Reconstitution of ATP-dependent 3'-cleavage of mammalian pre-mRNA reveals a central role of RBBP6

Moritz Schmidt[1,5], Florian Kluge[1,5], Felix Sandmeir[2,5], Peter Schäfer[1], Christian Tüting[1,4], Christian Ihling[3], Uwe Kühn[1,6], Elena Conti[2,6,7], Elmar Wahle[1,6,7,8]

[1]Institute of Biochemistry and Biotechnology, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, 06099 Halle, Germany

[2]Department of Structural Cell Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

[3]Institute of Pharmacy, Martin Luther University Halle-Wittenberg, 06099 Halle, Germany

Present address:

[4]ZIK HALOMEM and Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, 06099 Halle, Germany

[5]These authors contributed equally.

[6]Senior author

[7]Corresponding authors: conti@biochem.mpg.de, ewahle@biochemtech.uni-halle.de

[8]Lead contact

Summary

The 3' ends of almost all eukaryotic mRNAs are generated in an essential two-step processing reaction, endonucleolytic cleavage of an extended precursor followed by the addition of a poly(A) tail. By reconstituting the reaction from overproduced and purified proteins, we provide a minimal list of fourteen polypeptides essential and two stimulatory for RNA cleavage. In a reaction depending on the polyadenylation signal AAUAAA, the reconstituted system cleaves pre-mRNA at a single preferred site corresponding to the one used *in vivo*. Among the proteins, Cleavage Factor I stimulates cleavage but is not essential, consistent with its prominent role in alternative polyadenylation. RBBP6 is required, structural data showing it to contact and presumably activate the endonuclease CPSF73 through its DWNN domain. The C-terminal domain of RNA polymerase II is dispensable. ATP, but not its hydrolysis, supports RNA cleavage by binding to the hClp1 subunit of cleavage factor II with submicromolar affinity.

Introduction

The 3' ends of eukaryotic mRNAs are generated by an essential post-transcriptional processing reaction: An extended primary transcript is cleaved by an endonuclease, then a poly(A) tail is added to the upstream cleavage fragment whereas the downstream fragment is degraded. 3' end processing is co-transcriptional and contributes to transcription termination by RNA polymerase II downstream of the cleavage site (Eaton et al., 2020; Porrua et al., 2016; Proudfoot, 2016). Poly(A) tails are important in the initiation of translation (Jackson et al., 2010) and the control of mRNA half-life (Chen and Shyu, 2011; Eisen et al., 2020). Alternative polyadenylation is widespread in higher eukaryotes and can change the amino acid sequence of the encoded protein or, more frequently, generate mRNAs with different regulatory sites in their 3' UTRs (Gruber and Zavolan, 2019; Tian and Manley, 2017). The only known eukaryotic mRNAs not decorated with a poly(A) tail are histone mRNAs in animal cells, the precursors of which are cleaved but not polyadenylated (Marzluff and Koreski, 2017; Sun et al., 2020).

Cleavage and polyadenylation are carried out by a large protein complex (Kumar et al., 2019; Xiang et al., 2014) containing, in mammals, four heterooligomeric complexes: Cleavage and Polyadenylation Specificity Factor (CPSF), Cleavage Factors I and II (CF I and II), and Cleavage stimulation Factor (CstF). CPSF consists of two subcomplexes: The first, mammalian polyadenylation specificity factor (mPSF) (Clerici et al., 2017; Clerici et al., 2018; Schönemann et al., 2014; Sun et al., 2018), contains CPSF160, WDR33, hFip1 and CPSF30. WDR33 and CPSF30 jointly recognize the polyadenylation signal AAUAAA, located upstream of the cleavage site. Poly(A) polymerase (PAP) is not stably associated with mPSF, but relies on this factor in AAUAAA-dependent polyadenylation. PAP is also required for pre-mRNA cleavage (Christofori and Keller, 1989; Takagaki et al., 1988). The second subcomplex of CPSF has been termed 'mammalian cleavage factor' (mCF) (Sun et al., 2020; Zhang et al., 2020), but we will use the term 'cleavage module' (Schönemann et al., 2014) to avoid confusion with cleavage factors CF I and CF II. The cleavage module is composed of the endonuclease CPSF73, the related but catalytically inactive protein CPSF100, symplekin, and perhaps CstF64 (Mandel et al., 2006; Sullivan et al., 2009; Sun et al., 2020; Yang et al., 2020; Zhang et al., 2020). CF I consists of a dimer of CF I-25 and two copies of CF I-59 or CF I-68 (Kim et al., 2010; Ruegsegger et al., 1998; Yang et al., 2011) and recognizes UGUA motifs upstream of the polyadenylation signal (Brown and Gilmartin, 2003; Venkataraman et al., 2005; Yang et al., 2011; Yang et al., 2010). CF II is a heterodimer of hPcf11 and hClp1 (Schäfer et al., 2018). Pcf11 binds the RNA polymerase II CTD and is important for the coupling of 3' end processing to transcription termination (Kamieniarz-Gdula et al., 2019; Meinhart and Cramer, 2004); Clp1 has an RNA 5' kinase activity of unknown function (Weitzer et al., 2015). CstF, which recognizes GU-rich sequence elements downstream of the cleavage site, is a hexamer composed of two copies each of CstF77, CstF64, and CstF50 (Bai et al., 2007; Takagaki and Manley, 1994; Yang et al., 2018). In addition to these proteins, RBBP6 is likely to be involved, based on its sequence similarity to yeast 3' processing factor Mpe1, its presence in affinity-purified 3' processing complexes (Shi et al., 2009) and additional experimental evidence (Di Giammartino et al., 2014). The proteins are represented schematically in **Fig. 1A**. In addition,

