in presence of indicated protein factors.

Next, different nucleotides were screened to test whether they might have a similar stimulatory effect as ATP (**Fig. 22C**). Using a CPA machinery with wildtype PAPOA, ATP and AMPPNP supported polyadenylation, while 3'-dATP, ddATP, AMPNPP, and AMPCPP allowed efficient cleavage, although AMPNPP and AMPCPP to a lower extend. Interestingly, neither AMP nor GTP, CTP or UTP stimulated the endonuclease reaction, indicating that both the adenosine as well as a (modified) triphosphate group were required for the enhancing effect (**Fig. 22C**).

Lastly, RNase protection experiments were used to assess whether the CPA machinery stably assembled on a pre-mRNA substrate and whether it remained bound after the cleavage reaction (**Fig. 22D**). The CPA machinery was assembled on a body-labeled L3 substrate before benzonase addition to degrade unprotected RNA. Overall, the conditions with and without nucleotide resulted in highly similar RNA migration patterns. While all modules protected smaller fragments, addition of CPSF or the complete CPA complex resulted in recovery of longer RNA. In particular, the cleavage complex protected RNA of approximately 30 nucleotides, a length roughly spanning the sequence from the PAS to the cleavage site (**Fig. 22D**, lanes 8 and 14). However, in all samples, only a small subset of RNA molecules was protected from degradation, indicating that complex formation on the substrate is not efficient under these conditions, consistent with the low yields of the pull-down experiments (**Fig. 21G, 22D**).

3.2.3 RBBP6 plays a central role in pre-mRNA cleavage

The biochemical reconstitution experiments suggested that RBBP6 might have an analogous role in mammalian pre-mRNA cleavage as Mpe1 in yeast, helping to activate the CPSF3 endonuclease in the context of a correct pre-mRNA cleavage site. Yeast Mpe1 interacts directly with the CPSF3 ortholog Ysh1 and shares significant sequence similarity with the N-terminal part of RBBP6, suggesting that RBBP6 might engage CPSF3 in a similar manner (Hill et al., 2019). In pull-down experiments using purified proteins, RBBP6 efficiently co-purified with the seven-subunit CPSF complex, even without addition of an RNA substrate (**Fig. 23A**). Interestingly, a reproducible, albeit weaker interaction with the four-subunit mCF alone was observed, suggesting the presence of a second binding site in mPSF (**Fig. 23A**).

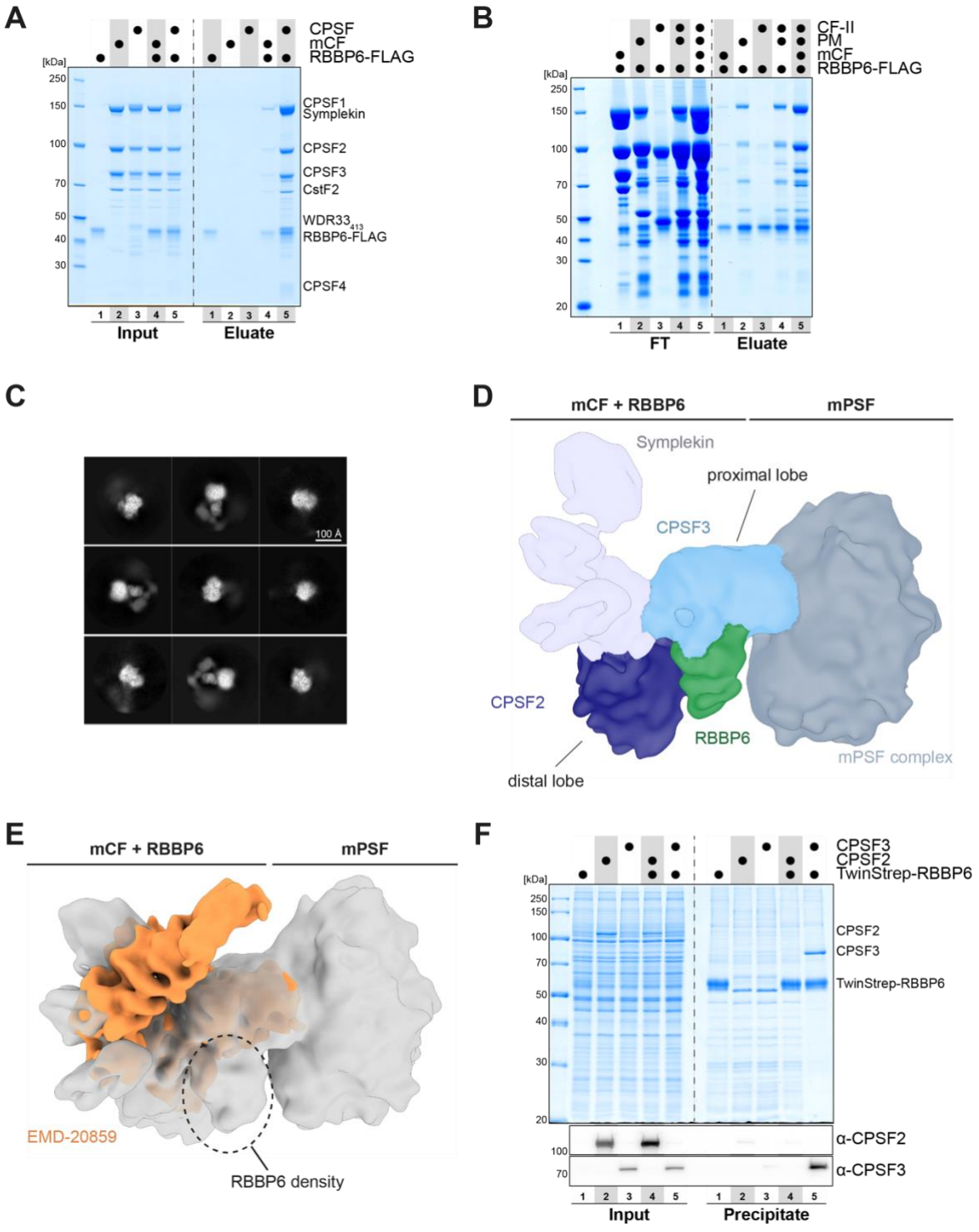

Figure 23: RBBP6 is recruited to CPSF. (A) Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that RBBP6 directly interacts with CPSF and, albeit weaker, mCF. **(B)** Coomassie-stained SDS-PAGE analysis of pull-down experiment in presence of RNA showing that RBBP6 interacts with mCF and PM. **(C)** Representative 2D class averages of picked particles. Scale bar ~ 100 Å. **(D)** Filtered and segmented cryo-EM map of the eight-subunit RBBP6-CPSF complex. The proximal and distal lobes of mCF are indicated. **(E)** Cryo-EM map of mCF alone (EMD-20859) in orange fitted into the cryo-EM map of the RBBP6-CPSF complex. Both cryo-EM maps were filtered to 10 Å before superposition. The same orientation as in (D) is shown and the extra density protruding from the proximal lobe is indicated by a black circle. **(F)** Coomassie-stained SDS-PAGE (top panel) and Western

blot (bottom panel) analysis of pull-down experiment showing that RBBP6 directly interacts with CPSF3 and not CPSF2.

To test this, pull-down experiments were performed trying to co-purify mCF, PM (mPSF + PAPOA) or CF-II with RBBP6 as bait protein. In presence of a substrate pre-mRNA, PM was co-purified even more efficiently than mCF, while CF-II did not interact with RBBP6 (**Fig. 23B**).

Based on these results, an eight-subunit RBBP6-CPSF complex was reconstituted for single-particle cryo-EM analysis. As for CPSF alone, the 2D class averages clearly revealed secondary structure features, but preferred orientation and inherent flexibility limited the overall resolution of the 3D reconstruction (**Fig. 23C, D**). The cryo-EM map, filtered to 10 Å, showed similar features as CPSF in isolation, a cloverleaf structure next to a more rigid mPSF complex (**Fig. 23D, E**) (Zhang et al., 2020). However, extra density was protruding from the proximal lobe, indicating that this density might correspond to RBBP6, the additional subunit present in the sample (**Fig. 23E**). Moreover, this also pointed to the proximal lobe as corresponding to CPSF3. To test this hypothesis, TwinStrep-tagged RBBP6 was overexpressed in HEK293T cells with either CPSF2 or CPSF3 for co-precipitation experiments. In these experiments, RBBP6 indeed co-purified reproducibly only with CPSF3 and not with CPSF2 (**Fig. 23F**).

3.2.4 RBBP6 activates the CPSF3 endonuclease activity

Next, cross-linking mass spectrometry (XL-MS) was performed to confirm the structural arrangement in the context of the complete RBBP6-bound CPSF complex. After treatment with BS3, cross-links between CPSF3 and the N-terminal DWNN domain of RBBP6 were repeatedly detected (**Fig. 24A**). Based on these biochemical data, AlphaFold2 was used to computationally predict a structural model of the CPSF3-RBBP6 complex (Jumper et al., 2021). In these calculations, the DWNN domain of RBBP6 docks onto the CPSF3 metallo-ß-lactamase domain, in close proximity to the active site opening, similar to the binding mode observed for the yeast orthologs (**Fig. 24B, C**) (Hill et al., 2019). Mapping the mass spectrometry cross-links onto the structural model revealed that they were well within the expected distance range of BS3, further supporting the AlphaFold2 prediction (**Fig. 24D**). The RBBP6-

CPSF3 model from AlphaFold2 fitted into the size and shape of the proximal lobe and features of the protruding density. The protruding density extended further and additional mass spectrometry cross-links were identified between the region of RBBP6 downstream of the DWNN domain and other subunits of the CPSF complex, specifically CPSF2, WDR33 and Symplekin (**Fig. 24A**). This suggested the presence of additional contact sites in this more flexible part, consistent with the relative strengths of the interactions observed in the pull-down assays (**Fig. 23A, B**).

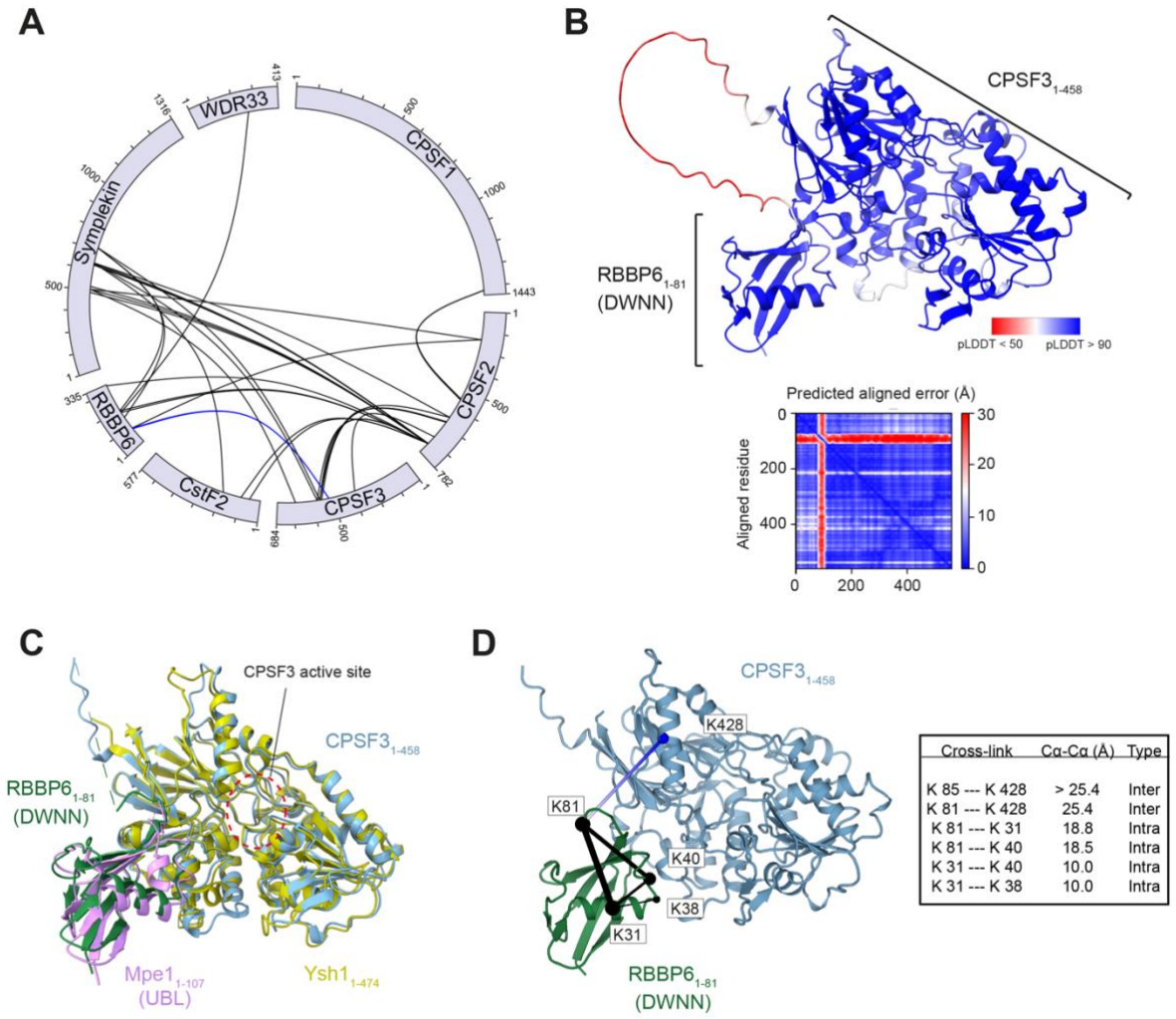

Figure 24: RBBP6 contacts the metallo-ß-lactamase domain of CPSF3. (A) Intermolecular cross-links of the RBBP6-CPSF complex. Intermolecular cross-links between CPSF3 and RBBP6 shown in (D) are highlighted in blue. Intramolecular cross-links are omitted for clarity. **(B)** Computationally predicted structural model of the CPSF3$_{1-458}$-RBBP6$_{1-81}$ complex colored by its per-residue confidence score (pLDDT) (Jumper et al., 2021). The artificial linker sequence connecting the two domains is shown in red with low confidence score (top panel). Predicted aligned error plot with blue indicating high confidence in predicting relative positions (bottom panel). **(C)** Overlay of the AlphaFold2 model of CPSF3-RBBP6 with the experimentally derived structure of yeast Ysh1-Mpe1 (PDB: 6I1D) with an RMSD of 0.884 Å. Approximate location of the CPSF3 active site is indicated by a red dashed circle. **(D)** Cross-links between RBBP6$_{1-81}$ and CPSF3$_{1-458}$ mapped onto the computationally predicted model of the complex. Inter- and intramolecular cross-links are in blue and black, respectively, with the

thickness of the line indicating their score (thicker = higher score), and the respective residues indicated as spheres. Cross-links to an unmodeled region are shown in white (placed to the closest visible Cα atom). The table lists measured distances for visualized cross-links.

To further validate the model, mutations aimed at disrupting the interaction between CPSF3 and RBBP6 were introduced. Specifically, positively charged residues in the CPSF3-binding interface of RBBP6 were changed to Aspartic acid (**Fig. 25A**).

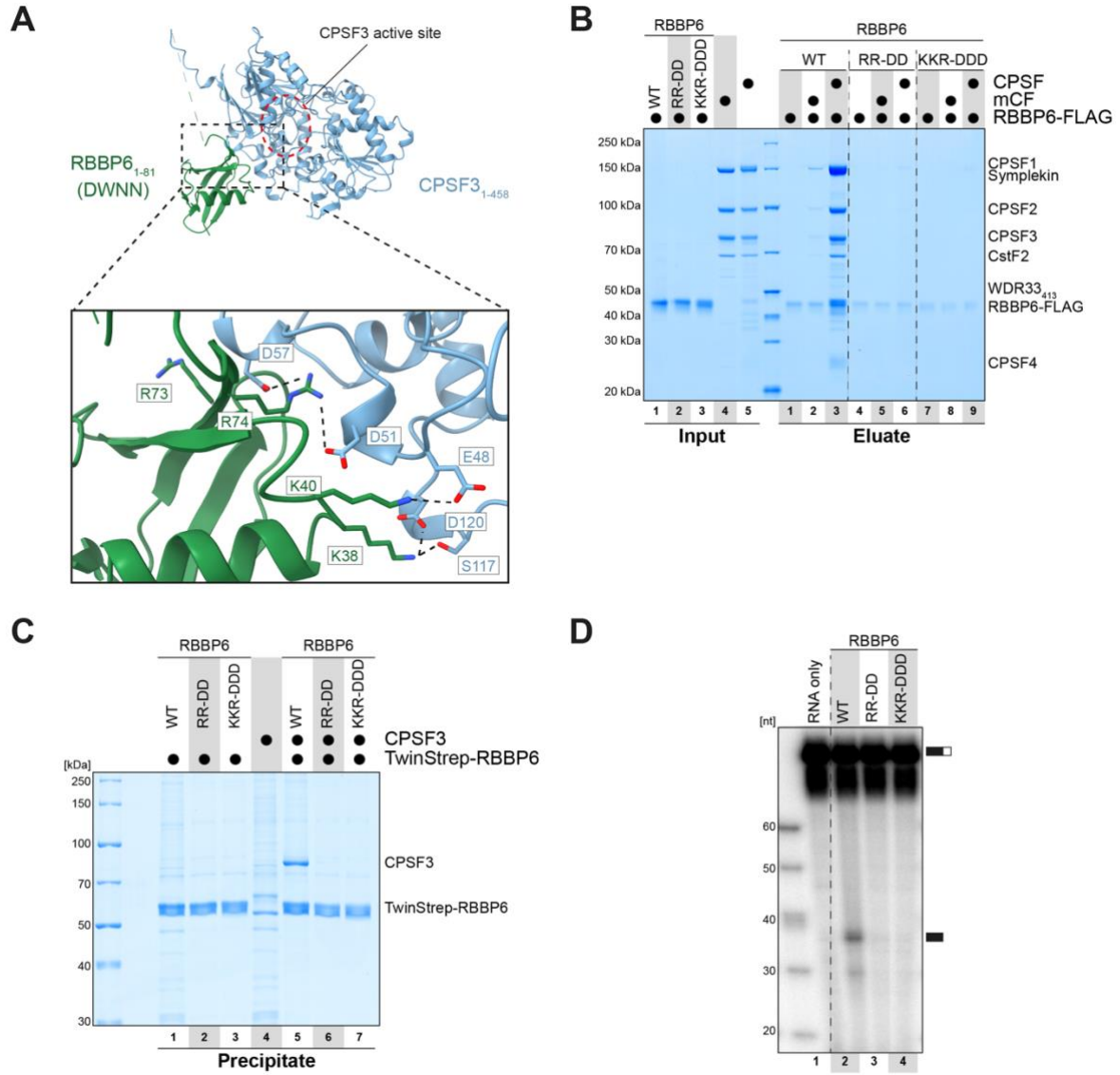

Figure 25: Characterization of the RBBP6-CPSF interaction. (A) Computationally predicted structural model of the CPSF3$_{1-458}$-RBBP6$_{1-81}$ complex with close-up of the putative binding interface. The CPSF3 active site is indicated. Possible ionic interactions are shown with black dotted lines. Labeled positively charged RBBP6 residues were changed to Aspartic acid. **(B)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that reverse-charged mutations in the predicted RBBP6-CPSF3 interface disrupt the interaction of RBBP6 with mCF and CPSF. **(C)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that the reverse-charged mutations in RBBP6 disrupt the interaction to CPSF3. **(D)** Radioactive pre-mRNA cleavage assay using either wildtype or mutated RBBP6 variants showing that only wildtype RBBP6 supported cleavage activity.

These reverse-charged mutations disrupted the interaction both in pull-down experiments using purified proteins as well as in co-precipitation experiments in HEK293T cell lysates (**Fig. 25B, C**).

Consistently, the mutated RBBP6 proteins did not support pre-mRNA cleavage activity (**Fig. 25D**). These data indicate that the identified interface between the RBBP6 DWNN domain and CPSF3 is indeed the major binding site with which RBBP6 is recruited to the CPSF complex (**Fig. 25B**) and locates the CPSF3 subunit at the proximal lobe adjacent to the mPSF core. Collectively, these data reveal that the RBBP6-CPSF3 interaction is indeed conserved from yeast to human and that *in vitro*, this binding site is essential to recruit RBBP6 to a fully assembled CPSF complex prior to pre-mRNA cleavage.