the nuclear poly(A) binding protein (PABPN1) participates in poly(A) tail extension (Kühn et al., 2017; Wahle, 1991), but may also play a role in poly(A) site choice (de Klerk et al., 2012; Jenal et al., 2012). Several other proteins have been suggested to be involved in 3' end formation. For example, the C-terminal domain (CTD) of RNA polymerase II has been reported to be required for 3' processing (Hirose and Manley, 1998), consistent with the co-transcriptional nature of the reaction *in vivo*.

3' end processing of yeast mRNA precursors has been reconstituted from a set of proteins similar but not identical to the mammalian factors (Hill et al., 2019) (**Table S1**). 3' end cleavage of mammalian histone mRNAs has also been reconstituted; required factors include the cleavage module described above (Sun et al., 2020). In contrast, reconstitution of the mammalian 3' end processing system has been limited to the second step, polyadenylation, which requires mPSF, poly(A) polymerase and PABPN1 (Schönemann et al., 2014). Investigations of the cleavage reaction have relied on combinations of 'native' proteins purified to various degrees or on the use of unfractionated nuclear extract. Thus, the list of proteins essential for the reaction is not known with certainty, and even the involvement of ATP and perhaps other small molecules has been controversial. In order to clarify these issues, we have reconstituted mammalian pre-mRNA cleavage from recombinant proteins.

Results

Reconstitution of pre-mRNA cleavage

With the aim of reconstituting the pre-mRNA 3' processing reaction, we overproduced and purified established and putative mammalian 3' processing factors (**Fig. 1B; Suppl. Fig. 2D**). A mixture of proteins identified by classical resolution and reconstitution experiments, mPSF, cleavage module, CF I, CF II, CstF, and PAPα, proved unable to catalyze specific cleavage of an RNA containing the SV40 late polyadenylation site. Only upon inclusion of RBBP6 did the proteins, in the presence of 3'-dATP, cleave the RNA into an upstream fragment of the expected size plus a heterogeneous set of downstream fragments (**Fig. 1C)**. An RNA with a point mutation in its AAUAAA signal was not cleaved. Accumulation of the upstream fragment

proceeded without a pronounced lag phase at a slowly decreasing rate. The efficiency of the reaction was variable; 30 – 40% of the precursor were cleaved under optimal conditions. The reaction proceeded similarly in nuclear extract, except that the downstream fragment was degraded and the precursor RNA was slightly shortened from its 3' end (**Fig. 1C**). In these experiments, chain-terminating 3'-dATP prevented polyadenylation of the RNA to allow direct detection of the upstream cleavage product. When the reconstituted reaction was carried out with ATP instead, polyadenylation took place. RNase H/oligo(dT) digestion revealed polyadenylation both of the upstream cleavage product and of the precursor RNA, as reported for experiments in nuclear extract (Zarkower et al., 1986) (**Fig. 1D**). The efficiency of cleavage was slightly higher with ATP than with 3'-dATP (**Fig. 1D and Suppl. Fig. 2A)**.

In nuclear extract and *in vivo*, the SV40 late RNA is cleaved in the second of three consecutive CAA repeats (Sheets et al., 1987; Wickens and Stephenson, 1984) (www.ncbi.nlm.nih.gov/genbank; entry J02400.1). The cleavage site used in the reconstituted reaction was first analyzed by high resolution gel electrophoresis of the upstream cleavage fragment. Use of an uncapped RNA to avoid heterogeneity due to incomplete cap incorporation resulted in a single fragment at approximately the expected position (**Fig. 1E**), demonstrating dominant cleavage at a single site. In contrast, the reconstituted yeast system cleaved a model substrate RNA with equal probability at any of three adjacent phosphodiester bonds (Hill et al., 2019). As the SV40 late RNA is cleaved at one preferred site, the heterogeneous nature of the downstream fragment could be the result of transcript heterogeneity at the 3' end and/or of limited 5' exonuclease digestion subsequent to cleavage. In order to avoid complications by potential 3' end heterogeneity, we analyzed 5' ends of the downstream fragments by primer extension. A chemically synthesized RNA representing the expected downstream fragment (Sheets et al., 1987) served as a positive control. The dominant reverse transcription product obtained from authentic downstream cleavage fragments corresponded to the product obtained from the synthetic RNA (**Fig. 1F**). Most other bands in the neighborhood were background, being equally intense in a control reaction in which RNA cleavage was prevented by an active site mutation in CPSF73. However, two bands, three nucleotides upstream and

downstream, respectively, of the main product, were also dependent on CPSF73 activity. Presumably, they represent minor cleavage sites in either of the two neighboring CAA repeats. Together, the data show that cleavage occurs predominantly at the site mapped earlier. In contrast to the situation in nuclear extract (Sheets et al., 1987), little to no 5' end nibbling of the downstream fragment takes place. Still, at multiple time points of several independent experiments, average recovery of the downstream fragment was only ~ 10% of the upstream fragment. Thus, most of the downstream fragment must be degraded. However, when the synthetic downstream fragment, carrying a 5'-monophosphate like the authentic fragment, was 3'-labeled and incubated with the mixture of all cleavage factors, no significant degradation took place (**Suppl. Fig. 1**), suggesting that degradation of the 3' cleavage fragment only occurs in the context of the processing reaction.

Definition of a minimal cleavage complex: CF I is not essential

We used the reconstitution system to define the minimal set of factors essential for pre-mRNA cleavage. Most factors were indispensable; their individual omission abolished the reaction (**Fig. 2A**). PAP was among the essential proteins, but its catalytic activity was not required for RNA cleavage (**Suppl. Fig. 2A**). The only protein that could be omitted without complete loss of cleavage activity was CF I (**Fig. 2A**). A titration experiment confirmed that CF I was stimulatory for cleavage of the SV40 late RNA but not essential (**Fig. 2B**). Whereas the SV40 late RNA contains a single copy of the CF I binding motif, UGUA, adenovirus L3 contains the preferred configuration (Yang et al., 2011), two copies upstream of AAUAAA. CF I stimulated cleavage of wild-type L3 RNA. As expected, the effect was strongly reduced by mutation of the upstream copies of UGUA to UcUA, with or without mutation of a third copy immediately upstream of the cleavage site (**Fig. 2B)**. Importantly, the mutations had no significant effect on the efficiency of cleavage when no CF I was added. Thus, there was no functionally relevant amount of CF I present as a contamination. CF I has been reported also to stimulate polyadenylation (Brown and Gilmartin, 2003), but in our hands the factor had no detectable

effect on the polyadenylation of a 'pre-cleaved' L3 RNA even when mPSF and PAP were present at limiting concentrations (**Suppl. Fig. 3**).

The C-terminal domain (CTD) of RNA polymerase II has been reported to be essential for pre-mRNA cleavage *in vitro* (Hirose and Manley, 1998; Ryan et al., 2002). However, cleavage in the reconstituted system proceeded without addition of the CTD, and MS analysis did not detect the large subunit of RNA polymerase II as a contamination. The CTD was overproduced in *E. coli* or Sf21 cells and purified. Addition of either preparation to reconstituted cleavage reactions had a small (~ 1.3fold) stimulatory effect of uncertain significance (**Suppl. Fig. 2B**). The Sf21-made CTD was shown to be phosphorylated and reacted with an antibody specific for the Ser2-phosphorylated CTD (**Suppl. Fig. 2C**). Thus, the phosphorylation of Ser2, which is important for CTD function in 3' end processing (Ahn et al., 2004; Davidson et al., 2014; Meinhart and Cramer, 2004), was present in the Sf21-derived preparation.

Five additional proteins (PABPN1, SSU72, a heterotetrameric protein phosphatase 1 complex, CDC73 and XRN2) potentially participating in the 3' cleavage reaction were also tested, but had little effect on the reaction (**Suppl. Fig. 2D, E**).

RBBP6 plays a central role in pre-mRNA cleavage

Fig. 2A confirms that RBBP6 is essential for pre-mRNA cleavage. The protein contains three structured domains at its N-terminus: a ubiquitin-like or DWNN domain, a CCHC zinc knuckle and an E3 ubiquitin-protein ligase RING domain (Kappo et al., 2012), the rest consisting of low complexity/disordered regions (**Fig. 3A**). In agreement with (Di Giammartino et al., 2014), N-terminal fragments were active in pre-mRNA cleavage: Fragments containing the first two or all three structured domains (amino acids 1-253 or 1-340) had a similar activity as a longer fragment extending into the unstructured part (amino acids 1-780), but full-length protein was active at lower concentrations (**Fig. 3A, B**). The DWNN domain alone (aa 1 – 81) (Pugh et al., 2006) was inactive (**data not shown**). The shortest active N-terminal fragment (1-253) was lacking the RING domain. Also, an internal deletion of the domain (amino acids 253-326) from full-length RBBP6 did not diminish the protein's activity (**Fig. 3A, B**). Thus, a potential E3 ligase