3.2.5 Purification of the CPA complex and EM trials

After biochemical characterization of the active endonuclease, electron microscopy was used to gain structural insight into the minimal 3' end processing complex. The aforementioned FLAG pull-down approach was used to assemble the CPA complex on a substrate RNA and subsequently treated with Bio-beads to remove residual detergent to avoid interference during preparation of negative stain EM grids (**Fig. 26A**). However, despite preparing grids without and with chemical cross-linking using BS3, micrographs showed mostly small particles spiked with some larger assemblies or aggregates (**Fig. 26B**). Concordantly, 2D class averages did not reveal any defined particles (**Fig. 26C**). This suggested either complex disassembly or formation of different sub-complexes during the pull-down experiment, consistent with the ability of RBBP6 to interact with several modules (**Fig. 23B**). Hence, a second purification step was introduced to enrich the fully assembled CPA machinery. Both previously used RNA substrates, L3 and SV40, were 5' labeled with desthiobiotin (DBT) by splint ligation, which allowed specific purification of desthiobiotin-RNA using Streptavidin beads (**Fig. 26D**).

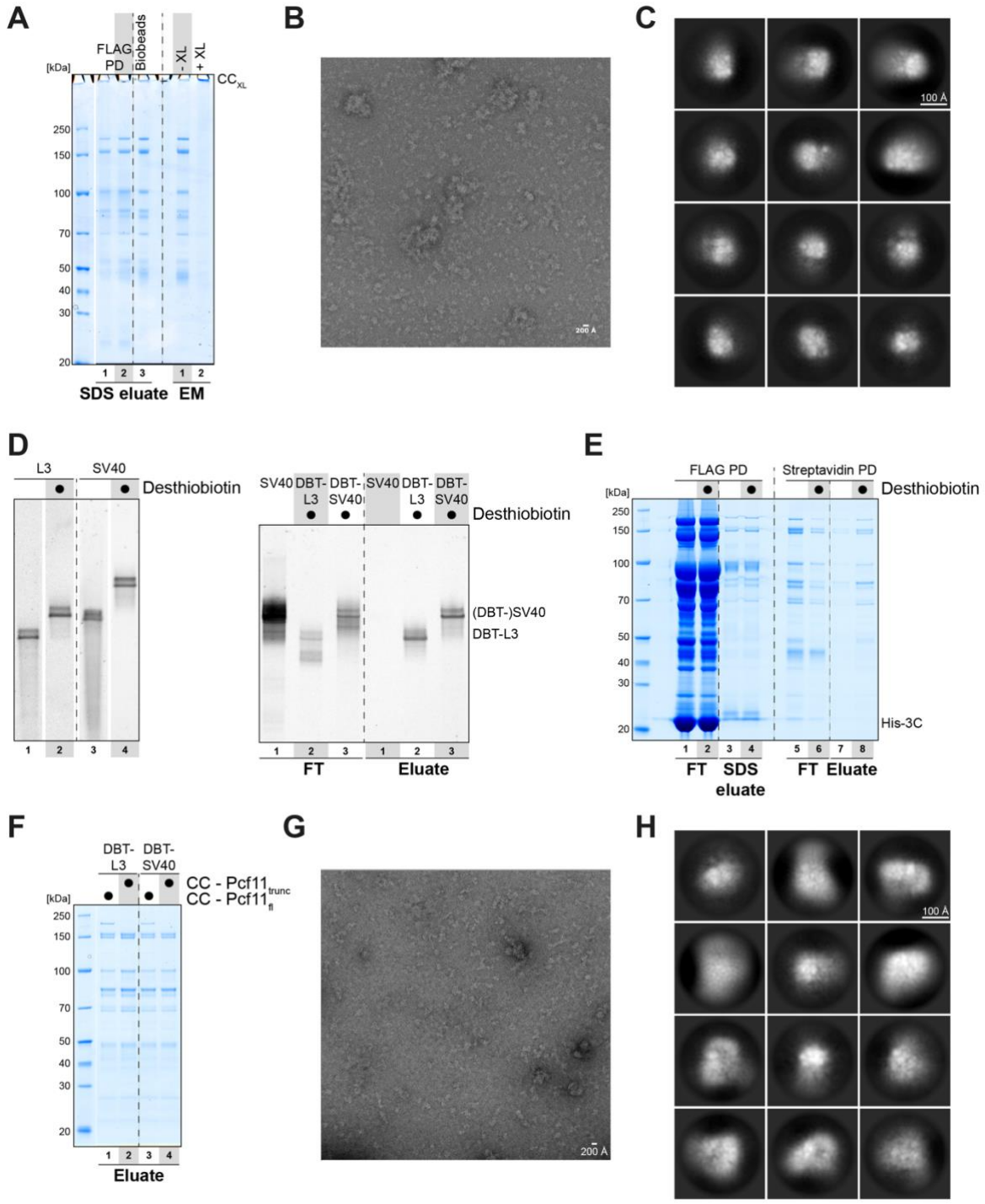

Figure 26: Purification of CPA machinery using pull-down approaches. (A) Coomassie-stained SDS-PAGE analysis of FLAG pull-down experiment performed to prepare negative stain EM grids. Non-cross-linked and cross-linked samples were used. **(B)** Representative negative stain micrograph of FLAG pull-down eluate. Scale bar ~ 200 Å. **(C)** Representative negative stain 2D class averages of picked particles of FLAG pull-down eluate. Scale bar ~ 100 Å. **(D)** Denaturing urea PAGE of RNA substrates before and after desthiobiotinylation by splint ligation (left panel). Denaturing urea PAGE of pull-down experiment showing that desthiobiotin-RNA can be purified specifically (right panel). **(E)** Coomassie-stained SDS-PAGE analysis of two-step pull-down experiment showing that proteins bound to desthiobiotin-RNA can be purified specifically. **(F)** Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that neither the used RNA substrate nor the Pcf11 construct influence co-purification efficiency. **(G)** Representative negative stain micrograph of Streptavidin pull-down eluate.

Scale bar ~ 200 Å. **(H)** Representative negative stain 2D class averages of picked particles of Streptavidin pull-down eluate. Scale bar ~ 100 Å.

The CPA complex was assembled on desthiobiotin-RNA while His-3C was added during the FLAG pull-down step to remove TwinStrep tags, which competed for binding to Streptavidin beads. After the native FLAG elution was added to these beads, proteins bound to desthiobiotinylated RNA could be purified specifically (**Fig. 26E**, lanes 7 and 8). Next, different RNA and protein constructs were compared to optimize the sample prior to EM trials (**Fig. 26F**). Surprisingly however, although SV40 RNA was cleaved more efficiently in pre-mRNA cleavage assays, Streptavidin pull-down efficiency in these conditions was entirely independent of the used substrate, suggesting cleavage rates are not only influenced by assembly rates on a pre-mRNA molecule (**Fig. 22A, 26F**). The two-step pull-down strategy was utilized to prepare negative stain grids, while a deposited Streptavidin mono-layer was used to perform the specific RNA purification step directly on the grids (Han et al., 2016). However, neither the micrographs nor the 2D class averages showed any defined particles, which could represent the CPA machinery, clearly indicating that a different sample preparation method was needed for successful EM studies (**Fig. 26G, H**).

Although the 3' end machinery could be purified with the pull-down approach, assembly using gradient centrifugation or columns, size exclusion and ion-exchange, was not successful, either due to incomplete complex formation or unspecific interaction with the column material (not shown). Hence, protein constructs were optimized by truncation of unstructured regions, which were not needed for specific interaction within the minimal CPA complex. In addition to shortening WDR33, Pcf11, and RBBP6, the disordered FIP1 C-terminus was removed since it is absent in yeast Fip1 and also in human isoform 4, which was recently shown to be active in *in vitro* pre-mRNA cleavage assays (Boreikaite et al., 2022). Surprisingly, even though an HEK293T cell expression test indicated lower expression levels of the short FIP1 construct (**Fig. 27A**), purification yields of PM with truncated FIP1 were almost 10x higher compared to PM with full-length FIP1. Interestingly, injection of stoichiometric amounts of all cleavage complex subunits without any RNA resulted now in formation of one slightly tailing main peak followed by a second, shoulder-like peak (**Fig. 27B**). Furthermore, the retention volume, SDS-PAGE analysis of the peak fractions and overlay with the SEC profiles of single CPA modules suggested the formation of a 13-

component complex without PAPOA, consistent with its transient interaction with mPSF (**Fig. 27B, C**) (Kaufmann et al., 2004).

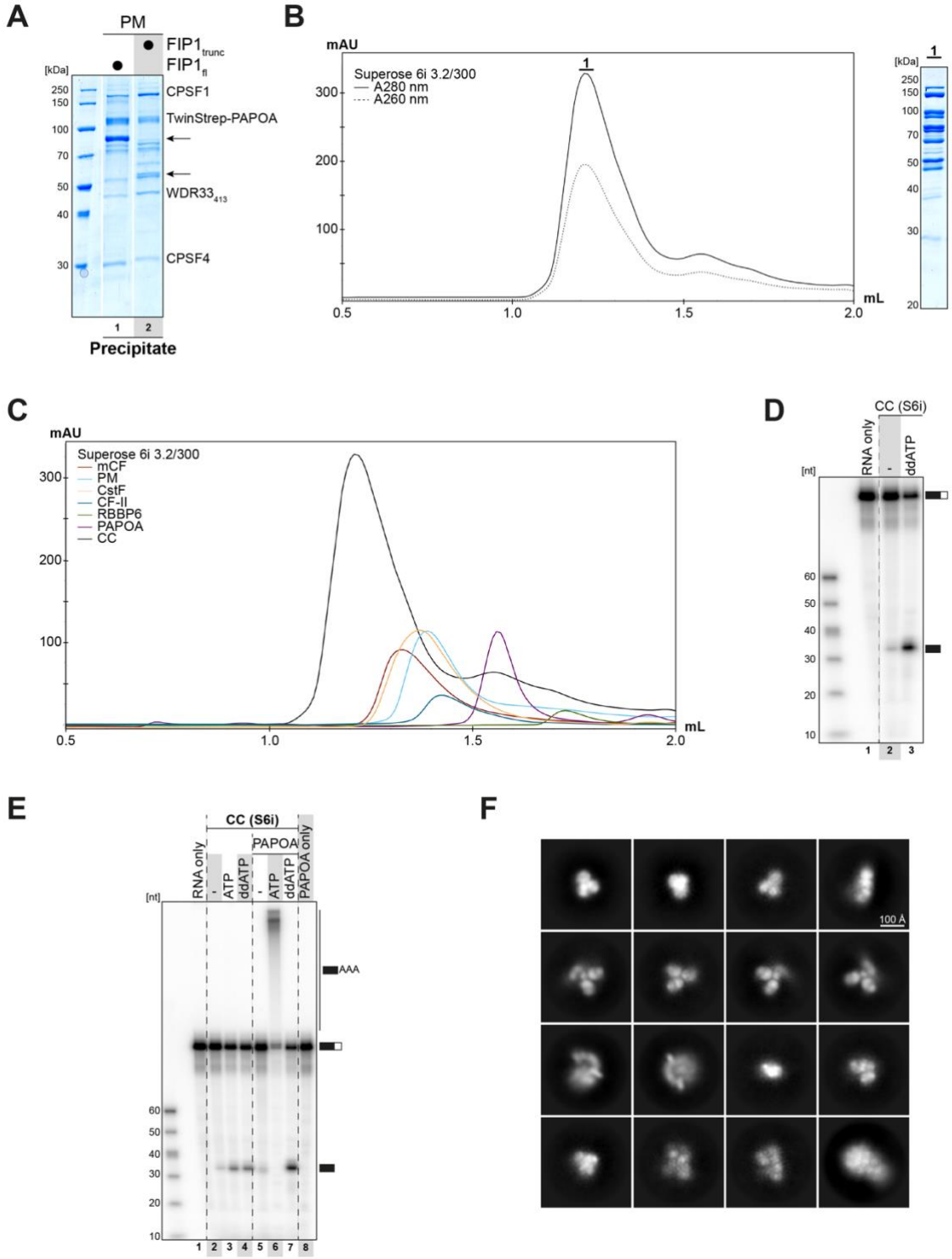

Figure 27: Purification of the minimal CPA machinery using SEC. (A) Coomassie-stained SDS-PAGE analysis of HEK293T expression test of PM with different FIP1 constructs, which are indicated with arrows. **(B)** Analytical SEC profile of CC showing a tailing peak (left panel) and Coomassie-stained SDS-PAGE analysis of the peak fraction (right panel). **(C)** Analytical SEC assay with overlayed SEC

$UV_{280\ nm}$ traces of CC and single modules. **(D)** Radioactive pre-mRNA cleavage assay showing that the SEC peak fraction possessed endonuclease activity and was stimulated by ddATP. **(E)** Radioactive pre-mRNA cleavage assay showing that the CC SEC peak fraction was only stimulated by addition of PAPOA in presence of ddATP. **(F)** Representative negative stain 2D class averages of picked particles of CC SEC peak fraction. Scale bar ~ 100 Å.

Next, the peak fractions were used in a radioactive pre-mRNA cleavage assay as control experiment whether a nuclease-active complex was formed. The RNA was cleaved in absence of any nucleotides while ddATP greatly stimulated the activity, as seen before (**Fig. 22A, 27D**). This suggested that an active endonuclease complex could be assembled using size exclusion chromatography and that PAPOA was either present in sub-stochiometric amounts or dispensable for the cleavage activity. In a follow-up experiment, the CPA machinery including a PAPOA catalytic mutant was run over the SEC column. The peak fractions were active in pre-mRNA cleavage assays, while subsequent addition of wildtype PAPOA into the reaction mix resulted in comparable nuclease activity in absence of any nucleotides (**Fig. 27E**, lanes 2 and 5). The RNA was either cleaved more efficiently or polyadenylated in presence of ATP (**Fig. 27E**, lanes 3 and 6). Again, ddATP enhanced endonuclease activity, surprisingly however, to a larger extend upon addition of PAPOA (**Fig. 27E**, lanes 4 and 7). This indicated that both ATP and PAPOA could stimulate CPSF3 endonuclease activity, although the latter only in combination with a nucleotide.

Lastly, the gel-filtrated sample was used to prepare negative stain grids to gain first structural information of a cleavage-competent machinery. However, 2D class averages did not reveal particles representing the intact complex (**Fig. 27F**). Instead, the averages were reminiscent of 2D classes of the single modules in isolation, indicating that the complex dissociated prior or during grid preparation.

3.3 Reconstitution of pre-mRNA polyadenylation

3.3.1 Regulation of poly(A) polymerase activity by poly(A)-binding proteins

The activity of poly(A) polymerase in pre-mRNA polyadenylation was extensively studied *in vitro* and *in vivo* over the last decades. Nevertheless, many important aspects are still poorly understood, including some aspects of regulation by poly(A)-binding proteins, the mechanism of poly(A) tail length control or underlying structural details.

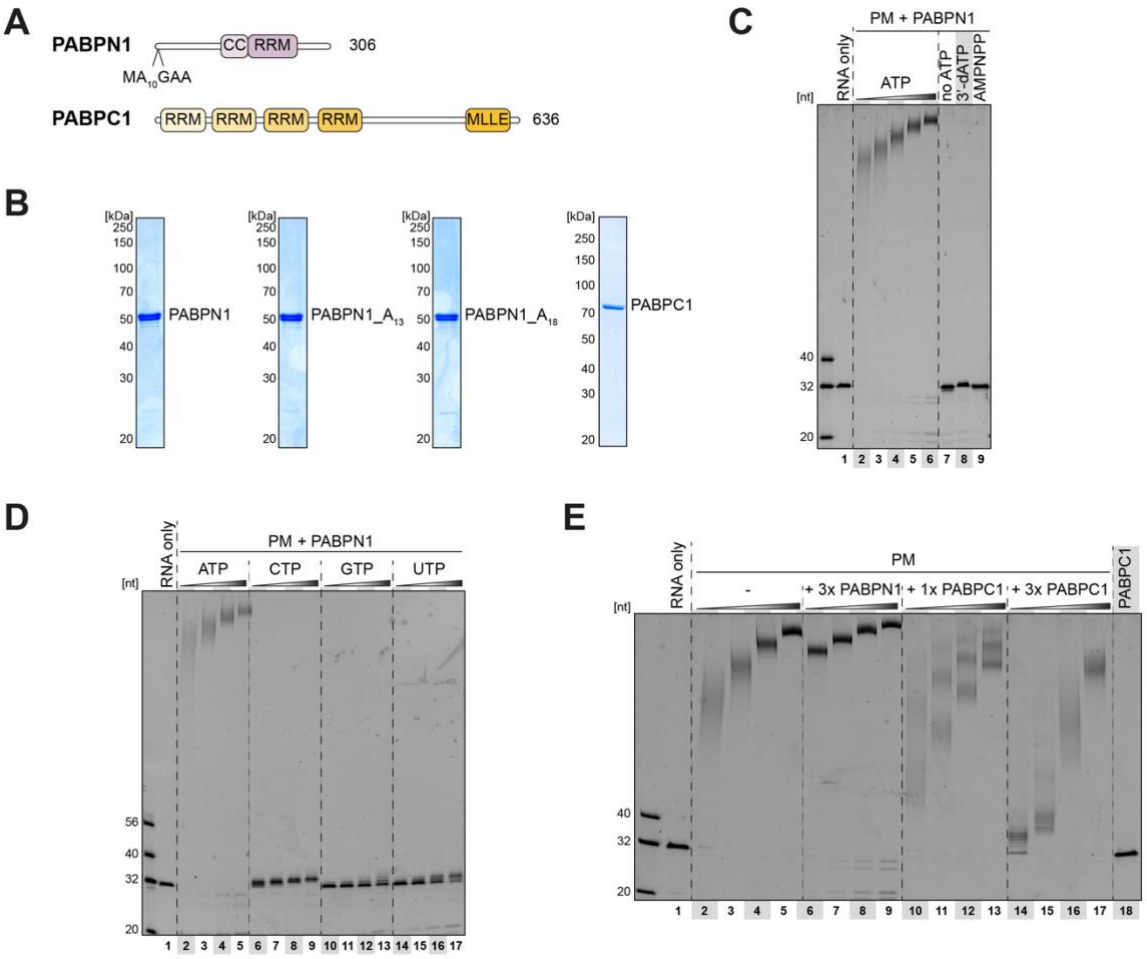

Figure 28: Regulation of PAPOA activity by poly(A)-binding proteins. (A) Domain organization of poly(A)-binding proteins. The N-terminal Alanine stretch of PABPN1 is indicated. **(B)** Coomassie-stained SDS-PAGE analysis of purified poly(A)-binding proteins. **(C)** Fluorescent polyadenylation assay in presence of PM and PABPN1 showing that only ATP supported efficient polyadenylation. **(D)** Fluorescent polyadenylation assay of PM and PABPN1 in presence of indicated nucleotides showing that all other nucleotides except ATP are barely incorporated by PAPOA. **(E)** Fluorescent polyadenylation assay in presence of PM and indicated poly(A)-binding proteins showing that PABPN1 stimulates and PABPC1 inhibits PAPOA activity.

Polyadenylation in the nucleus by PAPOA is regulated by the poly(A)-binding protein PABPN1 (**Fig. 28A, B**). As expected, the polymerase module (PM) very efficiently polyadenylated a fluorescently labeled model substrate in presence of ATP (**Fig. 28C**). Omission of ATP or replacement by non-hydrolyzable AMPNPP completely abolished the activity, while 3'-dATP allowed addition of a single AMP before acting as chain-terminating nucleotide (**Fig. 28C**). While ATP was very efficiently incorporated to elongate the pre-mRNA primer, the other RNA nucleotides CTP, GTP, and UTP cannot substitute for ATP, highlighting that the PAPOA *in vivo* specificity towards ATP can be reproduced *in vitro* (**Fig. 28D**) (Legnini et al., 2019).

While PABPN1-regulated polyadenylation was extensively characterized before, a previous report described nuclear polyadenylation dependent on the primarily cytoplasmic poly(A)-binding protein PABPC1 (Hosoda et al., 2006). Thus, the *in vitro* activity of the polymerase module alone or in complex with the poly(A)-binding proteins PABPN1 or PABPC1 was compared (**Fig. 28E**). In a time-course experiment, the polymerase module efficiently elongated the pre-mRNA primer, while addition of PABPN1 significantly stimulated the activity. In contrast, stochiometric amounts of PABPC1 reduced polyadenylation and even resulted in a different, step-wise polyadenylation pattern (**Fig. 28E**, lanes 10-13). Moreover, increased amounts of PABPC1 enhanced this inhibitory effect, suggesting that PABPC1 competed for RNA binding with the polymerase module (**Fig. 28E**).

3.3.2 Structural characterization of the polymerase module

While mPSF and the catalytic domain of PAPOA were structurally characterized in isolation, it remains unclear how PAPOA is recruited by mPSF or regulated by PABPN1 (Martin et al., 2000; Sun et al., 2018). Accordingly, single-particle cryo-EM was used to study the polymerase module-RNA complex in absence or presence of PABPN1.