function is not essential for the role of RBBP6 in RNA cleavage. The zinc knuckle of RBBP6 was inactivated, in the context of $RBBP6_{1-780}$, by mutation of two cysteine residues (C161G/C164G). The mutant protein was nearly inactive in pre-mRNA cleavage (**Fig. 3C**), consistent with observations in yeast (Lee and Moore, 2014). The ring domain of RBBP6 dimerizes at high concentrations ($K_D \approx 0.1$ mM) (Kappo et al., 2012). As the ring domain is dispensable, dimerization is unlikely to be important for cleavage. Accordingly, a point mutation inhibiting dimerization (N312D) (Kappo et al., 2012) did not inactivate $RBBP6_{1-780}$ (**Fig. 3C**). As also reported for Mpe1 (Lee and Moore, 2014), $RBBP6_{1-780}$ bound RNA (**Fig. 3D**). The activity co-migrated with the protein through an anion exchange column (**Fig. 3E**). The zinc finger mutation had only a minor effect on the RNA binding activity of RBBP6, perhaps due to a high arginine content in the C-terminal half of the protein fragment (**data not shown**).

The experiments suggested that RBBP6 might have a role in mammalian pre-mRNA cleavage analogous to the one of Mpe1 in yeast, helping to activate the CPSF73 endonuclease in the context of a correct pre-mRNA cleavage site. Yeast Mpe1 interacts directly with Ysh1, the orthologue of CPSF73 (Hill et al., 2019). Mpe1 shares significant sequence similarity with the N-terminal part of RBBP6. We therefore tested whether $RBBP6_{1-335}$, which includes the three structured domains and provides RBBP6 function in the cleavage reaction, may engage CPSF73 in an interaction similar to Mpe1-Ysh1. In pull-down experiments using purified proteins, $RBBP6_{1-335}$ efficiently co-purified with a seven-subunit CPSF complex, even without addition of an RNA substrate (**Fig. 4A, lane 5**). Interestingly, we observed a reproducible, albeit weaker interaction with a four-subunit cleavage module (containing CstF64) alone (**Fig. 4A, lane 4**). Based on these results, we reconstituted an eight-subunit CPSF-$RBBP6_{1-335}$ complex for single-particle cryo-EM analysis.

The 2D class averages clearly showed structural details, but preferred orientation and inherent flexibility limited the overall resolution of the 3D reconstruction (**Fig. 4B** and **Suppl. Fig. 4A-C**). The resulting cryo-EM map, filtered to 10 Å, showed similar features as the previously reported reconstruction of CPSF in isolation (**Suppl. Fig 4A**) (Zhang et al., 2020). The assembly contained a rather rigid mPSF core adjacent to a cloverleaf structure

characterized by two globular lobes (corresponding to CPSF100 and CPSF73) and an extended lobe (corresponding to symplekin). The proximal lobe, which directly interacts with mPSF, was previously interpreted as corresponding to CPSF100 and the distal solvent-facing lobe as corresponding to CPSF73 (Zhang et al., 2020). In our cryo-EM reconstruction, extra density was observed protruding from the proximal lobe, suggesting this density may correspond to the additional subunit present in the sample, RBBP6$_{1-335}$, and in turn pointing to the proximal lobe as corresponding to CPSF73 rather than CPSF100. To test this hypothesis, we overexpressed RBBP6$_{1-335}$ with either CPSF100 or CPSF73 in HEK cells and performed co-precipitation experiments. RBBP6$_{1-335}$ indeed reproducibly co-purified with CPSF73 and not with CPSF100 (**Fig. 4C**).

Cross-linking mass spectrometry (XL-MS) was used to examine the structural arrangement in the context of the RBBP6-bound CPSF complex. After treatment with Bis(sulfosuccinimidyl)suberate (BS3), intermolecular cross-links between the RBBP6 DWNN domain and CPSF73 were repeatedly detected (**Suppl. Fig. 4D**). Based on these data, AlphaFold2 calculations were performed to predict a structural model of the CPSF73-RBBP6 complex (**Suppl. Fig. 4E**) (Jumper et al., 2021; Mirdita et al., 2021). In the resulting model, the RBBP6 DWNN domain docks onto the metallo-ß-lactamase domain of CPSF73 in close proximity to the active site opening (**Fig. 4E**), similar to the binding observed for the yeast orthologues (Hill et al., 2019). Mapping the cross-links onto the model revealed that they were well within the expected distance range of BS3, supporting the AlphaFold2 prediction (**Fig. 4D**). The CPSF73-RBBP6 model fitted into the size and shape of the proximal lobe and features of the protruding density. We note that the protruding density extends further and that additional cross-links were identified between the region of RBBP6 downstream of the DWNN domain and other subunits of the CPSF complex, specifically CPSF100, WDR33 and symplekin (**Suppl Fig. 4D**). This suggested additional contact sites in this more flexible part, consistent with the relative strengths of the interactions observed in the pull-down assays (**Fig. 4A**).