Despite testing many different conditions and assemblies (PAPOA wildtype or catalytic mutant, various RNAs and nucleotides, absence or presence of PABPN1 in different stoichiometries, different buffers and detergents, with and without cross-linking), no intact complex could be visualized. The 2D class averages of cryo-EM screening datasets clearly revealed structural details but already indicated the disassembly of

the complex (**Fig. 29A**). Fitting of available structural models into the respective 3D maps revealed that mPSF and the PAPOA catalytic domain could be reconstructed independently (**Fig. 29B, C**). The mPSF reconstruction showed secondary structure features (6.6 Å resolution) and an available model bound to an AAUAAA RNA was readily fitted into the map with high confidence (**Fig. 29B**) (Clerici et al., 2018).

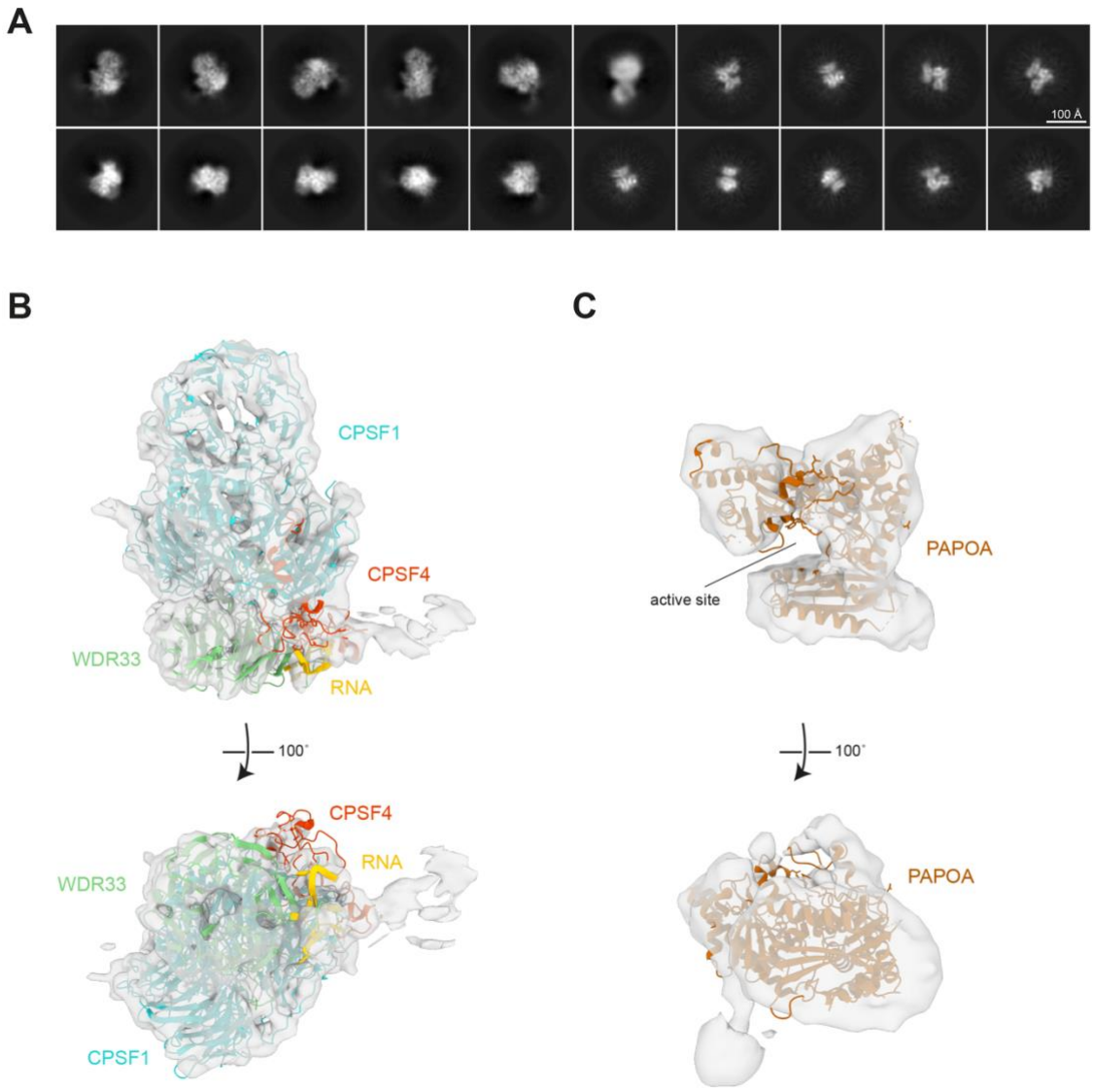

Figure 29: Structural characterization of PM. (A) Representative cryo-EM 2D class averages of picked particles. Scale bar ~ 100 Å. **(B)** Cryo-EM 3D reconstruction of mPSF particles with fitted structural model of mPSF-AAUAAA RNA (PDB: 6FBS) in two different orientations. **(C)** Cryo-EM 3D reconstruction of PAPOA particles with fitted structural model of the PAPOA catalytic domain (PDB: 1Q79) in two different orientations. The PAPOA active site is indicated.

The PAPOA particles could only be reconstructed to slightly lower resolution (9.3 Å resolution), mainly because of the smaller molecular weight (59 kDa for the PAPOA

catalytic domain) and missing views. Nevertheless, the catalytic domain could be docked reliably into the cryo-EM map while the unstructured C-terminus and PABPN1 were not visible (**Fig. 29C**) (Martin et al., 2004). Hence, while the available cryo-EM data clearly indicated the presence of a high-quality sample, structural details regarding PAPOA recruitment or regulation by PABPN1 remained elusive.

3.3.3 PolyA RNPs

As mentioned previously, it is still poorly understood how poly(A) tail length control is achieved, the mechanism how processive polyadenylation is terminated once the characteristic *in vivo* length is reached. Therefore, the effect of potential regulatory factors on polyadenylation by PM and PABPN1 was assessed to reconstitute this mechanism *in vitro* (**Fig. 30A**). In the studied reaction conditions, the growing poly(A) tail significantly exceeded 400 nucleotides, far longer than *in vivo*, suggesting that polyadenylation might continue until ATP is depleted (Eckmann et al., 2011). Addition of various protein factors did not change this behavior, showing that controlled and specific termination of poly(A) tail synthesis could not be reconstituted *in vitro* (**Fig. 30A**).

The current model of terminating polyadenylation is based on low-resolution EM data of glutaraldehyde-stabilized poly(A)-PABPN1 RNPs (Keller et al., 2000). Non-cross-linked PABPN1 RNPs with different poly(A) RNAs (90A and 200A) were reconstituted and subjected to single-particle cryo-EM analysis to validate the available data (**Fig. 30B-D**). The 2D class averages revealed the presence of globular and compact particles, indicating that the RNPs were collapsing into an almost spherical form (**Fig. 30B**). 3D reconstructions showed overall similar shapes, although with varying sizes (~ 70 x 70 x 95 Å for 90A, ~ 90 x 100 x 120 Å for 200A). Interestingly, the exact trace of the poly(A) RNA could not be recognized, suggesting that similar, but differing particles were averaged. Moreover, although a structural model of the PABPN1 RRM could be placed into some bulges, the overall resolution did not allow a confident structural fitting of RNA or protein molecules (**Fig. 30C, D**).

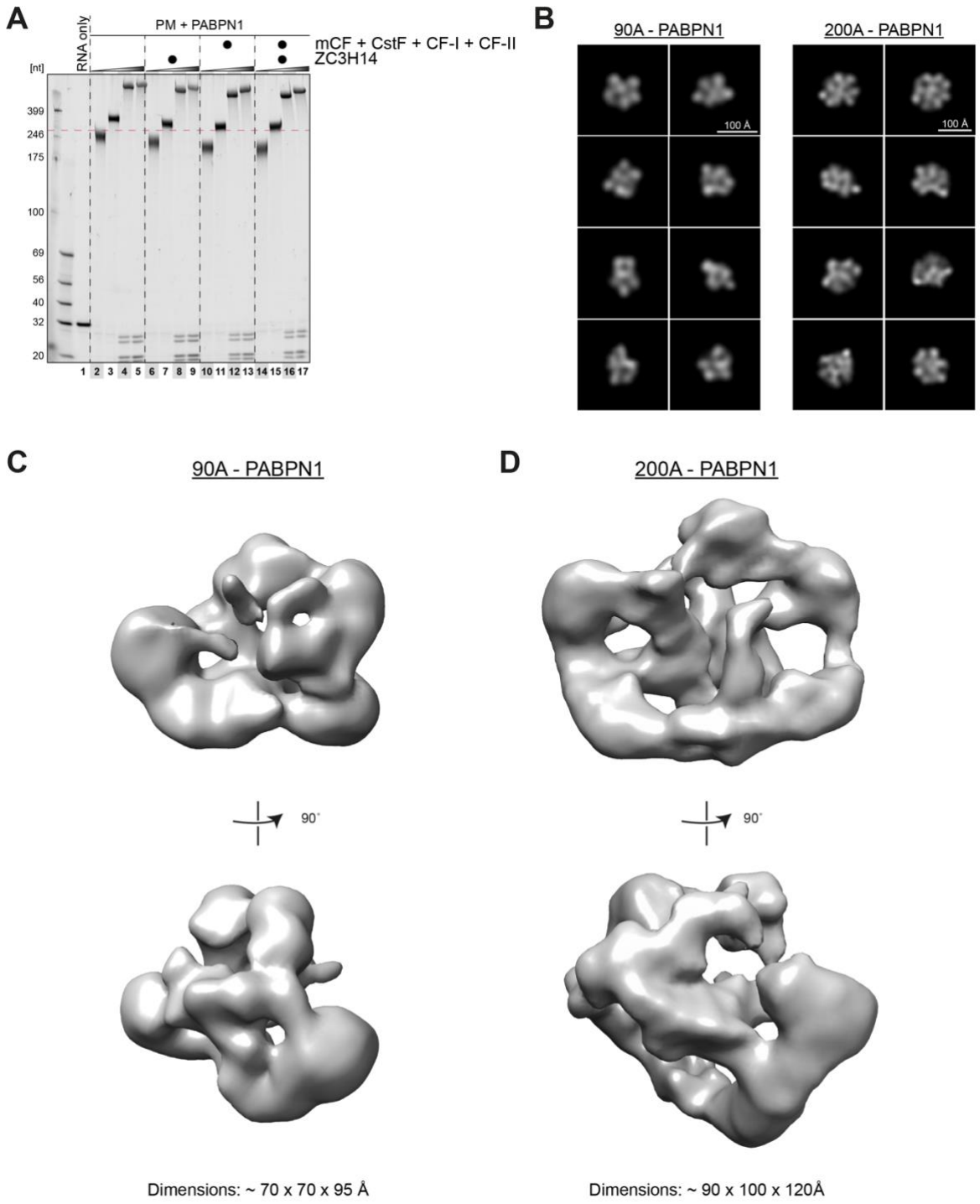

Figure 30: Poly(A)-PABPN1 RNPs. (A) Fluorescent polyadenylation assay showing that RNA substrates are hyperadenylated *in vitro*. The red dashed line approximates the migration pattern of a 250A RNA. **(B)** Representative cryo-EM 2D class averages of picked particles of 90A-PABPN1 (left panel) and 200A-PABPN1 (right panel) RNPs. Scale bars ~ 100 Å. **(C)** Cryo-EM 3D reconstruction of 90A-PABPN1 RNP particles in two different orientations. Approximate dimension of the particles is indicated below. **(D)** Cryo-EM 3D reconstruction of 200A-PABPN1 RNP particles in two different orientations. Approximate dimension of the particles is indicated below.

3.3.4 Characterization of OPMD mutants

OPMD is a muscular dystrophy, which is caused by an extension of the N-terminal Alanine stretch of PABPN1 (**Fig. 28A**) (Brais et al., 1998). While formation of intranuclear aggregation sequestering PABPN1 and potentially other proteins is thought to be causing the disease phenotypes, it is still unclear whether the role of PABPN1 in nuclear pre-mRNA polyadenylation might contribute to the disease (Banerjee et al., 2013). Hence, two disease mutants with different Alanine extensions (A13 and A18) were purified to characterize them biochemically and structurally (**Fig. 28B**). First, RNA-binding behaviors of wildtype and mutated PABPN1 variants were compared (**Fig. 31A**). Interestingly, in gel-shift binding assays (EMSA), the different PABPN1 preparations showed highly similar migration patterns, both for a 15A RNA as well as for a 35A RNA, which allowed cooperative RNA binding (**Fig. 31A**).

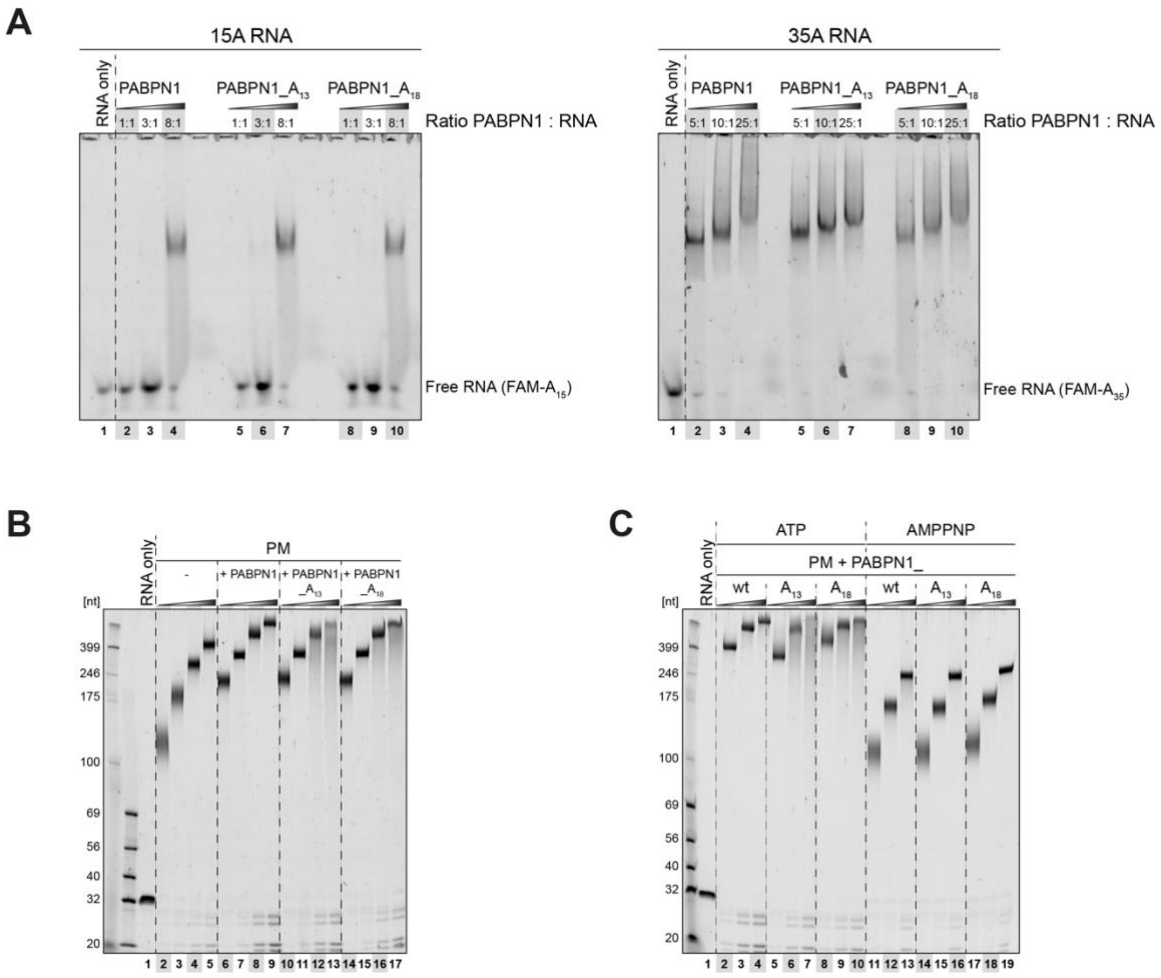

Figure 31: Characterization of PABPN1 OPMD variants. (A) Fluorescent EMSA with 15A (left panel) or 35A (right panel) RNA showing that the PABPN1 variants exhibit similar RNA-binding properties. **(B)** Fluorescent polyadenylation assay of PM with PABPN1 variants showing different gel migration

patterns for very long RNAs. **(C)** Fluorescent polyadenylation assay of PM with PABPN1 variants in presence of either ATP or AMPPNP showing that the observed differences in gel migration patterns are RNA-length dependent.

In polyadenylation time-course experiments, all three PABPN1 variants comparably enhanced PAPOA activity in early time points (**Fig. 31B**). These results suggested that the PABPN1-RNA and PABPN1-PAPOA interactions were not affected by the N-terminal Alanine extensions. However, longer incubation times resulted in reproducible differences in the RNA migration patterns as the otherwise defined bands were replaced by smears for samples containing the PABPN1 mutants (**Fig. 31B**). Next, ATP was replaced by AMPPNP to test whether this behavior was caused by unspecific nuclease contamination, which would only become apparent over time (**Fig. 31C**). Although PAPOA can readily hydrolyze AMPPNP to elongate the growing poly(A) tail, polyadenylation was significantly slower. While ATP reproduced the PABPN1 mutant-dependent band smear, all variants showed similar behavior in presence of AMPPNP (**Fig. 31C**). This suggested an effect dependent on the length of the synthesized RNA, not on the incubation time. However, it remained unclear what caused this effect and whether it would be relevant *in vivo* as the generated poly(A) tails were significantly longer than expected in cells.

4. Discussion

Mammalian 3' end processing is a crucial step during eukaryotic gene expression, yet many important biochemical and structural aspects are poorly understood. Here, by establishing or extending existing *in vitro* reconstitution systems, the presented work provided novel insights into both catalytic reactions, endonucleolytic cleavage and polyadenylation.

4.1 CPSF3 is stored catalytically inactive until recruited to a substrate RNA

The endonuclease CPSF3, responsible for cleaving all pre-mRNA substrates within cells, is likely kept in a catalytically inhibited state. MCF, which forms a stable complex in human cells, and CPSF are inactive *in vitro*, likely because the CPSF3 RNA channel reaching towards the active site remains in a closed conformation (Mandel et al., 2006b). As CPSF3 itself exhibits only weak sequence specificity, storing the endonuclease in an inhibited state seems crucial to prevent spurious cleavage events at non-target sites (Sheets et al., 1990; Sun et al., 2020b). Instead, activation of mCF only after recruitment to the desired pre-mRNA substrate allows intricate regulatory networks to fine-tune cleavage site selection and subsequently, transcription termination (Proudfoot, 2016).

Replication-dependent histone pre-mRNAs are processed by a distinct set of proteins, but including mCF, termed histone pre-mRNA cleavage complex (HCC), as responsible endonuclease (Dominski and Marzluff, 2007). After recruitment by an U7 snRNP-FLASH-SLBP complex, CPSF3 transitions into an active conformation, distinct to the state observed in mCF or CPSF (Sun et al., 2020b). Likely, canonical pre-mRNA cleavage is achieved by a similar mechanism, activating CPSF3 only in context of the correct substrate. It is tempting to envision that this also restricts the inherent conformational flexibility currently observed for all mCF or CPSF complexes (**Fig. 15, 17, 23**) (Zhang et al., 2020). Hence, storing CPSF3 in an inactive state within mCF does not only prevent widespread off-target cleavage but also allows fine-tuning of its activity by recruitment with different RNA-recognizing machineries dependent on the substrate class.

4.2 CPSF acts as scaffold for relaying substrate recognition

The CPA machinery consists of several RNA-interacting modules, thereby allowing high specificity towards defined cleavage sites within the 3' UTR of pre-mRNA molecules despite their diversity regarding the exact nucleotide sequence. Hence, recognition of their binding motifs needs to be relayed to activate CPSF3, most likely via the CPSF submodules mPSF and mCF as scaffolds. CF-II interacts directly with mCF, possibly via the Symplekin CTD, which is composed of helical HEAT repeats known for protein-protein interactions (**Fig. 15, 19**) (Kennedy et al., 2009). Negative stain EM 2D class averages of the mCF-CF-II complex showed a third protein lobe next to CPSF2 and CPSF3, consistent with the close spatial proximity of the Symplekin CPSF3-binding and CTD domains (**Fig. 19, 20, 32**). In histone pre-mRNA processing, FLASH connects the Symplekin CTD to SLBP bound to the 3' end stem-loop structure, a crucial factor for RNA recognition and cleavage site positioning (Sun et al., 2020b). A similar binding network could be envisioned for the mCF-CF-II interaction, consistent with the idea of different protein factors but similar activation mechanisms between canonical and histone pre-mRNA processing.

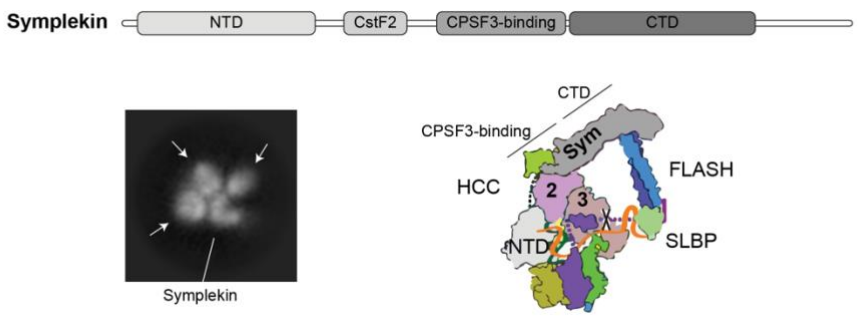

Figure 32: Symplekin acts as scaffold in 3' end processing machineries. Domain organization of Symplekin (top panel). Representative negative stain 2D class average of mCF-CF-II with the three globular lobes indicated with arrows next to Symplekin (bottom left panel). Schematic model of active histone pre-mRNA processing complex highlighting that the Symplekin (Sym) CTD binds to FLASH, which helps to recruit the HCC to the substrate pre-mRNA (bottom right panel). Adapted from (Sun et al., 2020b).

As reported previously, CstF does not directly interact with mCF but is recruited via mPSF (**Fig. 19**) (Ruepp et al., 2011b; Zhang et al., 2020). However, the CstF2 subunit is present in both CstF and mCF modules due to a mutually exclusive interaction of CstF2 with CstF3 or Symplekin (Ruepp et al., 2011b). Currently, it remains unknown how many copies of CstF2 are present in the fully assembled CPA machinery or

whether the different binding partners represent distinct stages during the cleavage cycle. Moreover, the exact function of CstF2 in mCF or HCC is still unclear as it is not necessary for *in vitro* cleavage of any pre-mRNA substrate class and could not be visualized by cryo-EM in context of the histone pre-mRNA processing machinery (Schmidt et al., 2022; Sun et al., 2020b). To address these open questions, more structural insights into the activated pre-mRNA cleavage complex are necessary.

4.3 *In vitro* reconstitution of pre-mRNA cleavage defines a minimal complex

The reconstitution of the pre-mRNA endonuclease activity using purified proteins described in this work allowed for the first time to characterize the reaction using a highly defined and controlled *in vitro* system. A minimal complex containing 14 subunits was defined to be necessary and sufficient for specific and efficient pre-mRNA cleavage. While mCF, mPSF, PAPOA, CstF, CF-II, and RBBP6 were required for nuclease activity, CF-I and Ssu72 were dispensable, consistent with their prominent role in alternative polyadenylation and coupling of transcription to pre-mRNA processing, respectively (**Fig. 21**) (Pereira-Castro and Moreira, 2021; Tian and Manley, 2017; Xiang et al., 2010). However, a concurrent reconstitution described *in vitro* pre-mRNA cleavage without the requirement of PAPOA (Boreikaite et al., 2022). Although it remains to be determined which subtle differences in the reaction setups are responsible for this discrepancy, it is tempting to speculate that PAPOA is present during the cleavage event, potentially in an inactivated state. The presence of PAPOA would allow a direct hand-over of the newly established 3' end from the CPSF3 to the PAPOA active sites for subsequent polyadenylation, obstructing any exonucleolytic attack on a possibly unbound RNA end.

Similar disparities were described for the requirement of ATP (Boreikaite et al., 2022; Schmidt et al., 2022), which could at least partly be explained by the reaction conditions, as exemplified here. While ATP addition clearly stimulated the activity for both RNA substrates, reactions only with the less efficiently cleaved L3 RNA would suggest that ATP was strictly required for nuclease activity (**Fig. 22**). Even though the nucleotide might not be needed *in vitro*, subsequent polyadenylation absolutely requires ATP, making efficient cleavage only in presence of ATP and PAPOA an

attractive regulatory mechanism to ensure its availability. Interestingly, ATP-dependent simulation was enhanced by the addition of PAPOA, pointing towards a mechanism of coupling pre-mRNA cleavage and polyadenylation (**Fig. 27**).

Although the tested pre-mRNA substrates were cleaved specifically, overall activity was relatively low, especially considering that an excess of protein over RNA was used. The underlying problem was likely inefficient complex assembly on the RNA, as indicated by the low recovery rates in RNase protection and pull-down experiments (**Fig. 21, 22**). It remains unclear whether this is an intrinsic property of the CPA machinery in order to fine-tune cleavage site selection in presence of multiple alternative polyadenylation cassettes or represents an artefact of the minimal *in vitro* system. *In vivo*, 3' end processing is coupled to transcription and the CPA complex is recruited by the RNA polymerase II CTD domain, facilitating substrate recognition and probably stimulating the endonuclease activity (Proudfoot, 2016). Comparison of both pre-mRNA substrates, L3 and SV40, showed significantly better processing efficiencies using the SV40 RNA (**Fig. 22**). Sequence analysis revealed that SV40 is mostly U- and G-rich downstream of the cleavage site, while L3 exhibits a high C content, clearly deviating from the optimal consensus sequence (Neve et al., 2017; Tian and Graber, 2012). This suggested that different binding affinities of CstF and CF-II, which recognize (G)U- and G-rich downstream motifs, respectively, modulate cleavage efficiencies in this minimal *in vitro* system, similarly to usage of weak and strong poly(A) sites *in vivo* (Neve et al., 2017; Tian and Graber, 2012).

4.4 RBBP6 is critical for CPSF3 endonuclease activation

The combined structural and biochemical data provided additional insights into the architecture of the CPA machinery by finally allowing to assign CPSF3 as corresponding to the CPSF proximal lobe (**Fig. 23**). This places the endonuclease next to the mPSF module and suggests a likely RNA substrate path, potentially rationalizing the *in vivo* distance of 21-22 nucleotides between AAUAAA PAS and cleavage site (Gruber et al., 2016; Sun et al., 2020b; Tian et al., 2005; Zhang et al., 2020). Previously, RBBP6 was mostly implicated in alternative polyadenylation (Di Giammartino et al., 2014), however, the reported data indicate a more prominent role, being directly involved in nuclease activation. Thus, its role in pre-mRNA cleavage and

its recruitment to the endonuclease are conserved from yeast to human (**Fig. 23, 24, 25, 33**) (Hill et al., 2019).

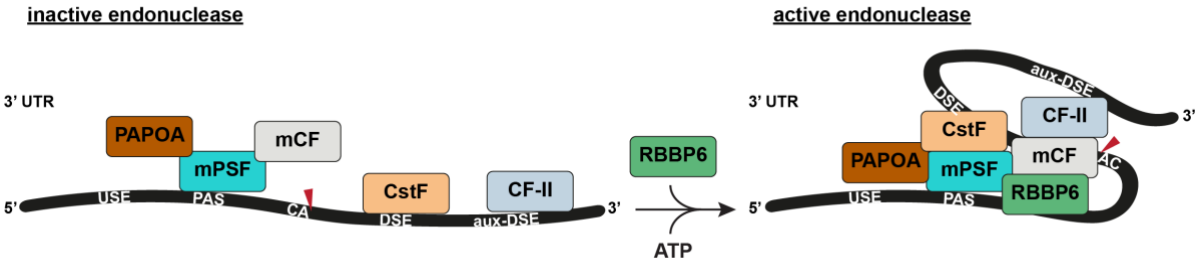

Figure 33: Model for activation of CPSF3 cleavage. Coordinated assembly of mPSF, mCF, PAPOA, CstF, CF-II, and RBBP6 on the pre-mRNA substrate is required. RBBP6 and ATP activate and stimulate CPSF3 endonuclease activity, respectively.

Although the CPSF3-RBBP6 interface represents the main binding site, additional contacts to mPSF are required for mRNA cleavage (**Fig. 23, 24, 25**) (Boreikaite et al., 2022). This might provide a rationale for the inhibitory effect of RBBP6 isoform 3, a shortened isoform including the N-terminal DWNN domain but lacking almost all residues C-terminally (Di Giammartino et al., 2014). As the DWNN domain is probably sufficient for CPSF3 binding, the additional contact sites are likely involved in allowing endonucleolytic cleavage, potentially by stabilizing an active conformation or orienting the RNA substrate.

Interestingly, although an N-terminal RBBP6 construct is sufficient for its activity in pre-mRNA cleavage, it contains a very long and mostly unstructured C-terminus, which harbors binding sites for transcription factors such as Rb or p53 (Simons et al., 1997). As they are absent in the corresponding yeast protein Mpe1, it is tempting to speculate that they were acquired in mammals to allow more intricate regulatory networks coupling transcription to diverse RNA processing pathways.

4.5 Comparison to the Integrator complex

The pre-mRNA 3' end machineries share common features with the Integrator complex, which was originally discovered as specialized 3' end processing complex for snRNAs (Baillat et al., 2005). Both machineries, with similar molecular weight and number of subunits, contain nuclease and phosphatase activities, thereby combining RNA processing with modulation of the Pol II CTD phosphorylation pattern (Pfleiderer

and Galej, 2022; Zheng et al., 2020). They share a common active site architecture, featuring a tightly intertwined pseudo-dimer of an inactive and an active nuclease, CPSF2-3 for CPA and INTS9-11 in Integrator, which are homologous proteins and likely have a common evolutionary origin (Baillat et al., 2005; Dominski et al., 2005; Pfleiderer and Galej, 2021, 2022; Zheng et al., 2020; Sun et al., 2020b). Both endonucleases are kept in an inactive state requiring additional factors to gain catalytic competence. INTS11 is activated through an association with promotor-proximally paused Pol II by negative transcription elongation factor SPT5, which pushes into the INTS11 ß-CASP lid to allow access to the active center (Fianu et al., 2021; Zheng et al., 2021). For histone pre-mRNA cleavage, proteins of the Lsm ring induce CPSF3 activation after substrate recognition; however, as these factors are not present for canonical 3' end processing, the exact mechanism of active site opening remains elusive (Sun et al., 2020b). Overall, all three metazoan 3' end processing machineries combine a common active center with additional factors specializing it to certain cellular tasks (Pfleiderer and Galej, 2022). Functionally however, pre-mRNA cleavage is a crucial event required for mRNA maturation, while Integrator has a mostly negative, abortive role on transcription by contributing to promotor-proximal transcription attenuation via RNA cleavage and PP2A-mediated CTD dephosphorylation during Pol II pausing (Elrod et al., 2019; Fianu et al., 2021; Hsin and Manley, 2012; Lykke-Andersen et al., 2021; Tatomer et al., 2019).

4.6 *In vitro* polyadenylation reproduces cellular PAPOA characteristics

As pre-mRNA processing occurs co-transcriptionally and is template-independent, PAPOA requires high specificity towards ATP to discriminate against the other nucleotides, which are similarly available. Despite the absence of any competing nucleotides, PAPOA only uses ATP efficiently *in vitro*, clearly reproducing the *in vivo* specificity as intrinsic characteristic of the polymerase (**Fig. 28**) (Legnini et al., 2019). In the last years, it became increasingly clear that cellular poly(A) tails contain non-A nucleotides, although the function of these mixed tails is still unclear (Legnini et al., 2019; Lim et al., 2018). Interestingly, FLAM-seq, an RNA sequencing method combining full-length poly(A) tail and mRNA sequencing, determined non-A nucleotide

frequencies in different human cell types (A >> C > U > G), which closely match the *in vitro* properties (**Fig. 28**) (Legnini et al., 2019). This raises the question whether mixed tailing actually serves a regulatory purpose or only represents misincorporation events by PAPOA.

For efficient and processive PAPOA activity, mPSF and PABPN1 are required for enhanced RNA substrate binding, allowing the possibility that the primarily cytoplasmic PABPC1 replaces PABPN1 to stimulate polyadenylation (Hosoda et al., 2006; Wahle, 1991b). However, while addition of PABPN1 reproduced the reported PAPOA stimulation, PABPC1 clearly inhibited poly(A) tail synthesis (**Fig. 28**). Interestingly, the polyadenylation pattern with PABPC1 was reminiscent of the stepwise poly(A) tail removal of yeast Pan2-Pan3 in presence of Pab1 (Schäfer et al., 2019). Deadenylase activity of Pan2-Pan3 is slowed down when obstructed by poly(A)-bound Pab1, but once the 3' terminal Pab1 is removed, degradation continues rapidly, generating a stepwise decay pattern. This suggests a likewise mechanism in PABPC1-dependent inhibition of polyadenylation. After synthesis of a poly(A) stretch allowing binding, PABPC1 might compete more efficiently for the RNA 3' end, thereby inhibiting PAPOA activity and generating stepwise poly(A) elongation. As these *in vitro* results are in conflict with previously reported data, it remains to be determined whether PABPC1 might have different effects dependent on post-translational modifications, specific pre-mRNA substrates or cellular contexts (Hosoda et al., 2006).

In cells, polyadenylation needs to be terminated in a controlled manner to decorate mRNAs with poly(A) tails of defined length, a poorly understood aspect of PAPOA activity. Using the minimal set of proteins required for efficient and processive activity, PAPOA, mPSF, and PABPN1, *in vitro* polyadenylation generated RNAs significantly longer than 400A, exceeding cellular synthesis lengths (**Fig. 30**) (Eckmann et al., 2011). In yeast, *in vitro* reconstitution of poly(A) tail length control required the CPA machinery and poly(A)-binding proteins Pab1 and Nab2 in defined stoichiometries (Turtola et al., 2021). However, experimental conditions presented here lacked either PABPN1 (**Fig. 21**) or components required for pre-mRNA cleavage (**Fig. 30**), suggesting that one or several factors needed for poly(A) tail length control were missing.

4.7 PABPN1-poly(A) RNPs form spherical particles

PABPN1-coating of the poly(A) tail during its synthesis is thought to generate spherical particles, based on low resolution negative stain data of chemically-fixed RNPs containing different poly(A) RNAs (Keller et al., 2000). These compact shapes are thought to allow PAPOA to maintain contact to mPSF despite the growing nucleotide sequence between the AAUAAA PAS and the RNA 3' end. Cryo-EM studies presented in this work revealed for the first time the architecture of non-cross-linked polyA-PABPN1 RNPs (**Fig. 30**). Despite likely sufficient resolution, no clear RNA path could be defined, probably because similar but differing shapes were averaged during cryo-EM data processing. As PABPN1 is known to oligomerize using its unstructured N- and C-termini, the RNPs might collapse mediated by these interactions between different PABPN1 protomers (Keller et al., 2000; Kühn et al., 2003; Meyer et al., 2002). Experiments using PABPN1 deposition during polyadenylation are needed to examine whether this gradual assembly would result in more homogeneous particles allowing higher resolution reconstructions. It is conceivable that the spherical shape of the RNP is the determining characteristic recognized by factors involved in poly(A) tail length control or downstream processes in nucleus and cytoplasm. In yeast deadenylation, Pan2-Pan3 recognizes an extended 90A-Pab1 RNP, suggesting that poly(A) tail modifying enzymes might be recruited or regulated by the poly(A) RNP 3D structure (Schäfer et al., 2019).

Additionally, it remains unclear whether the formation of the globular particles is involved in termination of processive poly(A) tail synthesis. While both RNPs, 90A and 200A, form similar shapes with distinct dimensions, they are still shorter than the average poly(A) tail in human nuclei (**Fig. 30**) (Eckmann et al., 2011). Experiments using RNA exceeding this length are lacking to visualize potential differences; however, generation of RNA homopolymers with this size remains technically very challenging.

4.8 Role of pre-mRNA polyadenylation in OPMD

Extensions of the N-terminal PABPN1 Alanine stretch caused differences in the polyadenylation behavior, although the PABPN1-RNA and PABPN1-PAPOA interactions were likely not influenced, based on similar RNA binding and PAPOA

stimulation properties (**Fig. 31**). Only very long poly(A) tails were affected, suggesting that altered 3D assemblies might be involved. Hence, cryo-EM reconstructions of the mutated PABPN1 RNPs are needed to address these open questions. However, the *in vivo* significance of the observed disparities remains unclear, considering the RNA size exceeding nuclear poly(A) tail lengths. Overall, although *in vitro* polyadenylation assays are useful experiments to study diverse aspects of pre-mRNA 3' end processing, OPMD is probably caused by cellular protein aggregation, which clearly limits the interpretability of these *in vitro* studies using simplified reaction setups (Banerjee et al., 2013).

5. Outlook

As this thesis described the *in vitro* reconstitution of pre-mRNA cleavage, both enzymatic reactions involved in 3' end processing can finally be studied using purified proteins, extending the experimental framework available to assay this essential co-transcriptional mechanism.

One major obstacle in further understanding the endonuclease reaction is the lack of structural insights visualizing the activated complex before and after cleavage of its substrate. Given the size of the machinery and its intricate regulation, structural studies will be needed to dissect many important aspects, including the detailed mechanism of CPSF3 activation, the role of ATP in nuclease stimulation, how RNA motif recognition is relayed by different modules, or the hand-over of the newly generated 3' end between the catalytic sites. But considering the successful assembly and first negative stain screening attempts of the active CPA machinery, obtaining these structural insights seems now within reach.

Furthermore, subtle optimization should suffice to extend the pre-mRNA cleavage assays to study diverse aspects of 3' end processing, including alternative polyadenylation, where *in vitro* experiments could complement the plethora of RNA sequencing data produced in recent years. Moreover, addition of purified RNA polymerase II should allow to increase the complexity of the system in order to investigate various facets of co-transcriptional RNA processing. As transcription termination is facilitated by the exonucleolytic degradation by Xrn2, its recruitment might involve CPA proteins since they likely remain bound to the downstream fragment.

Building on the presented results, structural analysis of PAPOA bound to mPSF should extend our knowledge about poly(A) tail synthesis and, especially in complex with poly(A)-PABPN1 RNPs, length control. Given the advances in biochemical characterization of both enzymatic activities and the availability of many high-quality recombinant factors, understanding these open questions should be feasible.

References

Adelman, K., Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. 13, 720–731. https://doi.org/10.1038/nrg3293

Aguilera, A., García-Muse, T., 2012. R loops: from transcription byproducts to threats to genome stability. Mol. Cell 46, 115–124. https://doi.org/10.1016/j.molcel.2012.04.009

Ahearn, J.M., Bartolomei, M.S., West, M.L., Cisek, L.J., Corden, J.L., 1987. Cloning and sequence analysis of the mouse genomic locus encoding the largest subunit of RNA polymerase II. J. Biol. Chem. 262, 10695–10705.

Ahn, S.H., Kim, M., Buratowski, S., 2004. Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. Mol. Cell 13, 67–76. https://doi.org/10.1016/S1097-2765(03)00492-1

Aik, W.S., Lin, M.-H., Tan, D., Tripathy, A., Marzluff, W.F., Dominski, Z., Chou, C.-Y., Tong, L., 2017. The N-terminal domains of FLASH and Lsm11 form a 2:1 heterotrimer for histone pre-mRNA 3'-end processing. PLOS ONE 12, e0186034. https://doi.org/10.1371/journal.pone.0186034

Alexander, R.D., Innocente, S.A., Barrass, J.D., Beggs, J.D., 2010. Splicing-dependent RNA polymerase pausing in yeast. Mol. Cell 40, 582–593. https://doi.org/10.1016/j.molcel.2010.11.005

Apponi, L.H., Leung, S.W., Williams, K.R., Valentini, S.R., Corbett, A.H., Pavlath, G.K., 2010. Loss of nuclear poly(A)-binding protein 1 causes defects in myogenesis and mRNA biogenesis. Hum. Mol. Genet. 19, 1058–1065. https://doi.org/10.1093/hmg/ddp569

Aravind, L., 1999. An evolutionary classification of the metallo-beta-lactamase fold proteins. In Silico Biol. 1, 69–91.