To test our model, we introduced mutations aimed at disrupting the interaction between RBBP6 and CPSF73: Positively charged residues in the CPSF73-binding interface of $RBBP6_{1-335}$ were changed to aspartic acid (**Fig. 4E**). These mutations disrupted the interaction in pull-down experiments using purified proteins as well as in co-precipitation experiments in HEK cell lysates (**Fig. 4F** and **Suppl. Fig 4F**) and obliterated the activity of $RBBP6_{1-335}$ in pre-mRNA cleavage (**Fig. 4G**). Thus, the identified interface between the RBBP6 DWNN domain and CPSF73 is indeed the major binding site recruiting RBBP6 to the CPSF complex; the interaction is essential for RNA cleavage. At the same time, the data locate the CPSF73 subunit at the proximal lobe adjacent to the mPSF core. Collectively, these data reveal that the RBBP6-CPSF73 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.

Subunits of CstF and the cleavage module active in pre-mRNA cleavage

Reconstituted cleavage reactions were further used to define the compositions of active CstF and cleavage module. Hexameric CstF can contain CstF64 or its paralog CstF64τ or a mixture of both (Dass et al., 2001; Yao et al., 2013). We prepared CstF containing only CstF64 [CstF (64/64)] or only the 64tau variant [CstF (τ/τ)] or a mixture of both [CstF (64/τ)] (**Fig. 5A**). The existence of mixed complexes was demonstrated by the ability of FLAG-tagged CstF64τ to carry along untagged CstF64 as well as CstF77 and CstF50 in a pull-down experiment (**Fig. 5A**, right panel). All three versions of the complex supported cleavage activity (**Fig. 5B**), as anticipated from earlier studies suggesting largely overlapping functions of CstF64 and 64τ (Ruepp et al., 2011; Yao et al., 2013). A slightly lower activity of τ-containing complexes in pre-mRNA cleavage and RNA binding (**Suppl. Fig. 5A; Table S2**) was repeatedly observed with independent preparations. In addition to CF I, the CstF subunit CstF50 is the only component in mammalian 3' end processing that has no yeast ortholog. Partial CstF complexes lacking CstF50 (CstFΔ50 or CstFτΔ50) were nearly inactive. Complementation with individually purified CstF50 restored the activity (**Figs. 5A, B**), showing that the subunit is important for

pre-mRNA cleavage. CstF64 or CstF50 alone or in combination were inactive, suggesting an essential role of CstF77 (**Fig. 5A, B**). In this case, complementation was not possible since we were unable to purify CstF77 by itself.

The cleavage module could be prepared from three or four subunits (CM3 or CM4, respectively), containing CstF64 or not (**Fig. 1A**). Both preparations supported pre-mRNA cleavage (**Fig. 5C**). Thus, the cleavage module lacking CstF64 is fully competent for pre-mRNA cleavage, as also observed in histone 3' processing (Sun et al., 2020; Yang et al., 2020). In XL-MS experiments employing the amine-reactive cross-linker DSBU (Muller et al., 2010) in the context of a CM3-containing cleavage reaction, the dominant partners of CstF64 were the other two CstF subunits, but cross-links to the CM subunit symplekin were also reproducibly observed (**Suppl. Fig. 6)**, suggesting that CstF64 can interact with symplekin (Ruepp et al., 2011; Takagaki and Manley, 2000) under reaction conditions. As expected, a double point mutation in the active site of CPSF73 (D75K, H76A) (Mandel et al., 2006) abolished the activity of the cleavage module (**Fig. 5C**). CM3 bound the SV40 late RNA with a K_{50} of ~ 3 nM (**Suppl. Fig. 5B, C; Table S2**). In spite of the presence of RNA-binding CstF64, the RNA affinity of CM4 was barely different from that of CM3 (**Suppl. Fig. 5B; Table S2**).

Cross-linking under reaction conditions also suggested the existence of novel homodimers in the cleavage complex: Homotypic cross-links of hFip1 were consistent with the binding of one hFip1 monomer each to zinc fingers 4 and 5 of CPSF30 (Hamilton and Tong, 2020) and with the presence of more than one copy of Fip1 in the yeast equivalent of CPSF (Casanal et al., 2017). Several homotypic cross-links seen for both subunits of CF II (**Suppl. Fig. 6**) suggest the formation of at least a tetramer under reaction conditions, whereas the isolated factor behaves like a heterodimer (Schäfer et al., 2018).

Pre-mRNA cleavage is ATP-dependent

Experiments in crude nuclear extract initially suggested that ATP was required for RNA cleavage (Moore and Sharp, 1985). In contrast, a later study concluded that ATP was dispensable, but that creatine phosphate, originally added to replenish ATP degraded by

ATPases in the nuclear extract, played an essential role independent of ATP regeneration (Hirose and Manley, 1997). 3' processing of yeast precursor RNAs in a reconstituted system did not require the addition of either ATP or creatine phosphate (Hill et al., 2019).

Neither creatine phosphate nor creatine kinase were included in our standard reconstitution assays. Addition of either or both had no effect as compared to several control compounds (**Fig. 6A**). Thus, creatine phosphate is not essential for cleavage.