Arhin, G.K., Boots, M., Bagga, P.S., Milcarek, C., Wilusz, J., 2002. Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. Nucleic Acids Res. 30, 1842–1850. https://doi.org/10.1093/nar/30.8.1842

Awasthi, S., Alwine, J.C., 2003. Association of polyadenylation cleavage factor I with U1 snRNP. RNA 9, 1400–1409. https://doi.org/10.1261/rna.5104603

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., Söding, J., Cramer, P., 2017. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol. Cell 66, 38-49.e6. https://doi.org/10.1016/j.molcel.2017.02.009

Bagga, P.S., Ford, L.P., Chen, F., Wilusz, J., 1995. The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. Nucleic Acids Res. 23, 1625–1631. https://doi.org/10.1093/nar/23.9.1625

Bai, Y., Auperin, T.C., Chou, C.-Y., Chang, G.-G., Manley, J.L., Tong, L., 2007. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. Mol. Cell 25, 863–875. https://doi.org/10.1016/j.molcel.2007.01.034

Baillat, D., Hakimi, M.-A., Näär, A.M., Shilatifard, A., Cooch, N., Shiekhattar, R., 2005. Integrator, a multiprotein mediator of small nuclear RNA processing, associates

with the C-terminal repeat of RNA polymerase II. Cell 123, 265–276. https://doi.org/10.1016/j.cell.2005.08.019
Banerjee, A., Apponi, L.H., Pavlath, G.K., Corbett, A.H., 2013. PABPN1: molecular function and muscle disease. FEBS J. 280, 4230–4250. https://doi.org/10.1111/febs.12294
Barabino, S.M., Hübner, W., Jenny, A., Minvielle-Sebastia, L., Keller, W., 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev. 11, 1703–1716. https://doi.org/10.1101/gad.11.13.1703
Barabino, S.M.L., Ohnacker, M., Keller, W., 2000. Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J. 19, 3778–3787. https://doi.org/10.1093/emboj/19.14.3778
Barillà, D., Lee, B.A., Proudfoot, N.J., 2001. Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. 98, 445–450. https://doi.org/10.1073/pnas.98.2.445
Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., Gautheret, D., 2000. Patterns of variant polyadenylation signal usage in human genes. Genome Res. 10, 1001–1010. https://doi.org/10.1101/gr.10.7.1001
Berkovits, B.D., Mayr, C., 2015. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. Nature 522, 363–367. https://doi.org/10.1038/nature14321
Bienroth, S., Keller, W., Wahle, E., 1993. Assembly of a processive messenger RNA polyadenylation complex. EMBO J. 12, 585–594. https://doi.org/10.1002/j.1460-2075.1993.tb05690.x
Birchmeier, C., Schümperli, D., Sconzo, G., Birnstiel, M.L., 1984. 3’ editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. U. S. A. 81, 1057–1061. https://doi.org/10.1073/pnas.81.4.1057
Biyani, N., Righetto, R.D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K.N., Stahlberg, H., 2017. Focus: The interface between data collection and data processing in cryo-EM. J. Struct. Biol. 198, 124–133. https://doi.org/10.1016/j.jsb.2017.03.007
Bogard, N., Linder, J., Rosenberg, A.B., Seelig, G., 2019. A Deep Neural Network for Predicting and Engineering Alternative Polyadenylation. Cell 178, 91-106.e23. https://doi.org/10.1016/j.cell.2019.04.046
Boreikaite, V., Elliott, T.S., Chin, J.W., Passmore, L.A., 2022. RBBP6 activates the pre-mRNA 3′ end processing machinery in humans. Genes Dev. genesdev;gad.349223.121v1. https://doi.org/10.1101/gad.349223.121
Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chrétien, N., Tomé, F.M., Lafrenière, R.G., Rommens, J.M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A.D., Heutink, P., Mathieu, J., Duranceau, A., Codère, F., Fardeau, M., Rouleau, G.A., Korcyn, A.D., 1998. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat. Genet. 18, 164–167. https://doi.org/10.1038/ng0298-164
Bresson, S.M., Conrad, N.K., 2013. The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. PLoS Genet. 9, e1003893. https://doi.org/10.1371/journal.pgen.1003893

Brown, K.M., Gilmartin, G.M., 2003. A Mechanism for the Regulation of Pre-mRNA 3′ Processing by Human Cleavage Factor Im. Mol. Cell 12, 1467–1476. https://doi.org/10.1016/S1097-2765(03)00453-2

Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K.J., Ji, F., Schwarz, B.A., Kalocsay, M., Elledge, S.J., Chen, Y., Sadreyev, R.I., Gygi, S.P., Hu, G., Shi, Y., Hochedlinger, K., 2018. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. Cell 172, 106-120.e21. https://doi.org/10.1016/j.cell.2017.11.023

Buratowski, S., 2009. Progression through the RNA polymerase II CTD cycle. Mol. Cell 36, 541–546. https://doi.org/10.1016/j.molcel.2009.10.019

Calado, A., Tome, F.M.S., Brais, B., Rouleau, G.A., Kuhn, U., Wahle, E., Carmo-Fonseca, M., 2000. Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA. Hum. Mol. Genet. 9, 2321–2328. https://doi.org/10.1093/oxfordjournals.hmg.a018924

Callebaut, I., Moshous, D., Mornon, J.-P., de Villartay, J.-P., 2002. Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Res. 30, 3592–3601. https://doi.org/10.1093/nar/gkf470

Casañal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K., Dornan, G.L., Skehel, M., Robinson, C.V., Passmore, L.A., 2017. Architecture of eukaryotic mRNA 3′-end processing machinery. Science 358, 1056–1059. https://doi.org/10.1126/science.aao6535

Ceelie, H., Spaargaren-van Riel, C.C., Bertina, R.M., Vos, H.L., 2004. G20210A is a functional mutation in the prothrombin gene; effect on protein levels and 3'-end formation. J. Thromb. Haemost. JTH 2, 119–127. https://doi.org/10.1111/j.1538-7836.2003.00493.x

Chan, S., Choi, E.-A., Shi, Y., 2011. Pre-mRNA 3′-end processing complex assembly and function: Pre-mRNA 3′-end processing complex assembly. Wiley Interdiscip. Rev. RNA 2, 321–335. https://doi.org/10.1002/wrna.54

Chan, S.L., Huppertz, I., Yao, C., Weng, L., Moresco, J.J., Yates, J.R., Ule, J., Manley, J.L., Shi, Y., 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3′ processing. Genes Dev. 28, 2370–2380. https://doi.org/10.1101/gad.250993.114

Chang, J.W., Yeh, H.S., Yong, J., 2017. Alternative Polyadenylation in Human Diseases. Endocrinol. Metab. Seoul Korea 32, 413–421. https://doi.org/10.3803/EnM.2017.32.4.413

Chen, J.-M., Férec, C., Cooper, D.N., 2006. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. Hum. Genet. 120, 1–21. https://doi.org/10.1007/s00439-006-0180-7

Chen, Z., Li, Y., Krug, R.M., 1999. Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. EMBO J. 11.

Chibi, M., Meyer, M., Skepu, A., G Rees, D.J., Moolman-Smook, J.C., Pugh, D.J.R., 2008. RBBP6 interacts with multifunctional protein YB-1 through its RING finger domain, leading to ubiquitination and proteosomal degradation of YB-1. J. Mol. Biol. 384, 908–916. https://doi.org/10.1016/j.jmb.2008.09.060

Cho, E.-J., Kobor, M.S., Kim, M., Greenblatt, J., Buratowski, S., 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase

II C-terminal domain. Genes Dev. 15, 3319–3329. https://doi.org/10.1101/gad.935901
Cho, E.J., Takagi, T., Moore, C.R., Buratowski, S., 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. 11, 3319–3326. https://doi.org/10.1101/gad.11.24.3319
Clerici, M., Faini, M., Aebersold, R., Jinek, M., 2017. Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. eLife 6, e33111. https://doi.org/10.7554/eLife.33111
Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., Jinek, M., 2018. Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat. Struct. Mol. Biol. 25, 135–138. https://doi.org/10.1038/s41594-017-0020-6
Colgan, D.F., Murthy, K.G., Prives, C., Manley, J.L., 1996. Cell-cycle related regulation of poly(A) polymerase by phosphorylation. Nature 384, 282–285. https://doi.org/10.1038/384282a0
Connelly, S., Manley, J.L., 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev. 2, 440–452. https://doi.org/10.1101/gad.2.4.440
Coppola, J.A., Field, A.S., Luse, D.S., 1983. Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. Proc. Natl. Acad. Sci. U. S. A. 80, 1251–1255. https://doi.org/10.1073/pnas.80.5.1251
Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., Bentley, D.L., 2019. Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. Mol. Cell 76, 896-908.e4. https://doi.org/10.1016/j.molcel.2019.09.031
Coseno, M., Martin, G., Berger, C., Gilmartin, G., Keller, W., Doublié, S., 2008. Crystal structure of the 25 kDa subunit of human cleavage factor I m. Nucleic Acids Res. 36, 3474–3483. https://doi.org/10.1093/nar/gkn079
Cramer, P., Bushnell, D.A., Kornberg, R.D., 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863–1876. https://doi.org/10.1126/science.1059493
Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A., Moreira, A., 2014. Implications of polyadenylation in health and disease. Nucl. Austin Tex 5, 508–519. https://doi.org/10.4161/nucl.36360
Dalziel, M., Kolesnichenko, M., das Neves, R.P., Iborra, F., Goding, C., Furger, A., 2011. Alpha-MSH regulates intergenic splicing of MC1R and TUBB3 in human melanocytes. Nucleic Acids Res. 39, 2378–2392. https://doi.org/10.1093/nar/gkq1125
Dalziel, M., Nunes, N.M., Furger, A., 2007. Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3’ end processing. Mol. Cell. Biol. 27, 1568–1580. https://doi.org/10.1128/MCB.01821-06
Danckwardt, S., Gehring, N.H., Neu-Yilik, G., Hundsdoerfer, P., Pforsich, M., Frede, U., Hentze, M.W., Kulozik, A.E., 2004. The prothrombin 3’end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations. Blood 104, 428–435. https://doi.org/10.1182/blood-2003-08-2894

Das, K., Ma, L.-C., Xiao, R., Radvansky, B., Aramini, J., Zhao, L., Marklund, J., Kuo, R.-L., Twu, K.Y., Arnold, E., Krug, R.M., Montelione, G.T., 2008. Structural basis for suppression of a host antiviral response by influenza A virus. Proc. Natl. Acad. Sci. 105, 13093–13098. https://doi.org/10.1073/pnas.0805213105

Davidson, L., Muniz, L., West, S., 2014. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. Genes Dev. 28, 342–356. https://doi.org/10.1101/gad.231274.113

de Klerk, E., Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O., Trollet, C., Dickson, G., den Dunnen, J.T., van der Maarel, S.M., Raz, V., 't Hoen, P.A.C., 2012. Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res. 40, 9089–9101. https://doi.org/10.1093/nar/gks655

de Vries, H., Rüegsegger, U., Hübner, W., Friedlein, A., Langen, H., Keller, W., 2000. Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. 19, 5895–5904. https://doi.org/10.1093/emboj/19.21.5895

Derti, A., Garrett-Engele, P., MacIsaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., Babak, T., 2012. A quantitative atlas of polyadenylation in five mammals. Genome Res. 22, 1173–1183. https://doi.org/10.1101/gr.132563.111

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., Eick, D., Andrau, J.-C., 2014. Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife 3, e02105. https://doi.org/10.7554/eLife.02105

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., Barabino, S.M.L., 2004. Distinct Sequence Motifs within the 68-kDa Subunit of Cleavage Factor Im Mediate RNA Binding, Protein-Protein Interactions, and Subcellular Localization. J. Biol. Chem. 279, 35788–35797. https://doi.org/10.1074/jbc.M403927200

Di Giammartino, D.C., Li, W., Ogami, K., Yashinskie, J.J., Hoque, M., Tian, B., Manley, J.L., 2014. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3′ UTRs. Genes Dev. 28, 2248–2260. https://doi.org/10.1101/gad.245787.114

Di Giammartino, D.C., Nishida, K., Manley, J.L., 2011. Mechanisms and consequences of alternative polyadenylation. Mol. Cell 43, 853–866. https://doi.org/10.1016/j.molcel.2011.08.017

Dominski, Z., Carpousis, A.J., Clouet-d'Orval, B., 2013. Emergence of the β-CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. Biochim. Biophys. Acta 1829, 532–551. https://doi.org/10.1016/j.bbagrm.2013.01.010

Dominski, Z., Marzluff, W.F., 2007. Formation of the 3' end of histone mRNA: getting closer to the end. Gene 396, 373–390. https://doi.org/10.1016/j.gene.2007.04.021

Dominski, Z., Yang, X., Marzluff, W.F., 2005. The Polyadenylation Factor CPSF-73 Is Involved in Histone-Pre-mRNA Processing. Cell 123, 37–48. https://doi.org/10.1016/j.cell.2005.08.002

Dominski, Z., Zheng, L.X., Sanchez, R., Marzluff, W.F., 1999. Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. Mol. Cell. Biol. 19, 3561–3570. https://doi.org/10.1128/MCB.19.5.3561

Dower, K., Kuperwasser, N., Merrikh, H., Rosbash, M., 2004. A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. RNA N. Y. N 10, 1888–1899. https://doi.org/10.1261/rna.7166704

Drexler, H.L., Choquet, K., Churchman, L.S., 2020. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell 77, 985-998.e8. https://doi.org/10.1016/j.molcel.2019.11.017

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., West, S., 2018. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev. 32, 127–139. https://doi.org/10.1101/gad.308528.117

Eaton, J.D., West, S., 2020. Termination of Transcription by RNA Polymerase II: BOOM! Trends Genet. TIG 36, 664–675. https://doi.org/10.1016/j.tig.2020.05.008

Eckmann, C.R., Rammelt, C., Wahle, E., 2011. Control of poly(A) tail length: Control of poly(A) tail length. Wiley Interdiscip. Rev. RNA 2, 348–361. https://doi.org/10.1002/wrna.56

Edmonds, M., 1990. Polyadenylate polymerases. Methods Enzymol. 181, 161–170. https://doi.org/10.1016/0076-6879(90)81118-e

Edwalds-Gilbert, G., Veraldi, K.L., Milcarek, C., 1997. Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic Acids Res. 25, 2547–2561. https://doi.org/10.1093/nar/25.13.2547

Eick, D., Geyer, M., 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. Chem. Rev. 113, 8456–8490. https://doi.org/10.1021/cr400071f

Elkon, R., Drost, J., van Haaften, G., Jenal, M., Schrier, M., Oude Vrielink, J.A.F., Agami, R., 2012. E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biol. 13, R59. https://doi.org/10.1186/gb-2012-13-7-r59

Elkon, R., Ugalde, A.P., Agami, R., 2013. Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. 14, 496–506. https://doi.org/10.1038/nrg3482

Elrod, N.D., Henriques, T., Huang, K.-L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., Adelman, K., 2019. The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. Mol. Cell 76, 738-752.e7. https://doi.org/10.1016/j.molcel.2019.10.034

Ezeokonkwo, C., Zhelkovsky, A., Lee, R., Bohm, A., Moore, C.L., 2011. A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. RNA 17, 652–664. https://doi.org/10.1261/rna.2273111

Ferraresi, P., Marchetti, G., Legnani, C., Cavallari, E., Castoldi, E., Mascoli, F., Ardissino, D., Palareti, G., Bernardi, F., 1997. The heterozygous 20210 G/A prothrombin genotype is associated with early venous thrombosis in inherited thrombophilias and is not increased in frequency in artery disease. Arterioscler. Thromb. Vasc. Biol. 17, 2418–2422. https://doi.org/10.1161/01.atv.17.11.2418

Fianu, I., Chen, Y., Dienemann, C., Dybkov, O., Linden, A., Urlaub, H., Cramer, P., 2021. Structural basis of Integrator-mediated transcription regulation. Science 374, 883–887. https://doi.org/10.1126/science.abk0154

Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., Piccirilli, J.A., 2013. RNA catalyses nuclear pre-mRNA splicing. Nature 503, 229–234. https://doi.org/10.1038/nature12734

Fitzgerald, M., Shenk, T., 1981. The sequence 5'-AAUAAA-3'forms parts of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24, 251–260. https://doi.org/10.1016/0092-8674(81)90521-3

Furger, A., Schaller, A., Schümperli, D., 1998. Functional importance of conserved nucleotides at the histone RNA 3' processing site. RNA 4, 246–256.
Furuichi, Y., Shatkin, A.J., 2000. Viral and cellular mRNA capping: past and prospects. Adv. Virus Res. 55, 135–184. https://doi.org/10.1016/s0065-3527(00)55003-9
Galli, G., Hofstetter, H., Stunnenberg, H.G., Birnstiel, M.L., 1983. Biochemical complementation with RNA in the Xenopus oocyte: a small RNA is required for the generation of 3' histone mRNA termini. Cell 34, 823–828. https://doi.org/10.1016/0092-8674(83)90539-1
Gehring, N.H., Frede, U., Neu-Yilik, G., Hundsdoerfer, P., Vetter, B., Hentze, M.W., Kulozik, A.E., 2001. Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. Nat. Genet. 28, 389–392. https://doi.org/10.1038/ng578
Ghazy, M.A., He, X., Singh, B.N., Hampsey, M., Moore, C., 2009. The Essential N Terminus of the Pta1 Scaffold Protein Is Required for snoRNA Transcription Termination and Ssu72 Function but Is Dispensable for Pre-mRNA 3′-End Processing. Mol. Cell. Biol. 29, 2296–2307. https://doi.org/10.1128/MCB.01514-08
Ghosh, A., Lima, C.D., 2010. Enzymology of RNA cap synthesis. Wiley Interdiscip. Rev. RNA 1, 152–172. https://doi.org/10.1002/wrna.19
Ghosh, A., Shuman, S., Lima, C.D., 2011. Structural insights to how mammalian capping enzyme reads the CTD code. Mol. Cell 43, 299–310. https://doi.org/10.1016/j.molcel.2011.06.001
Gick, O., Krämer, A., Vasserot, A., Birnstiel, M.L., 1987. Heat-labile regulatory factor is required for 3' processing of histone precursor mRNAs. Proc. Natl. Acad. Sci. U. S. A. 84, 8937–8940. https://doi.org/10.1073/pnas.84.24.8937
Gil, A., Proudfoot, N.J., 1987. Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. Cell 49, 399–406. https://doi.org/10.1016/0092-8674(87)90292-3
Gilmartin, G.M., Nevins, J.R., 1991. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol. Cell. Biol. 11, 2432–2438. https://doi.org/10.1128/mcb.11.5.2432-2438.1991
Gilmartin, G.M., Nevins, J.R., 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. Genes Dev. 3, 2180–2190. https://doi.org/10.1101/gad.3.12b.2180
Girbig, M., Misiaszek, A.D., Müller, C.W., 2022. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. Nat. Rev. Mol. Cell Biol. https://doi.org/10.1038/s41580-022-00476-9
Glover-Cutter, K., Kim, S., Espinosa, J., Bentley, D.L., 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. 15, 71–78. https://doi.org/10.1038/nsmb1352
Graham, R.R., Kyogoku, C., Sigurdsson, S., Vlasova, I.A., Davies, L.R.L., Baechler, E.C., Plenge, R.M., Koeuth, T., Ortmann, W.A., Hom, G., Bauer, J.W., Gillett, C., Burtt, N., Cunninghame Graham, D.S., Onofrio, R., Petri, M., Gunnarsson, I., Svenungsson, E., Rönnblom, L., Nordmark, G., Gregersen, P.K., Moser, K., Gaffney, P.M., Criswell, L.A., Vyse, T.J., Syvänen, A.-C., Bohjanen, P.R., Daly, M.J., Behrens, T.W., Altshuler, D., 2007. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc. Natl. Acad. Sci. U. S. A. 104, 6758–6763. https://doi.org/10.1073/pnas.0701266104

Gruber, A.J., Schmidt, R., Ghosh, S., Martin, G., Gruber, A.R., van Nimwegen, E., Zavolan, M., 2018. Discovery of physiological and cancer-related regulators of 3' UTR processing with KAPAC. Genome Biol. 19, 44. https://doi.org/10.1186/s13059-018-1415-3

Gruber, A.J., Schmidt, R., Gruber, A.R., Martin, G., Ghosh, S., Belmadani, M., Keller, W., Zavolan, M., 2016. A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. Genome Res. 26, 1145–1159. https://doi.org/10.1101/gr.202432.115

Gruber, A.J., Zavolan, M., 2019. Alternative cleavage and polyadenylation in health and disease. Nat. Rev. Genet. 20, 599–614. https://doi.org/10.1038/s41576-019-0145-z

Gruber, A.R., Martin, G., Keller, W., Zavolan, M., 2014. Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors: Alternative polyadenylation of messenger RNA precursors. Wiley Interdiscip. Rev. RNA 5, 183–196. https://doi.org/10.1002/wrna.1206

Gruber, A.R., Martin, G., Keller, W., Zavolan, M., 2012. Cleavage factor Im is a key regulator of 3′ UTR length. RNA Biol. 9, 1405–1412. https://doi.org/10.4161/rna.22570

Guéguéniat, J., Dupin, A.F., Stojko, J., Beaurepaire, L., Cianférani, S., Mackereth, C.D., Minvielle-Sébastia, L., Fribourg, S., 2017. Distinct roles of Pcf11 zinc-binding domains in pre-mRNA 3′-end processing. Nucleic Acids Res. 45, 10115–10131. https://doi.org/10.1093/nar/gkx674

Guo, A., Gu, H., Zhou, J., Mulhern, D., Wang, Y., Lee, K.A., Yang, V., Aguiar, M., Kornhauser, J., Jia, X., Ren, J., Beausoleil, S.A., Silva, J.C., Vemulapalli, V., Bedford, M.T., Comb, M.J., 2014. Immunoaffinity enrichment and mass spectrometry analysis of protein methylation. Mol. Cell. Proteomics MCP 13, 372–387. https://doi.org/10.1074/mcp.O113.027870

Hamilton, K., Sun, Y., Tong, L., 2019. Biophysical characterizations of the recognition of the AAUAAA polyadenylation signal. RNA 9.