The ATP-dependence of cleavage in the reconstituted reaction was assayed in the presence of either wild-type PAP, permitting polyadenylation, or catalytically disabled PAP D115A, revealing cleavage without the complication of polyadenylation. Cleavage was clearly dependent on the addition of ATP or one of several analogs (**Fig. 6B,** compare minus ATP lane to other lanes). In the presence of ATP, both the 5' cleavage product and the substrate RNA were polyadenylated with WT PAP, as discussed above, but use of the mutant revealed cleavage. Methyleno-ATP analogs with non-cleavable α-β or β-γ bonds (APCPP and APPCP, respectively) supported RNA cleavage, in agreement with results obtained in crude extract (Moore and Sharp, 1985). APPCP also served as a substrate for poly(A) polymerase as expected (Bienroth et al., 1993). Activity of the methyleno analogs in cleavage eliminates the possibility that the essential role of ATP is to serve as a substrate for PAP or the RNA kinase activity of hClp1 and agrees with the fact that mutational inactivation of either enzyme does not prevent cleavage (**Suppl. Fig. 2A**) (Schäfer et al., 2018). Arabino-ATP behaved like 3'-dATP (**Fig. 6B)**. At the 0.5 mM concentration used, ~1000fold above the K_A for ATP (see below), GTP, CTP, UTP and ITP supported cleavage to a significant extent.

Three constituents of the cleavage complex are known to bind ATP: CF I-25 (Coseno et al., 2008), PAP, and hClp1. CF I-25 is not responsible for the ATP-dependence of cleavage, since ATP is essential whereas CF I is not. Importantly, a cleavage reaction lacking CF I was still ATP-dependent, ruling out CF I-25 as the relevant ATP binder (**Fig. 6C)**. While use of ATP as a substrate by PAP or hClp1 cannot explain the ATP dependence of cleavage, the nucleotide might act as an allosteric effector for either enzyme. Therefore, mutations in their ATP binding sites were tested. Amino acids D113 and D115 in PAP are involved in ground-

state binding of Mg^{2+}-ATP (Martin et al., 2000). Thus, mutants D113A and D115A are expected to reduce the affinity of PAP for ATP. This is qualitatively supported by strongly reduced UV cross-linking to a photo-activatable ATP analog in both mutants (Martin et al., 1999). If ATP binding to PAP were responsible for the ATP-dependence of cleavage, a reduced affinity for ATP should be reflected in a requirement for higher ATP concentrations in the cleavage reaction. However, the K_A values (concentrations of 3'-dATP that resulted in half-maximal activation of cleavage) were 46.3 ± 7.4 µM with wild-type PAP, 69.3 ± 10.6 µM with the D113A mutant and 81.5 ± 10.2 µM with the D115A mutant (**Fig. 6D**). This minor change in concentration dependence makes it unlikely that ATP binding to the PAP active site is responsible for the ATP-dependence of cleavage.

Based on a comparison to *C. elegans* Clp1 (Dikfidan et al., 2014), amino acids R288 and R293 of human Clp1 bind the γ-phosphate of ATP, whereas E22 forms a hydrogen bond to the 6-amino group of the base. Clp1 variants with mutations in these positions were purified as complexes with hPcf11 (**Fig. 6E**) and first tested with respect to their RNA kinase activity. The R288A, R293K double mutant (Dikfidan et al., 2014) was inactive, and activities of the E22A and E22Q mutants were approximately twofold reduced under standard reaction conditions. ITP is lacking the 6-amino group hydrogen bonding to E22 in ATP, and the E22A mutation facilitated the use of ITP in the kinase reaction (**Fig. 6F**). The same mutations were then tested, in the presence of PAP D115A, for the ATP dependence of RNA cleavage: The K_A value was 0.5 ± 0.1 µM ATP with wild-type hClp1, ~hundredfold lower than the K_A of 3'-dATP. The RR mutant reproducibly showed significant cleavage activity without ATP addition; the K_A was 5.7 ± 3.4 µM ATP. The E22Q mutant had reduced cleavage activity at all ATP concentrations tested; the K_A was 95.9 ± 19.8 µM. K_A with the E22A mutant was 10.9 ± 2.7 µM ATP (**Fig. 6G**). The E22A mutation also improved cleavage in the presence of ITP, in parallel with its effect in the kinase reaction **(Fig. 6H).** The fact that mutations in the ATP binding site of hClp1 alter the ATP concentration dependence and, less dramatically, the nucleotide specificity of pre-mRNA cleavage strongly suggests that ATP facilitates cleavage by serving as an allosteric effector of hClp1.

Discussion

By reconstituting mammalian pre-mRNA 3' processing from recombinant proteins, we have provided a minimal parts list for the reaction. With the caveat that not every individual subunit of heterooligomeric factors has been tested, the following components are necessary and sufficient for the reaction: mPSF (four subunits), cleavage module (three subunits), CstF (three subunits), CF II (two subunits), RBBP6 and PAP, fourteen polypeptides in total. PABPN1 is involved in polyadenylation, but not essential for pre-mRNA cleavage. The list of required factors is now a very close match to the equivalent system from *S. cerevisiae* (Hill et al., 2019): Mammalian CstF50 and yeast Hrp1 (CF IB) are the only essential 3' processing factors unique to their respective systems. In addition, yeast Pta1 is part of the phosphatase module of CPF, which is dispensable for 3' processing (Casanal et al., 2017; Hill et al., 2019; Lidschreiber et al., 2018; Nedea et al., 2003), whereas the mammalian Pta1 orthologue, symplekin, is part of the cleavage module and presumably essential for cleavage. Orthologues to other subunits of the yeast phosphatase module were not required. The reconstituted mammalian system catalyzes cleavage of predominantly one phosphodiester bond at the expected position. The selection of the exact cleavage site seems to be more precise in comparison to the yeast system (Hill et al., 2019).