Hamilton, K., Tong, L., 2020. Molecular mechanism for the interaction between human CPSF30 and hFip1. Genes Dev. 34, 1753–1761. https://doi.org/10.1101/gad.343814.120

Han, B.-G., Watson, Z., Kang, H., Pulk, A., Downing, K.H., Cate, J., Glaeser, R.M., 2016. Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. J. Struct. Biol. 195, 238–244. https://doi.org/10.1016/j.jsb.2016.06.009

Hanada, T., Weitzer, S., Mair, B., Bernreuther, C., Wainger, B.J., Ichida, J., Hanada, R., Orthofer, M., Cronin, S.J., Komnenovic, V., Minis, A., Sato, F., Mimata, H., Yoshimura, A., Tamir, I., Rainer, J., Kofler, R., Yaron, A., Eggan, K.C., Woolf, C.J., Glatzel, M., Herbst, R., Martinez, J., Penninger, J.M., 2013. CLP1 links tRNA metabolism to progressive motor-neuron loss. Nature 495, 474–480. https://doi.org/10.1038/nature11923

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., Churchman, L.S., 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell Rep. 15, 2147–2158. https://doi.org/10.1016/j.celrep.2016.05.010

Harris, M.E., Böhni, R., Schneiderman, M.H., Ramamurthy, L., Schümperli, D., Marzluff, W.F., 1991. Regulation of histone mRNA in the unperturbed cell cycle:

evidence suggesting control at two posttranscriptional steps. Mol. Cell. Biol. 11, 2416–2424. https://doi.org/10.1128/mcb.11.5.2416-2424.1991

He, X., Khan, A.U., Cheng, H., Pappas Jr., D.L., Hampsey, M., Moore, C.L., 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. Genes Dev. 17, 1030–1042. https://doi.org/10.1101/gad.1075203

Hellquist, A., Zucchelli, M., Kivinen, K., Saarialho-Kere, U., Koskenmies, S., Widen, E., Julkunen, H., Wong, A., Karjalainen-Lindsberg, M.-L., Skoog, T., Vendelin, J., Cunninghame-Graham, D.S., Vyse, T.J., Kere, J., Lindgren, C.M., 2007. The human GIMAP5 gene has a common polyadenylation polymorphism increasing risk to systemic lupus erythematosus. J. Med. Genet. 44, 314–321. https://doi.org/10.1136/jmg.2006.046185

Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J., Proudfoot, N.J., 1983. α-Thalassaemia caused by a polyadenylation signal mutation. Nature 3.

Hill, C.H., Boreikaitė, V., Kumar, A., Casañal, A., Kubík, P., Degliesposti, G., Maslen, S., Mariani, A., von Loeffelholz, O., Girbig, M., Skehel, M., Passmore, L.A., 2019. Activation of the Endonuclease that Defines mRNA 3′ Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. Mol. Cell 73, 1217-1231.e11. https://doi.org/10.1016/j.molcel.2018.12.023

Hino, H., Araki, K., Uyama, E., Takeya, M., Araki, M., Yoshinobu, K., Miike, K., Kawazoe, Y., Maeda, Y., Uchino, M., Yamamura, K., 2004. Myopathy phenotype in transgenic mice expressing mutated PABPN1 as a model of oculopharyngeal muscular dystrophy. Hum. Mol. Genet. 13, 181–190. https://doi.org/10.1093/hmg/ddh017

Hirose, Y., Manley, J.L., 1997. Creatine Phosphate, Not ATP, Is Required for 3′ End Cleavage of Mammalian Pre-mRNA in Vitro. J. Biol. Chem. 272, 29636–29642. https://doi.org/10.1074/jbc.272.47.29636

Ho, C.K., 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. https://doi.org/10.1016/s1097-2765(00)80468-2

Hosoda, N., Lejeune, F., Maquat, L.E., 2006. Evidence that Poly(A) Binding Protein C1 Binds Nuclear Pre-mRNA Poly(A) Tails. Mol. Cell. Biol. 26, 3085–3097. https://doi.org/10.1128/MCB.26.8.3085-3097.2006

Hosp, F., Scheltema, R.A., Eberl, H.C., Kulak, N.A., Keilhauer, E.C., Mayr, K., Mann, M., 2015. A Double-Barrel Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) System to Quantify 96 Interactomes per Day. Mol. Cell. Proteomics MCP 14, 2030–2041. https://doi.org/10.1074/mcp.O115.049460

Hsin, J.-P., Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 26, 2119–2137. https://doi.org/10.1101/gad.200303.112

Hu, J., Lutz, C.S., Wilusz, J., Tian, B., 2005. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA 11, 1485–1493. https://doi.org/10.1261/rna.2107305

Huber, Z., Monarez, R.R., Dass, B., MacDonald, C.C., 2005. The mRNA encoding tauCstF-64 is expressed ubiquitously in mouse tissues. Ann. N. Y. Acad. Sci. 1061, 163–172. https://doi.org/10.1196/annals.1336.017

Hutchins, L.N., Murphy, S.M., Singh, P., Graber, J.H., 2008. Position-dependent motif characterization using non-negative matrix factorization. Bioinformatics 24, 2684–2690. https://doi.org/10.1093/bioinformatics/btn526

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Vrielink, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K., Rubinsztein, D.C., Agami, R., 2012. The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. Cell 149, 538–553. https://doi.org/10.1016/j.cell.2012.03.022

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., Dreyfuss, G., 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468, 664–668. https://doi.org/10.1038/nature09479

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wiśniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., Proudfoot, N.J., 2019. Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Mol. Cell 74, 158-172.e9. https://doi.org/10.1016/j.molcel.2019.01.027

Kandala, D.T., Mohan, N., A, V., Ap, S., G, R., Laishram, R.S., 2016. CstF-64 and 3′-UTR cis-element determine Star-PAP specificity for target mRNA selection by excluding PAPα. Nucleic Acids Res. 44, 811–823. https://doi.org/10.1093/nar/gkv1074

Kastner, B., Will, C.L., Stark, H., Lührmann, R., 2019. Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. Cold Spring Harb. Perspect. Biol. 11, a032417. https://doi.org/10.1101/cshperspect.a032417

Kaufmann, I., Martin, G., Friedlein, A., Langen, H., Keller, W., 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J. 23, 616–626. https://doi.org/10.1038/sj.emboj.7600070

Keller, R.W., Kühn, U., Aragón, M., Bornikova, L., Wahle, E., Bear, D.G., 2000. The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J. Mol. Biol. 297, 569–583. https://doi.org/10.1006/jmbi.2000.3572

Kennedy, S.A., Frazier, M.L., Steiniger, M., Mast, A.M., Marzluff, W.F., Redinbo, M.R., 2009. Crystal Structure of the HEAT Domain from the Pre-mRNA Processing Factor Symplekin. J. Mol. Biol. 392, 115–128. https://doi.org/10.1016/j.jmb.2009.06.062

Keon, B.H., Schäfer, S., Kuhn, C., Grund, C., Franke, W.W., 1996. Symplekin, a novel type of tight junction plaque protein. J. Cell Biol. 134, 1003–1018. https://doi.org/10.1083/jcb.134.4.1003

Kerwitz, Y., Kühn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E., Wahle, E., 2003. Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J. 22, 3705–3714. https://doi.org/10.1093/emboj/cdg347

Khatter, H., Vorländer, M.K., Müller, C.W., 2017. RNA polymerase I and III: similar yet unique. Curr. Opin. Struct. Biol. 47, 88–94. https://doi.org/10.1016/j.sbi.2017.05.008

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., Buratowski, S., 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432, 517–522. https://doi.org/10.1038/nature03041

Kim, S., Yamamoto, J., Chen, Y., Aida, M., Wada, T., Handa, H., Yamaguchi, Y., 2010. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation: Role of CFIm in alternative polyadenylation. Genes Cells 15, 1003–1013. https://doi.org/10.1111/j.1365-2443.2010.01436.x

Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Lubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N., Kashlev, M., 2008. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. Mol. Cell 30, 557–566. https://doi.org/10.1016/j.molcel.2008.04.017

Klykov, O., Steigenberger, B., Pektaş, S., Fasci, D., Heck, A.J.R., Scheltema, R.A., 2018. Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. Nat. Protoc. 13, 2964–2990. https://doi.org/10.1038/s41596-018-0074-x

Kolev, N.G., Steitz, J.A., 2005. Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. Genes Dev. 19, 2583–2592. https://doi.org/10.1101/gad.1371105

Kolev, N.G., Yario, T.A., Benson, E., Steitz, J.A., 2008. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3′-end maturation. EMBO Rep. 9, 1013–1018. https://doi.org/10.1038/embor.2008.146

Komarnitsky, P., Cho, E.J., Buratowski, S., 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. 14, 2452–2460. https://doi.org/10.1101/gad.824700

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A., Handa, H., 2006. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3′-UTRs. Nucleic Acids Res. 34, 6264–6271. https://doi.org/10.1093/nar/gkl794

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., Wahle, E., 2009. Poly(A) Tail Length Is Controlled by the Nuclear Poly(A)-binding Protein Regulating the Interaction between Poly(A) Polymerase and the Cleavage and Polyadenylation Specificity Factor. J. Biol. Chem. 284, 22803–22814. https://doi.org/10.1074/jbc.M109.018226

Kühn, U., Nemeth, A., Meyer, S., Wahle, E., 2003. The RNA Binding Domains of the Nuclear poly(A)-binding Protein. J. Biol. Chem. 278, 16916–16925. https://doi.org/10.1074/jbc.M209886200

Kühn, U., Wahle, E., 2004. Structure and function of poly(A) binding proteins. Biochim. Biophys. Acta BBA - Gene Struct. Expr. 1678, 67–84. https://doi.org/10.1016/j.bbaexp.2004.03.008

Kulak, N.A., Geyer, P.E., Mann, M., 2017. Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. Mol. Cell. Proteomics 16, 694–705. https://doi.org/10.1074/mcp.O116.065136

Kumar, A., Clerici, M., Muckenfuss, L.M., Passmore, L.A., Jinek, M., 2019. Mechanistic insights into mRNA 3′-end processing. Curr. Opin. Struct. Biol. 59, 143–150. https://doi.org/10.1016/j.sbi.2019.08.001

Kumar, A., Yu, C.W.H., Rodríguez-Molina, J.B., Li, X.-H., Freund, S.M.V., Passmore, L.A., 2021. Dynamics in Fip1 regulate eukaryotic mRNA 3′ end processing. Genes Dev. genesdev;gad.348671.121v1. https://doi.org/10.1101/gad.348671.121

Lackford, B., Yao, C., Charles, G.M., Weng, L., Zheng, X., Choi, E.-A., Xie, X., Wan, J., Xing, Y., Freudenberg, J.M., Yang, P., Jothi, R., Hu, G., Shi, Y., 2014. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. EMBO J. 33, 878–889. https://doi.org/10.1002/embj.201386537

Lee, S.D., Moore, C.L., 2014. Efficient mRNA Polyadenylation Requires a Ubiquitin-Like Domain, a Zinc Knuckle, and a RING Finger Domain, All Contained in the Mpe1 Protein. Mol. Cell. Biol. 34, 3955–3967. https://doi.org/10.1128/MCB.00077-14

Lee, S.-H., Singh, I., Tisdale, S., Abdel-Wahab, O., Leslie, C.S., Mayr, C., 2018. Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia. Nature 561, 127–131. https://doi.org/10.1038/s41586-018-0465-8

Legendre, M., Gautheret, D., 2003. Sequence determinants in human polyadenylation site selection. BMC Genomics 4, 7. https://doi.org/10.1186/1471-2164-4-7

Legnini, I., Alles, J., Karaiskos, N., Ayoub, S., Rajewsky, N., 2019. FLAM-seq: full-length mRNA sequencing reveals principles of poly(A) tail length control. Nat. Methods 16, 879–886. https://doi.org/10.1038/s41592-019-0503-y

Legrand, P., Pinaud, N., Minvielle-Sébastia, L., Fribourg, S., 2007. The structure of the CstF-77 homodimer provides insights into CstF assembly. Nucleic Acids Res. 35, 4515–4522. https://doi.org/10.1093/nar/gkm458

Levitt, N., Briggs, D., Gil, A., Proudfoot, N.J., 1989. Definition of an efficient synthetic poly(A) site. Genes Dev. 3, 1019–1025. https://doi.org/10.1101/gad.3.7.1019

Li, L., Deng, B., Xing, G., Teng, Y., Tian, C., Cheng, X., Yin, X., Yang, J., Gao, X., Zhu, Y., Sun, Q., Zhang, L., Yang, X., He, F., 2007. PACT is a negative regulator of p53 and essential for cell growth and embryonic development. Proc. Natl. Acad. Sci. U. S. A. 104, 7951–7956. https://doi.org/10.1073/pnas.0701916104

Li, W., You, B., Hoque, M., Zheng, D., Luo, W., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L., Tian, B., 2015. Systematic Profiling of Poly(A)+ Transcripts Modulated by Core 3' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. PLOS Genet. 11, e1005166. https://doi.org/10.1371/journal.pgen.1005166

Li, Z., Michael, I.P., Zhou, D., Nagy, A., Rini, J.M., 2013. Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. Proc. Natl. Acad. Sci. 110, 5004–5009. https://doi.org/10.1073/pnas.1218620110

Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., Mayr, C., 2013. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev. 27, 2380–2396. https://doi.org/10.1101/gad.229328.113

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., Bentley, D.L., 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol. Cell 9, 1101–1111. https://doi.org/10.1016/s1097-2765(02)00518-x

Lidschreiber, M., Leike, K., Cramer, P., 2013. Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol. Cell. Biol. 33, 3805–3816. https://doi.org/10.1128/MCB.00361-13

Lim, J., Kim, D., Lee, Y., Ha, M., Lee, M., Yeo, J., Chang, H., Song, J., Ahn, K., Kim, V.N., 2018. Mixed tailing by TENT4A and TENT4B shields mRNA from rapid deadenylation. Science 361, 701–704. https://doi.org/10.1126/science.aam5794

Logan, J., Falck-Pedersen, E., Darnell, J.E., Shenk, T., 1987. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc. Natl. Acad. Sci. U. S. A. 84, 8306–8310. https://doi.org/10.1073/pnas.84.23.8306

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., Varani, G., 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. 17, 1195–1201. https://doi.org/10.1038/nsmb.1893

Lykke-Andersen, S., Žumer, K., Molska, E.Š., Rouvière, J.O., Wu, G., Demel, C., Schwalb, B., Schmid, M., Cramer, P., Jensen, T.H., 2021. Integrator is a genome-wide attenuator of non-productive transcription. Mol. Cell 81, 514-529.e6. https://doi.org/10.1016/j.molcel.2020.12.014

MacDonald, C.C., 2019. Tissue-specific mechanisms of alternative polyadenylation: Testis, brain, and beyond (2018 update). WIREs RNA 10. https://doi.org/10.1002/wrna.1526

MacDonald, C.C., Wilusz, J., Shenk, T., 1994. The 64-Kilodalton Subunit of the CstF Polyadenylation Factor Binds to Pre-mRNAs Downstream of the Cleavage Site and Influences Cleavage Site Location. Mol. Cell. Biol. 14, 8.

Mandel, C.R., Gebauer, D., Zhang, H., Tong, L., 2006a. A serendipitous discovery that in situ proteolysis is essential for the crystallization of yeast CPSF-100 (Ydh1p). Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 62, 1041–1045. https://doi.org/10.1107/S1744309106038152

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., Tong, L., 2006b. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature 444, 953–956. https://doi.org/10.1038/nature05363

Martin, G., Gruber, A.R., Keller, W., Zavolan, M., 2012. Genome-wide Analysis of Pre-mRNA 3′ End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3′ UTR Length. Cell Rep. 1, 753–763. https://doi.org/10.1016/j.celrep.2012.05.003

Martin, G., Jenö, P., Keller, W., 1999. Mapping of ATP binding regions in poly(A) polymerases by photoaffinity labeling and by mutational analysis identifies a domain conserved in many nucleotidyltransferases. Protein Sci. Publ. Protein Soc. 8, 2380–2391. https://doi.org/10.1110/ps.8.11.2380

Martin, G., Keller, W., 1996. Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. EMBO J. 15, 2593–2603. https://doi.org/10.1002/j.1460-2075.1996.tb00617.x

Martin, G., Keller, W., Doublié, S., 2000. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. EMBO J. 19, 4193–4203. https://doi.org/10.1093/emboj/19.16.4193

Martin, G., Möglich, A., Keller, W., Doublié, S., 2004. Biochemical and Structural Insights into Substrate Binding and Catalytic Mechanism of Mammalian Poly(A) Polymerase. J. Mol. Biol. 341, 911–925. https://doi.org/10.1016/j.jmb.2004.06.047
Martinez-Rucobo, F.W., Kohler, R., van de Waterbeemd, M., Heck, A.J.R., Hemann, M., Herzog, F., Stark, H., Cramer, P., 2015. Molecular Basis of Transcription-Coupled Pre-mRNA Capping. Mol. Cell 58, 1079–1089. https://doi.org/10.1016/j.molcel.2015.04.004
Marzluff, W.F., 2005. Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. Curr. Opin. Cell Biol. 17, 274–280. https://doi.org/10.1016/j.ceb.2005.04.010
Marzluff, W.F., Duronio, R.J., 2002. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. Curr. Opin. Cell Biol. 14, 692–699. https://doi.org/10.1016/s0955-0674(02)00387-3
Masamha, C.P., 2022. The emerging roles of CFIm25 (NUDT21/CPSF5) in human biology and disease. Wiley Interdiscip. Rev. RNA e1757. https://doi.org/10.1002/wrna.1757
Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.-B., Li, W., Wagner, E.J., 2014. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature 510, 412–416. https://doi.org/10.1038/nature13261
Mayr, C., 2017. Regulation by 3′-Untranslated Regions. Annu. Rev. Genet. 51, 171–194. https://doi.org/10.1146/annurev-genet-120116-024704
Mayr, C., 2016. Evolution and Biological Roles of Alternative 3'UTRs. Trends Cell Biol. 26, 227–237. https://doi.org/10.1016/j.tcb.2015.10.012
Mbita, Z., Meyer, M., Skepu, A., Hosie, M., Rees, J., Dlamini, Z., 2012. De-regulation of the RBBP6 isoform 3/DWNN in human cancers. Mol. Cell. Biochem. 362, 249–262. https://doi.org/10.1007/s11010-011-1150-5
McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., Bentley, D.L., 1997a. 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev. 11, 3306–3318. https://doi.org/10.1101/gad.11.24.3306
McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L., 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385, 357–361. https://doi.org/10.1038/385357a0
McDevitt, M.A., Hart, R.P., Wong, W.W., Nevins, J.R., 1986. Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO J. 5, 2907–2913. https://doi.org/10.1002/j.1460-2075.1986.tb04586.x
McLennan, A.G., 2006. The Nudix hydrolase superfamily. Cell. Mol. Life Sci. CMLS 63, 123–143. https://doi.org/10.1007/s00018-005-5386-7
Meinhart, A., Cramer, P., 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature 430, 223–226. https://doi.org/10.1038/nature02679
Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., Cramer, P., 2005. A structural perspective of CTD function. Genes Dev. 19, 1401–1415. https://doi.org/10.1101/gad.1318105
Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., Jensen, T.H., 2016. Identification of a Nuclear Exosome Decay Pathway for Processed