Based on experiments employing native factors purified from nuclear extract, CF I was considered an essential 3' processing factor (Rüegsegger et al., 1996; Ruegsegger et al., 1998; Takagaki et al., 1989). Our experiments now show that CF I, which has no equivalent in *S. cerevisiae*, stimulates 3' processing, but is not essential. CF I also binds more weakly to RNA than any of the other 3' processing factors (**Table S2**). Both observations suggest that CF I selectively participates in the processing of some RNAs but not others. This is consistent with the prominent role of the protein in alternative polyadenylation (Gruber and Zavolan, 2019): Reduced expression of CF I had a much stronger effect on alternative polyadenylation than the depletion of any other individual 3' processing factor and consistently caused pronounced shifts to the use of proximal (upstream) polyadenylation sites in many genes

(Alcott et al., 2020; Kim et al., 2010; Kubo et al., 2006; Li et al., 2015; Martin et al., 2012; Masamha et al., 2014; Sasado et al., 2017; Zhu et al., 2018). The CF I binding motif, UGUA, was enriched in distal sites neglected and depleted in proximal sites favored upon knock-down, consistent with preferred cross-linking of CF I to distal sites (Li et al., 2015; Martin et al., 2012; Zhu et al., 2018). These data are explained by the model proposed by (Zhu et al., 2018): CF I is an activator of 3' processing dependent on its binding site, UGUA. Reduced availability of CF I leads to a decreased use of UGUA-containing (distal) poly(A) sites, whereas proximal sites lacking UGUA make no use of CF I anyway and, thus, are not affected by its loss; consequently, their relative use increases. The model of (Zhu et al., 2018) assumes that CF I is not essential for 3' processing; this tacit assumption is now confirmed by our data.

RBBP6 was not identified in the original resolution and reconstitution experiments that uncovered the roles of all other 3' processing factors. A proposed role in 3' processing (Di Giammartino et al., 2014; Shi et al., 2009) is confirmed by our experiments. The DWNN and zinc knuckle domains suffice for RBBP6 function; the RING domain and, thus, a potential ubiquitin ligase activity is not essential. Activation of the endonuclease activity of CPSF73 must be tightly coupled to its incorporation into a 3' processing complex assembled on the correct RNA site. Activation very likely involves a substantial structural rearrangement to open the active site for accommodation of a substrate RNA, as seen for the histone cleavage complex (Sun et al., 2020). As the RBBP6 DWNN domain is bound near the active site of CPSF73, we speculate that it might help to induce this structural change or might stabilize an open conformation (**Fig. 7**). Similarly, yeast Mpe1 has recently been proposed to couple endonuclease activation to poly(A) site recognition (Rodriguez-Molina et al., 2021).

In agreement with the original observations (Moore and Sharp, 1985), pre-mRNA cleavage was found to depend on ATP but not cleavage of its phosphoanhydride bonds. ATP very likely serves as an allosteric effector for hClp1 since several mutations in the ATP binding site of this protein increase the concentration of ATP required for RNA cleavage, one mutation allows significant cleavage in the absence of ATP, and another mutation modulates the nucleotide specificity, permitting better use of ITP both in RNA cleavage and in the RNA 5'

kinase reaction. An allosteric effect of ATP bound to hClp1 would be consistent with reports of mutations in the ATP binding site of yeast Clp1 affecting the interactions with other 3' processing factors, notably Pcf11 (Ghazy et al., 2012; Haddad et al., 2012; Holbein et al., 2011). However, all our hClp1 mutants could be purified in a stable complex with hPcf11. While cross-linking provided evidence of oligomerization of CF II in the context of the cleavage reaction, ATP did not induce the formation of larger CF II complexes as assayed by gel filtration (**data not shown**). The yeast pre-mRNA 3' cleavage reaction is independent of ATP addition (Hill et al., 2019). However, purified yClp1 contains tightly bound ATP (Ghazy et al., 2012; Noble et al., 2007); in one report, 57% of purified yClp1 had ATP bound (Holbein et al., 2011). Thus, a function of ATP in yeast pre-mRNA 3' processing might be provided by the Clp1-bound nucleotide. The K_A for ATP is ~0.5 µM, four orders of magnitude below the typical intracellular concentration (Berg et al., 2019). Thus, the relevant ATP binding site will be occupied at all times, and variations in ATP binding are unlikely to serve a regulatory purpose. It is possible that ATP is a constitutive ingredient of the 3' processing reaction with no regulatory significance. However, ATP hydrolysis, a reaction similar to the kinase activity of hClp1, might facilitate some aspect of the processing reaction that is not assayed in our experiments, in which non-hydrolyzable analogs were functional.