Transcripts. Mol. Cell 64, 520–533. https://doi.org/10.1016/j.molcel.2016.09.025
Meyer, S., Urbanke, C., Wahle, E., 2002. Equilibrium Studies on the Association of the Nuclear Poly(A) Binding Protein with Poly(A) of Different Lengths. Biochemistry 41, 6082–6089. https://doi.org/10.1021/bi0160866
Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., Vagner, S., 2006. An interaction between U2AF 65 and CF Im links the splicing and 3′ end processing machineries. EMBO J. 25, 4854–4864. https://doi.org/10.1038/sj.emboj.7601331
Miotto, B., Chibi, M., Xie, P., Koundrioukoff, S., Moolman-Smook, H., Pugh, D., Debatisse, M., He, F., Zhang, L., Defossez, P.-A., 2014. The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. Cell Rep. 7, 575–587. https://doi.org/10.1016/j.celrep.2014.03.030
Mitchelson, A., Simonelig, M., Williams, C., O'Hare, K., 1993. Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppression in Drosophila melanogaster by suppressor of forked occurs at the level of RNA stability. Genes Dev. 7, 241–249. https://doi.org/10.1101/gad.7.2.241
Mizrahi, N., Moore, C., 2000. Posttranslational phosphorylation and ubiquitination of the Saccharomyces cerevisiae Poly(A) polymerase at the S/G(2) stage of the cell cycle. Mol. Cell. Biol. 20, 2794–2802. https://doi.org/10.1128/MCB.20.8.2794-2802.2000
Mohanan, N.K., Shaji, F., Koshre, G.R., Laishram, R.S., 2022. Alternative polyadenylation: An enigma of transcript length variation in health and disease. Wiley Interdiscip. Rev. RNA 13, e1692. https://doi.org/10.1002/wrna.1692
Montell, C., Fisher, E.F., Caruthers, M.H., Berk, A.J., 1983. Inhibition of RNA cleavage but not polyadenylation by a point mutation in mRNA 3' consensus sequence AAUAAA. Nature 305, 600–605. https://doi.org/10.1038/305600a0
Moore, C.L., Sharp, P.A., 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. Cell 41, 845–855. https://doi.org/10.1016/S0092-8674(85)80065-9
Moore, M.J., Query, C.C., 2000. Joining of RNAs by splinted ligation, in: Methods in Enzymology. Elsevier, pp. 109–123. https://doi.org/10.1016/S0076-6879(00)17009-0
Moreno-Morcillo, M., Minvielle-Sébastia, L., Mackereth, C., Fribourg, S., 2011. Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure. RNA N. Y. N 17, 412–418. https://doi.org/10.1261/rna.2481011
Moteki, S., Price, D., 2002. Functional coupling of capping and transcription of mRNA. Mol. Cell 10, 599–609. https://doi.org/10.1016/s1097-2765(02)00660-3
Mowry, K.L., Oh, R., Steitz, J.A., 1989. Each of the conserved sequence elements flanking the cleavage site of mammalian histone pre-mRNAs has a distinct role in the 3'-end processing reaction. Mol. Cell. Biol. 9, 3105–3108. https://doi.org/10.1128/mcb.9.7.3105-3108.1989
Mowry, K.L., Steitz, J.A., 1987. Both conserved signals on mammalian histone pre-mRNAs associate with small nuclear ribonucleoproteins during 3' end formation in vitro. Mol. Cell. Biol. 7, 1663–1672. https://doi.org/10.1128/mcb.7.5.1663-1672.1987
Muckenfuss, L.M., Migenda Herranz, A.C., Boneberg, F.M., Clerici, M., Jinek, M., 2022. Fip1 is a multivalent interaction scaffold for processing factors in human mRNA 3' end biogenesis. eLife 11, e80332. https://doi.org/10.7554/eLife.80332

Murthy, K.G., Manley, J.L., 1995. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev. 9, 2672–2683. https://doi.org/10.1101/gad.9.21.2672

NandyMazumdar, M., Artsimovitch, I., 2015. Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - Shifting shapes and paradigms. BioEssays News Rev. Mol. Cell. Dev. Biol. 37, 324–334. https://doi.org/10.1002/bies.201400177

Neer, E.J., Schmidt, C.J., Nambudripad, R., Smith, T.F., 1994. The ancient regulatory-protein family of WD-repeat proteins. Nature 371, 297–300. https://doi.org/10.1038/371297a0

Nemeth, A., Krause, S., Blank, D., Jenny, A., Jenö, P., Lustig, A., Wahle, E., 1995. Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II. Nucleic Acids Res. 23, 4034–4041. https://doi.org/10.1093/nar/23.20.4034

Neugebauer, K.M., 2019. Nascent RNA and the Coordination of Splicing with Transcription. Cold Spring Harb. Perspect. Biol. 11, a032227. https://doi.org/10.1101/cshperspect.a032227

Neve, J., Patel, R., Wang, Z., Louey, A., Furger, A.M., 2017. Cleavage and polyadenylation: Ending the message expands gene regulation. RNA Biol. 14, 865–890. https://doi.org/10.1080/15476286.2017.1306171

Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.-R., Lis, J.T., 2004. Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. Mol. Cell 13, 55–65. https://doi.org/10.1016/s1097-2765(03)00526-4

Nilsen, T.W., Graveley, B.R., 2010. Expansion of the eukaryotic proteome by alternative splicing. Nature 463, 457–463. https://doi.org/10.1038/nature08909

Noble, C.G., Beuth, B., Taylor, I.A., 2007. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res. 35, 87–99. https://doi.org/10.1093/nar/gkl1010

Noe Gonzalez, M., Sato, S., Tomomori-Sato, C., Conaway, J.W., Conaway, R.C., 2018. CTD-dependent and -independent mechanisms govern co-transcriptional capping of Pol II transcripts. Nat. Commun. 9, 3392. https://doi.org/10.1038/s41467-018-05923-w

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., Proudfoot, N.J., 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell 161, 526–540. https://doi.org/10.1016/j.cell.2015.03.027

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., Carmo-Fonseca, M., 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Mol. Cell 72, 369-379.e4. https://doi.org/10.1016/j.molcel.2018.09.004

Nudler, E., 2012. RNA polymerase backtracking in gene regulation and genome instability. Cell 149, 1438–1445. https://doi.org/10.1016/j.cell.2012.06.003

Nunes, N.M., Li, W., Tian, B., Furger, A., 2010. A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. EMBO J. 29, 1523–1536. https://doi.org/10.1038/emboj.2010.42

Oberg, D., Fay, J., Lambkin, H., Schwartz, S., 2005. A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. J. Virol. 79, 9254–9269. https://doi.org/10.1128/JVI.79.14.9254-9269.2005

Oh, J.-M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I., Dreyfuss, G., 2017. U1 snRNP telescripting regulates a size-function-stratified human genome. Nat. Struct. Mol. Biol. 24, 993–999. https://doi.org/10.1038/nsmb.3473

Ohnacker, M., Barabino, S.M., Preker, P.J., Keller, W., 2000. The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. EMBO J. 19, 37–47. https://doi.org/10.1093/emboj/19.1.37

Orkin, S.H., Cheng, T.C., Antonarakis, S.E., Kazazian, H.H., 1985. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. EMBO J. 4, 453–456. https://doi.org/10.1002/j.1460-2075.1985.tb03650.x

Osley, M.A., 1991. The regulation of histone synthesis in the cell cycle. Annu. Rev. Biochem. 60, 827–861. https://doi.org/10.1146/annurev.bi.60.070191.004143

Osman, S., Cramer, P., 2020. Structural Biology of RNA Polymerase II Transcription: 20 Years On. Annu. Rev. Cell Dev. Biol. 36, 1–34. https://doi.org/10.1146/annurev-cellbio-042020-021954

Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B., Milos, P.M., 2010. Comprehensive Polyadenylation Site Maps in Yeast and Human Reveal Pervasive Alternative Polyadenylation. Cell 143, 1018–1029. https://doi.org/10.1016/j.cell.2010.11.020

Papasaikas, P., Valcárcel, J., 2016. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. Trends Biochem. Sci. 41, 33–45. https://doi.org/10.1016/j.tibs.2015.11.003

Paushkin, S.V., Patel, M., Furia, B.S., Peltz, S.W., Trotta, C.R., 2004. Identification of a Human Endonuclease Complex Reveals a Link between tRNA Splicing and Pre-mRNA 3′ End Formation. Cell 117, 311–321. https://doi.org/10.1016/S0092-8674(04)00342-3

Pereira-Castro, I., Moreira, A., 2021. On the function and relevance of alternative 3′-UTRs in gene expression regulation. WIREs RNA 12. https://doi.org/10.1002/wrna.1653

Pérez Cañadillas, J.M., Varani, G., 2003. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. EMBO J. 22, 2821–2830. https://doi.org/10.1093/emboj/cdg259

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., Ferrin, T.E., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 30, 70–82. https://doi.org/10.1002/pro.3943

Pfleiderer, M.M., Galej, W.P., 2022. Emerging insights into the function and structure of the Integrator complex. Transcription 1–15. https://doi.org/10.1080/21541264.2022.2047583

Pfleiderer, M.M., Galej, W.P., 2021. Structure of the catalytic core of the Integrator complex. Mol. Cell 81, 1246-1259.e8. https://doi.org/10.1016/j.molcel.2021.01.005

Pillai, R.S., Grimmler, M., Meister, G., Will, C.L., Lührmann, R., Fischer, U., Schümperli, D., 2003. Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. 17, 2321–2333. https://doi.org/10.1101/gad.274403

Pillai, R.S., Will, C.L., Lührmann, R., Schümperli, D., Müller, B., 2001. Purified U7 snRNPs lack the Sm proteins D1 and D2 but contain Lsm10, a new 14 kDa Sm D1-like protein. EMBO J. 20, 5470–5479. https://doi.org/10.1093/emboj/20.19.5470

Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M.C., Caracausi, M., 2019. Human protein-coding genes and gene feature statistics in 2019. BMC Res. Notes 12, 315. https://doi.org/10.1186/s13104-019-4343-8

Plaschka, C., Newman, A.J., Nagai, K., 2019. Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harb. Perspect. Biol. 11, a032391. https://doi.org/10.1101/cshperspect.a032391

Poon, L.L., Fodor, E., Brownlee, G.G., 2000. Polyuridylated mRNA synthesized by a recombinant influenza virus is defective in nuclear export. J. Virol. 74, 418–427. https://doi.org/10.1128/jvi.74.1.418-427.2000

Poort, S.R., Rosendaal, F.R., Reitsma, P.H., Bertina, R.M., 1996. A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 88, 3698–3703.

Porrua, O., Libri, D., 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. Nat. Rev. Mol. Cell Biol. 16, 190–202. https://doi.org/10.1038/nrm3943

Proudfoot, N.J., 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science 352, aad9926. https://doi.org/10.1126/science.aad9926

Proudfoot, N.J., 1989. How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem. Sci. 14, 105–110. https://doi.org/10.1016/0968-0004(89)90132-1

Pugh, D.J.R., Ab, E., Faro, A., Lutya, P.T., Hoffmann, E., Rees, D.J.G., 2006. DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. BMC Struct. Biol. 6, 1. https://doi.org/10.1186/1472-6807-6-1

Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296. https://doi.org/10.1038/nmeth.4169

Qian, Z.W., Wilusz, J., 1991. An RNA-binding protein specifically interacts with a functionally important domain of the downstream element of the simian virus 40 late polyadenylation signal. Mol. Cell. Biol. 11, 5312–5320. https://doi.org/10.1128/mcb.11.10.5312-5320.1991

Ramanathan, A., Robb, G.B., Chan, S.-H., 2016. mRNA capping: biological functions and applications. Nucleic Acids Res. 44, 7511–7526. https://doi.org/10.1093/nar/gkw551

Ramirez, A., Shuman, S., Schwer, B., 2008. Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme in vivo. RNA N. Y. N 14, 1737–1745. https://doi.org/10.1261/rna.1142908

Ran, Y., Deng, Y., Yao, C., 2021. U1 snRNP telescripting: molecular mechanisms and beyond. RNA Biol. 1–12. https://doi.org/10.1080/15476286.2021.1872963

Rasmussen, E.B., Lis, J.T., 1993. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. U. S. A. 90, 7923–7927. https://doi.org/10.1073/pnas.90.17.7923

Ren, F., Zhang, N., Zhang, L., Miller, E., Pu, J.J., 2020. Alternative Polyadenylation: a new frontier in post transcriptional regulation. Biomark. Res. 8, 67. https://doi.org/10.1186/s40364-020-00249-6

Reyes, A., Huber, W., 2018. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res. 46, 582–592. https://doi.org/10.1093/nar/gkx1165

Ridker, P.M., Hennekens, C.H., Miletich, J.P., 1999. G20210A mutation in prothrombin gene and risk of myocardial infarction, stroke, and venous thrombosis in a large cohort of US men. Circulation 99, 999–1004. https://doi.org/10.1161/01.cir.99.8.999

Rodríguez-Molina, J.B., O'Reilly, F.J., Fagarasan, H., Sheekey, E., Maslen, S., Skehel, J.M., Rappsilber, J., Passmore, L.A., 2022. Mpe1 senses the binding of pre-mRNA and controls 3′ end processing by CPF. Mol. Cell S1097276522003847. https://doi.org/10.1016/j.molcel.2022.04.021

Romeo, V., Griesbach, E., Schümperli, D., 2014. CstF64: Cell Cycle Regulation and Functional Role in 3′ End Processing of Replication-Dependent Histone mRNAs. Mol. Cell. Biol. 34, 4272–4284. https://doi.org/10.1128/MCB.00791-14

Romeo, V., Schümperli, D., 2016. Cycling in the nucleus: regulation of RNA 3' processing and nuclear organization of replication-dependent histone genes. Curr. Opin. Cell Biol. 40, 23–31. https://doi.org/10.1016/j.ceb.2016.01.015

Ross, N.T., Lohmann, F., Carbonneau, S., Fazal, A., Weihofen, W.A., Gleim, S., Salcius, M., Sigoillot, F., Henault, M., Carl, S.H., Rodríguez-Molina, J.B., Miller, H.R., Brittain, S.M., Murphy, J., Zambrowski, M., Boynton, G., Wang, Y., Chen, A., Molind, G.J., Wilbertz, J.H., Artus-Revel, C.G., Jia, M., Akinjiyan, F.A., Turner, J., Knehr, J., Carbone, W., Schuierer, S., Reece-Hoyes, J.S., Xie, K., Saran, C., Williams, E.T., Roma, G., Spencer, M., Jenkins, J., George, E.L., Thomas, J.R., Michaud, G., Schirle, M., Tallarico, J., Passmore, L.A., Chao, J.A., Beckwith, R.E.J., 2020. CPSF3-dependent pre-mRNA processing as a druggable node in AML and Ewing's sarcoma. Nat. Chem. Biol. 16, 50–59. https://doi.org/10.1038/s41589-019-0424-1

Rüegsegger, U., Beyer, K., Keller, W., 1996. Purification and Characterization of Human Cleavage Factor Im Involved in the 3′ End Processing of Messenger RNA Precursors. J. Biol. Chem. 271, 6107–6113. https://doi.org/10.1074/jbc.271.11.6107

Rüegsegger, U., Blank, D., Keller, W., 1998. Human Pre-mRNA Cleavage Factor Im Is Related to Spliceosomal SR Proteins and Can Be Reconstituted In Vitro from Recombinant Subunits. Mol. Cell 1, 243–253. https://doi.org/10.1016/S1097-2765(00)80025-8

Ruepp, M.-D., Schümperli, D., Barabino, S.M.L., 2011a. mRNA 3′ end processing and more-multiple functions of mammalian cleavage factor I-68: mRNA 3' end processing and more. Wiley Interdiscip. Rev. RNA 2, 79–91. https://doi.org/10.1002/wrna.35

Ruepp, M.-D., Schweingruber, C., Kleinschmidt, N., Schümperli, D., 2011b. Interactions of CstF-64, CstF-77, and symplekin: Implications on localisation and function. Mol. Biol. Cell 22, 91–104. https://doi.org/10.1091/mbc.e10-06-0543

Ryan, K., Calvo, O., Manley, J.L., 2004. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA 10, 565–573. https://doi.org/10.1261/rna.5214404

Sabath, I., Skrajna, A., Yang, X. -c., Dadlez, M., Marzluff, W.F., Dominski, Z., 2013. 3'-End processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. RNA 19, 1726–1744. https://doi.org/10.1261/rna.040360.113

Sadofsky, M., Connelly, S., Manley, J.L., Alwine, J.C., 1985. Identification of a sequence element on the 3' side of AAUAAA which is necessary for simian virus 40 late mRNA 3'-end processing. Mol. Cell. Biol. 5, 2713–2719. https://doi.org/10.1128/mcb.5.10.2713-2719.1985

Sadowski, M., Dichtl, B., Hübner, W., Keller, W., 2003. Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J. 22, 2167–2177. https://doi.org/10.1093/emboj/cdg200

Saijo, M., Sakai, Y., Kishino, T., Niikawa, N., Matsuura, Y., Morino, K., Tamai, K., Taya, Y., 1995. Molecular cloning of a human protein that binds to the retinoblastoma protein and chromosomal mapping. Genomics 27, 511–519. https://doi.org/10.1006/geno.1995.1084

Sakai, Y., Saijo, M., Coelho, K., Kishino, T., Niikawa, N., Taya, Y., 1995. cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. Genomics 30, 98–101. https://doi.org/10.1006/geno.1995.0017

Salisbury, J., Hutchison, K.W., Graber, J.H., 2006. A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. BMC Genomics 7, 55. https://doi.org/10.1186/1471-2164-7-55

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., Burge, C.B., 2008. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 320, 1643–1647. https://doi.org/10.1126/science.1155390

Sartini, B.L., Wang, H., Wang, W., Millette, C.F., Kilpatrick, D.L., 2008. Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. Biol. Reprod. 78, 472–482. https://doi.org/10.1095/biolreprod.107.064774

Schäfer, I.B., Yamashita, M., Schuller, J.M., Schüssler, S., Reichelt, P., Strauss, M., Conti, E., 2019. Molecular Basis for poly(A) RNP Architecture and Recognition by the Pan2-Pan3 Deadenylase. Cell 177, 1619-1631.e21. https://doi.org/10.1016/j.cell.2019.04.013

Schäfer, P., Tüting, C., Schönemann, L., Kühn, U., Treiber, T., Treiber, N., Ihling, C., Graber, A., Keller, W., Meister, G., Sinz, A., Wahle, E., 2018. Reconstitution of mammalian cleavage factor II involved in 3′ processing of mRNA precursors. RNA 18.

Scharl, E.C., Steitz, J.A., 1994. The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. EMBO J. 13, 2432–2440. https://doi.org/10.1002/j.1460-2075.1994.tb06528.x

Schmidt, M., Kluge, F., Sandmeir, F., Kühn, U., Schäfer, P., Tüting, C., Ihling, C., Conti, E., Wahle, E., 2022. Reconstitution of 3′ end processing of mammalian pre-mRNA reveals a central role of RBBP6. Genes Dev. genesdev;gad.349217.121v1. https://doi.org/10.1101/gad.349217.121

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., Wahle, E., 2014. Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. Genes Dev. 28, 2381–2393. https://doi.org/10.1101/gad.250985.114

Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y., Mastronarde, D.N., 2019. Software tools for automated transmission electron microscopy. Nat. Methods 16, 471–477. https://doi.org/10.1038/s41592-019-0396-9

Sheets, M.D., Ogg, S.C., Wickens, M.P., 1990. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. Nucleic Acids Res. 18, 5799–5805. https://doi.org/10.1093/nar/18.19.5799

Sheets, M.D., Stephenson, P., Wickens, M.P., 1987. Products of In Vitro Cleavage and Polyadenylation of Simian Virus 40 Late Pre-mRNAs. Mol. Cell. Biol. 7, 12.

Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., Shi, Y., 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA N. Y. N 17, 761–772. https://doi.org/10.1261/rna.2581711

Shetty, A., Kallgren, S.P., Demel, C., Maier, K.C., Spatt, D., Alver, B.H., Cramer, P., Park, P.J., Winston, F., 2017. Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. Mol. Cell 66, 77-88.e5. https://doi.org/10.1016/j.molcel.2017.02.023

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., Manley, J.L., 2009. Molecular Architecture of the Human Pre-mRNA 3′ Processing Complex. Mol. Cell 33, 365–376. https://doi.org/10.1016/j.molcel.2008.12.028

Shi, Y., Manley, J.L., 2015. The end of the message: multiple protein–RNA interactions define the mRNA polyadenylation site. Genes Dev. 29, 889–897. https://doi.org/10.1101/gad.261974.115

Shimazu, T., Horinouchi, S., Yoshida, M., 2007. Multiple Histone Deacetylases and the CREB-binding Protein Regulate Pre-mRNA 3′-End Processing. J. Biol. Chem. 282, 4470–4478. https://doi.org/10.1074/jbc.M609745200

Simons, A., Melamed-Bessudo, C., Wolkowicz, R., Sperling, J., Sperling, R., Eisenbach, L., Rotter, V., 1997. PACT: cloning and characterization of a cellular p53 binding protein that interacts with Rb. Oncogene 14, 145–155. https://doi.org/10.1038/sj.onc.1200825

Singh, P., Alley, T.L., Wright, S.M., Kamdar, S., Schott, W., Wilpan, R.Y., Mills, K.D., Graber, J.H., 2009. Global changes in processing of mRNA 3’ untranslated regions characterize clinically distinct cancer subtypes. Cancer Res. 69, 9422–9430. https://doi.org/10.1158/0008-5472.CAN-09-2236

Skourti-Stathaki, K., Kamieniarz-Gdula, K., Proudfoot, N.J., 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. Nature 516, 436–439. https://doi.org/10.1038/nature13787

Skrajna, A., Yang, X., Dadlez, M., Marzluff, W.F., Dominski, Z., 2018. Protein composition of catalytically active U7-dependent processing complexes assembled on histone pre-mRNA containing biotin and a photo-cleavable linker. Nucleic Acids Res. 46, 4752–4770. https://doi.org/10.1093/nar/gky133

Smith, J.J., Rücknagel, K.P., Schierhorn, A., Tang, J., Nemeth, A., Linder, M., Herschman, H.R., Wahle, E., 1999. Unusual sites of arginine methylation in Poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3. J. Biol. Chem. 274, 13229–13234. https://doi.org/10.1074/jbc.274.19.13229

So, B.R., Di, C., Cai, Z., Venters, C.C., Guo, J., Oh, J.-M., Arai, C., Dreyfuss, G., 2019. A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells. Mol. Cell 76, 590-599.e4. https://doi.org/10.1016/j.molcel.2019.08.007

Stein, G.S., Stein, J.L., Van Wijnen, A.J., Lian, J.B., 1996. Transcriptional control of cell cycle progression: the histone gene is a paradigm for the G1/S phase and proliferation/differentiation transitions. Cell Biol. Int. 20, 41–49. https://doi.org/10.1006/cbir.1996.0007

Steitz, J.A., Wolin, S.L., Rinke, J., Pettersson, I., Mount, S.M., Lerner, E.A., Hinterberger, M., Gottlieb, E., 1983. Small ribonucleoproteins from eukaryotes: structures and roles in RNA biogenesis. Cold Spring Harb. Symp. Quant. Biol. 47 Pt 2, 893–900. https://doi.org/10.1101/sqb.1983.047.01.103

Steitz, T.A., Steitz, J.A., 1993. A general two-metal-ion mechanism for catalytic RNA. Proc. Natl. Acad. Sci. U. S. A. 90, 6498–6502. https://doi.org/10.1073/pnas.90.14.6498

Stiller, J.W., Hall, B.D., 2002. Evolution of the RNA polymerase II C-terminal domain. Proc. Natl. Acad. Sci. U. S. A. 99, 6091–6096. https://doi.org/10.1073/pnas.082646199

Strub, K., Galli, G., Busslinger, M., Birnstiel, M.L., 1984. The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. EMBO J. 3, 2801–2807. https://doi.org/10.1002/j.1460-2075.1984.tb02212.x

Sun, Y., Hamilton, K., Tong, L., 2020a. Recent molecular insights into canonical pre-mRNA 3'-end processing. Transcription 11, 83–96. https://doi.org/10.1080/21541264.2020.1777047

Sun, Y., Zhang, Y., Aik, W.S., Yang, X.-C., Marzluff, W.F., Walz, T., Dominski, Z., Tong, L., 2020b. Structure of an active human histone pre-mRNA 3′-end processing machinery. Science 367, 700–703. https://doi.org/10.1126/science.aaz7758

Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., Tong, L., 2018. Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc. Natl. Acad. Sci. 115, E1419–E1428. https://doi.org/10.1073/pnas.1718723115

Takagaki, Y., Manley, J.L., 2000. Complex Protein Interactions within the Human Polyadenylation Machinery Identify a Novel Component. Mol. Cell. Biol. 20, 1515–1525. https://doi.org/10.1128/MCB.20.5.1515-1525.2000

Takagaki, Y., Manley, J.L., 1997. RNA recognition by the human polyadenylation factor CstF. Mol. Cell. Biol. 17, 3907–3914. https://doi.org/10.1128/MCB.17.7.3907

Takagaki, Y., Manley, J.L., 1994. A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. Nature 372, 471–474. https://doi.org/10.1038/372471a0

Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., Shenk, T., 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. Genes Dev. 4, 2112–2120. https://doi.org/10.1101/gad.4.12a.2112

Takagaki, Y., Ryner, L.C., Manley, J.L., 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev. 3, 1711–1724. https://doi.org/10.1101/gad.3.11.1711

Takagaki, Y., Ryner, L.C., Manley, J.L., 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. Cell 52, 731–742. https://doi.org/10.1016/0092-8674(88)90411-4

Takagaki, Y., Seipelt, R.L., Peterson, M.L., Manley, J.L., 1996. The Polyadenylation Factor CstF-64 Regulates Alternative Processing of IgM Heavy Chain Pre-

mRNA during B Cell Differentiation. Cell 87, 941–952. https://doi.org/10.1016/S0092-8674(00)82000-0
Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., Proudfoot, N.J., 2012. Gene loops enhance transcriptional directionality. Science 338, 671–675. https://doi.org/10.1126/science.1224350
Tatomer, D.C., Elrod, N.D., Liang, D., Xiao, M.-S., Jiang, J.Z., Jonathan, M., Huang, K.-L., Wagner, E.J., Cherry, S., Wilusz, J.E., 2019. The Integrator complex cleaves nascent mRNAs to attenuate transcription. Genes Dev. 15.
Tian, B., Graber, J.H., 2012. Signals for pre-mRNA cleavage and polyadenylation: Polyadenylation signals. Wiley Interdiscip. Rev. RNA 3, 385–396. https://doi.org/10.1002/wrna.116
Tian, B., Hu, J., Zhang, H., Lutz, C.S., 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 33, 201–212. https://doi.org/10.1093/nar/gki158
Tian, B., Manley, J.L., 2017. Alternative polyadenylation of mRNA precursors. Nat. Rev. Mol. Cell Biol. 18, 18–30. https://doi.org/10.1038/nrm.2016.116
Tian, B., Manley, J.L., 2013. Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem. Sci. 38, 312–320. https://doi.org/10.1016/j.tibs.2013.03.005
Tian, B., Pan, Z., Lee, J.Y., 2007. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. 17, 156–165. https://doi.org/10.1101/gr.5532707
Turtola, M., Manav, M.C., Kumar, A., Tudek, A., Mroczek, S., Krawczyk, P.S., Dziembowski, A., Schmid, M., Passmore, L.A., Casañal, A., Jensen, T.H., 2021. Three-layered control of mRNA poly(A) tail synthesis in Saccharomyces cerevisiae. Genes Dev. 35, 1290–1303. https://doi.org/10.1101/gad.348634.121
Twu, K.Y., Kuo, R.-L., Marklund, J., Krug, R.M., 2007. The H5N1 influenza virus NS genes selected after 1998 enhance virus replication in mammalian cells. J. Virol. 81, 8112–8121. https://doi.org/10.1128/JVI.00006-07
Twu, K.Y., Noah, D.L., Rao, P., Kuo, R.-L., Krug, R.M., 2006. The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target. J. Virol. 80, 3957–3965. https://doi.org/10.1128/JVI.80.8.3957-3965.2006
Vannini, A., Cramer, P., 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol. Cell 45, 439–446. https://doi.org/10.1016/j.molcel.2012.01.023
Vasserot, A.P., Schaufele, F.J., Birnstiel, M.L., 1989. Conserved terminal hairpin sequences of histone mRNA precursors are not involved in duplex formation with the U7 RNA but act as a target site for a distinct processing factor. Proc. Natl. Acad. Sci. U. S. A. 86, 4345–4349. https://doi.org/10.1073/pnas.86.12.4345
Venkataraman, K., Brown, K.M., Gilmartin, G.M., 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. 19, 1315–1327. https://doi.org/10.1101/gad.1298605
Vethantham, V., Rao, N., Manley, J.L., 2007. Sumoylation Modulates the Assembly and Activity of the Pre-mRNA 3′ Processing Complex. Mol. Cell. Biol. 27, 8848–8858. https://doi.org/10.1128/MCB.01186-07
Vo, L.T.A., Minet, M., Schmitter, J.-M., Lacroute, F., Wyers, F., 2001. Mpe1, a Zinc Knuckle Protein, Is an Essential Component of Yeast Cleavage and

Polyadenylation Factor Required for the Cleavage and Polyadenylation of mRNA. Mol. Cell. Biol. 21, 8346–8356. https://doi.org/10.1128/MCB.21.24.8346-8356.2001
Vorländer, M.K., Pacheco-Fiallos, B., Plaschka, C., 2022. Structural basis of mRNA maturation: Time to put it together. Curr. Opin. Struct. Biol. 75, 102431. https://doi.org/10.1016/j.sbi.2022.102431
Wahl, M.C., Will, C.L., Lührmann, R., 2009. The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701–718. https://doi.org/10.1016/j.cell.2009.02.009
Wahle, E., 1991a. Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. 266, 3131–3139. https://doi.org/10.1016/S0021-9258(18)49964-4
Wahle, E., 1991b. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell 66, 759–768. https://doi.org/10.1016/0092-8674(91)90119-J
Wahle, E., Keller, W., 1992. THE BIOCHEMISTRY OF 3′-END CLEAVAGE AND POLYADENYLATION OF MESSENGER RNA PRECURSORS. Annu Rev Biochem 24.
Wahle, E., Rüegsegger, U., 1999. 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. 23, 277–295. https://doi.org/10.1111/j.1574-6976.1999.tb00400.x
Wallace, A.M., Dass, B., Ravnik, S.E., Tonk, V., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., MacDonald, C.C., 1999. Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. Proc. Natl. Acad. Sci. U. S. A. 96, 6763–6768. https://doi.org/10.1073/pnas.96.12.6763
Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. https://doi.org/10.1038/nature07509
Wang, R., Zheng, D., Yehia, G., Tian, B., 2018. A compendium of conserved cleavage and polyadenylation events in mammalian genes. Genome Res. 28, 1427–1441. https://doi.org/10.1101/gr.237826.118
Wani, S., Yuda, M., Fujiwara, Y., Yamamoto, M., Harada, F., Ohkuma, Y., Hirose, Y., 2014. Vertebrate Ssu72 Regulates and Coordinates 3′-End Formation of RNAs Transcribed by RNA Polymerase II. PLoS ONE 9, e106040. https://doi.org/10.1371/journal.pone.0106040
Wassarman, D.A., Steitz, J.A., 1991. RNA splicing. Alive with DEAD proteins. Nature 349, 463–464. https://doi.org/10.1038/349463a0
Weiss, E.A., Gilmartin, G.M., Nevins, J.R., 1991. Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. EMBO J. 10, 215–219. https://doi.org/10.1002/j.1460-2075.1991.tb07938.x
Weitzer, S., Martinez, J., 2007. The human RNA kinase hClp1 is active on 3′ transfer RNA exons and short interfering RNAs. Nature 447, 222–226. https://doi.org/10.1038/nature05777
Wen, Y., Shatkin, A.J., 1999. Transcription elongation factor hSPT5 stimulates mRNA capping. Genes Dev. 13, 1774–1779. https://doi.org/10.1101/gad.13.14.1774

Werner, F., 2012. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. J. Mol. Biol. 417, 13–27. https://doi.org/10.1016/j.jmb.2012.01.031
West, S., Gromak, N., Proudfoot, N.J., 2004. Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432, 522–525. https://doi.org/10.1038/nature03035
Wickens, M., Stephenson, P., 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. Science 226, 1045–1051. https://doi.org/10.1126/science.6208611
Wigington, C.P., Williams, K.R., Meers, M.P., Bassell, G.J., Corbett, A.H., 2014. Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions: Poly(A) RNA-binding proteins and polyadenosine RNA. Wiley Interdiscip. Rev. RNA 5, 601–622. https://doi.org/10.1002/wrna.1233
Wilkinson, M.E., Charenton, C., Nagai, K., 2020. RNA Splicing by the Spliceosome. Annu. Rev. Biochem. 89, 359–388. https://doi.org/10.1146/annurev-biochem-091719-064225
Winstall, E., Sadowski, M., Kuhn, U., Wahle, E., Sachs, A.B., 2000. The Saccharomyces cerevisiae RNA-binding protein Rbp29 functions in cytoplasmic mRNA metabolism. J. Biol. Chem. 275, 21817–21826. https://doi.org/10.1074/jbc.M002412200
Wu, X., Liu, M., Downie, B., Liang, C., Ji, G., Li, Q.Q., Hunt, A.G., 2011. Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. Proc. Natl. Acad. Sci. U. S. A. 108, 12533–12538. https://doi.org/10.1073/pnas.1019732108
Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., Jacquier, A., 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121, 725–737. https://doi.org/10.1016/j.cell.2005.04.030
Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., Li, W., 2014. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. Nat. Commun. 5, 5274. https://doi.org/10.1038/ncomms6274
Xiang, K., Manley, J.L., Tong, L., 2012. An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72. Genes Dev. 26, 2265–2270. https://doi.org/10.1101/gad.198853.112
Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., Tong, L., 2010. Crystal structure of the human symplekin–Ssu72–CTD phosphopeptide complex. Nature 467, 729–733. https://doi.org/10.1038/nature09391
Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J.L., Tong, L., 2009. Structure and function of the 5'-->3' exoribonuclease Rat1 and its activating partner Rai1. Nature 458, 784–788. https://doi.org/10.1038/nature07731
Xu, X., Pérébaskine, N., Minvielle-Sébastia, L., Fribourg, S., Mackereth, C.D., 2015. Chemical shift assignments of a new folded domain from yeast Pcf11. Biomol. NMR Assign. 9, 421–425. https://doi.org/10.1007/s12104-015-9622-2
Xue, Z., Warren, R.L., Gibb, E.A., MacMillan, D., Wong, J., Chiu, R., Hammond, S.A., Yang, C., Nip, K.M., Ennis, C.A., Hahn, A., Reynolds, S., Birol, I., 2018. Recurrent tumor-specific regulation of alternative polyadenylation of cancer-related genes. BMC Genomics 19, 536. https://doi.org/10.1186/s12864-018-4903-7

Yan, C., Wan, R., Shi, Y., 2019. Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. Cold Spring Harb. Perspect. Biol. 11, a032409. https://doi.org/10.1101/cshperspect.a032409

Yan, J., Marr, T.G., 2005. Computational analysis of 3'-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. Genome Res. 15, 369–375. https://doi.org/10.1101/gr.3109605

Yang, C., Hager, P.W., Stiller, J.W., 2014. The identification of putative RNA polymerase II C-terminal domain associated proteins in red and green algae. Transcription 5, e970944. https://doi.org/10.4161/21541264.2014.970944

Yang, F., Hsu, P., Lee, S.D., Yang, W., Hoskinson, D., Xu, W., Moore, C., Varani, G., 2017. The C terminus of Pcf11 forms a novel zinc-finger structure that plays an essential role in mRNA 3'-end processing. RNA 23, 98–107. https://doi.org/10.1261/rna.058354.116

Yang, Q., Coseno, M., Gilmartin, G.M., Doublié, S., 2011a. Crystal Structure of a Human Cleavage Factor CFIm25/CFIm68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. Structure 19, 368–377. https://doi.org/10.1016/j.str.2010.12.021

Yang, Q., Doublié, S., 2011. Structural biology of poly(A) site definition: Poly(A) site definition. Wiley Interdiscip. Rev. RNA 2, 732–747. https://doi.org/10.1002/wrna.88

Yang, Q., Gilmartin, G.M., Doublié, S., 2011b. The structure of human Cleavage Factor Im hints at functions beyond UGUA-specific RNA binding: A role in alternative polyadenylation and a potential link to 5' capping and splicing. RNA Biol. 8, 748–753. https://doi.org/10.4161/rna.8.5.16040

Yang, Q., Gilmartin, G.M., Doublie, S., 2010. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3' processing. Proc. Natl. Acad. Sci. 107, 10062–10067. https://doi.org/10.1073/pnas.1000848107

Yang, W., Hsu, P.L., Yang, F., Song, J.-E., Varani, G., 2018. Reconstitution of the CstF complex unveils a regulatory role for CstF-50 in recognition of 3′-end processing signals. Nucleic Acids Res. 46, 493–503. https://doi.org/10.1093/nar/gkx1177

Yang, X., Sullivan, K.D., Marzluff, W.F., Dominski, Z., 2009. Studies of the 5′ Exonuclease and Endonuclease Activities of CPSF-73 in Histone Pre-mRNA Processing. Mol. Cell. Biol. 29, 31–42. https://doi.org/10.1128/MCB.00776-08

Yao, C., Choi, E.-A., Weng, L., Xie, X., Wan, J., Xing, Y., Moresco, J.J., Tu, P.G., Yates, J.R., Shi, Y., 2013. Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3' processing. RNA 19, 1781–1790. https://doi.org/10.1261/rna.042317.113

Yusa, K., Zhou, L., Li, M.A., Bradley, A., Craig, N.L., 2011. A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. 108, 1531–1536. https://doi.org/10.1073/pnas.1008322108

Zarudnaya, M.I., Kolomiets, I.M., Potyahaylo, A.L., Hovorun, D.M., 2003. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. Nucleic Acids Res. 31, 1375–1386. https://doi.org/10.1093/nar/gkg241

Zhang, F., Cole, C.N., 1987. Identification of a complex associated with processing and polyadenylation in vitro of herpes simplex virus type 1 thymidine kinase precursor RNA. Mol. Cell. Biol. 7, 3277–3286. https://doi.org/10.1128/mcb.7.9.3277-3286.1987

Zhang, H., Rigo, F., Martinson, H.G., 2015. Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Mol. Cell 59, 437–448. https://doi.org/10.1016/j.molcel.2015.06.008

Zhang, S., Aibara, S., Vos, S.M., Agafonov, D.E., Lührmann, R., Cramer, P., 2021. Structure of a transcribing RNA polymerase II–U1 snRNP complex. Science 371, 305–309. https://doi.org/10.1126/science.abf1870

Zhang, Y., Liu, L., Qiu, Q., Zhou, Q., Ding, J., Lu, Y., Liu, P., 2021. Alternative polyadenylation: methods, mechanism, function, and role in cancer. J. Exp. Clin. Cancer Res. CR 40, 51. https://doi.org/10.1186/s13046-021-01852-7

Zhang, Y., Sun, Y., Shi, Y., Walz, T., Tong, L., 2020. Structural Insights into the Human Pre-mRNA 3′-End Processing Machinery. Mol. Cell 77, 800-809.e6. https://doi.org/10.1016/j.molcel.2019.11.005

Zhao, J., Hyman, L., Moore, C., 1999. Formation of mRNA 3' Ends in Eukaryotes: Mechanism, Regulation, and Interrelationships with Other Steps in mRNA Synthesis. MICROBIOL MOL BIOL REV 63, 41.

Zhelkovsky, A.M., Kessler, M.M., Moore, C.L., 1995. Structure-function relationships in the Saccharomyces cerevisiae poly(A) polymerase. Identification of a novel RNA binding site and a domain that interacts with specificity factor(s). J. Biol. Chem. 270, 26715–26720. https://doi.org/10.1074/jbc.270.44.26715

Zheng, H., Jin, Q., Qi, Y., Liu, W., Ren, Y., Wang, X., Chen, F., Cheng, J., Chen, X., Xu, Y., 2021. Structural basis of INTAC-regulated transcription. BioRxiv. https://doi.org/10.1101/2021.11.29.470345

Zheng, H., Qi, Y., Hu, S., Cao, X., Xu, C., Yin, Z., Chen, X., Li, Y., Liu, W., Li, J., Wang, J., Wei, G., Liang, K., Chen, F.X., Xu, Y., 2020. Identification of Integrator-PP2A complex (INTAC), an RNA polymerase II phosphatase. Science 370. https://doi.org/10.1126/science.abb5872

Zhou, Z., Licklider, L.J., Gygi, S.P., Reed, R., 2002. Comprehensive proteomic analysis of the human spliceosome. Nature 419, 182–185. https://doi.org/10.1038/nature01031

Zhu, Y., Wang, X., Forouzmand, E., Jeong, J., Qiao, F., Sowd, G.A., Engelman, A.N., Xie, X., Hertel, K.J., Shi, Y., 2018. Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. Mol. Cell 69, 62-74.e4. https://doi.org/10.1016/j.molcel.2017.11.031

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., Scheres, S.H., 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. eLife 7, e42166. https://doi.org/10.7554/eLife.42166

Acknowledgements

This work would not have been possible without the help of many people.

First, I would like to express my deep gratitude towards Elena Conti for all the support and trust over the years. Thank you for letting me tackle this project in my own ways.

Thank you to Elmar Wahle and his lab for sharing ideas and data with me and for many insightful discussions.

Furthermore, I would like to thank my TAC members, Klaus Förstemann, Jürgen Plitzko, and Christian Benda, for great scientific input and advice throughout the years. This extends to my thesis committee for agreeing to evaluate this work and participate in my defense.

I acknowledge the International Max Planck Research School for Molecular Life Sciences (IMPRS-LS) for organizing interesting and helpful workshops and lectures.

I would like to thank the scientists in the MPIB Core and cryo-EM facilities, especially Daniel Bollschweiler and Tillman Schäfer for all their help during endless hours in the EM bunker and for keeping the microscopes in a well-maintained state. This extends also to Rajan Prabu for the scientific computing environment, Christian Benda for Streptavidin grids, and Christian Benda and Ingmar Schäfer for initial EM guidance.

A special thanks to all members of the Conti department for all the help and input over the years, especially to Daniela Wartini for countless hours in the mammalian cell culture, Fabien Bonneau for sharing his expertise on RNA biochemistry, Petra Birle and Tatjana Krywcun for cloning of many constructs, and Steffen Schüssler for protein preps. Thank you, Lukas Langer, for all the ideas, scientific and private discussions, coffee-walks and well-needed (often at least) distractions. Also, thank you to everyone in the lab for generating such a nice and open atmosphere and for all the time we shared inside and outside the lab.

And finally, I would like to thank my family and especially Ricarda for all the support and patience over the years. Thank you, Ricarda, for your understanding and for always having my back.