Acknowledgements: We thank Anne-Katrin Hoffmeister and Gudrun Scholz for technical assistance; Erik Fiedler for suggesting the use of TMAO; David Bentley, Walter Keller, David G. Skalnik and Steven West for cDNAs; Daniel Bollschweiler and Tillman Schäfer at the MPIB cryo-EM facility for help with EM data collection; Barbara Steigenberger at the MPIB biochemistry core facility for mass spectrometry; Christian Benda and J. Rajan Prabu for providing computational infrastructure for EM data processing; Daniela Wartini for help with cell culture; Fabien Bonneau, Courtney Long and members of the Conti lab for discussions. This work was supported by a grant from the German Research Foundation to EW and by funding from the Max Planck Gesellschaft, the European Commission (ERC Advanced

Investigator Grant EXORICO), the Novo Nordisk Foundation (Exo-Adapt) and the German Research Foundation (DFG Sonderforschungsbereich 1035, Projektnummer 201302640 and SFB/TRR 237) to E.C.

Author contributions:

Conceptualization, F. S., U. K., and E. W.; Methodology, F. S., U. K. and E. W.; Investigation, M. S., F. K., F. S., P. S., C. T., C. I., and U. K.; Writing - original draft, E. W., Writing – review and editing, F. S., U. K., E. C. and E. W. with input from all authors; Supervision, U. K., E. C., and E. W.; Funding acquisition, E. C. and E. W.

Data availability:

EM data will be deposited in the EMDB under the accession code EMD-xxxxx.

Declaration of interests:

The authors declare no competing interests.

Figure legends:

Fig. 1: Reconstitution of pre-mRNA 3' processing from sixteen purified polypeptides.

(A) Schematic representation of proteins participating in pre-mRNA cleavage.

(B) Proteins used in this study were separated by SDS-PAGE and stained with Coomassie. CPSF160 and WDR33 are not resolved, and CPSF30 runs as a doublet. Two versions of the cleavage module are shown, with or without CstF64 (CM3 or CM4; see main text). In the RBBP6 preparation, all major bands reacted with anti-PRPP6 in western blotting.

(C) Kinetics of cleavage of the SV40 late RNA in the presence of WT PAP and 3'-dATP. Left, reaction carried out with the mix of recombinant proteins described in the text. Right, reaction in nuclear extract. The substrate RNA (wild-type or mutant) is shortened by a few nucleotides from the 3' end in nuclear extract. Downstream fragments are shown in a longer exposure of the bottom part of the same gel. Quantification in the bottom panel (n = 3).

(D) Kinetics of cleavage and polyadenylation in the presence of ATP vs dATP. The SV40 late RNA was processed by the same protein mix with either 3'-dATP (left) or ATP (right). Only the upstream cleavage product is shown. Reactions were analyzed directly or after RNase H/oligo(dT) treatment. Brackets indicate the digestion products carrying oligo(A) stubs. Percent cleavage is shown at the bottom.

(E) Analysis of 5' cleavage fragment. A cleavage time course with uncapped RNA (right) reveals a single 5' product whereas capped RNA (left) generates two products, presumably reflecting incomplete capping. Products co-migrate with the one generated in nuclear extract. A partial alkaline hydrolysis ladder of an end-labeled RNA in the first lane demonstrates single-nucleotide resolution of the gel. This lane was cut from the same gel, and contrast was enhanced to reveal individual bands.

(F) Analysis of 3' cleavage fragment. Cleavage time courses were carried out with unlabeled substrate RNA and either wild-type (WT) CM3 or CM3 containing inactive CPSF73 (MUT). The cleavage site was mapped by primer extension. Products corresponding to uncleaved RNA

and the major 3' cleavage product are indicated. Minor cleavage products are labeled with asterisks. For the first lane, a synthetic 3' cleavage fragment was used as template.

Fig. 2: CF I is not an essential 3' processing factor

(A) CF I is the only non-essential component of the reconstituted cleavage reaction. The SV40 late precursor RNA was processed in nuclear extract (NXT) or with a mixture of recombinant proteins (PM). Individual components listed at the top were left out.

(B) Stimulation of cleavage by CF I depends on UGUA motifs. SV40 late and L3 RNAs with their UGUA motifs are sketched at the top. In L3-2xmut, the first two UGUA motifs were mutated to UCUA. In L3-3xmut, all three motifs were mutated. With each RNA, a titration of CF I was carried out. Cleavage efficiencies are presented on the right (average of n = 3 +/- SD). Cleavage efficiencies without CF I addition were 2.2 ± 0.3% for L3, 1.8 +/- 0.4% for L3-2xMUT and 1.9 +/- 0.4% for L3-3xMUT.

Fig. 3: Two structured domains of RBBP6 are sufficient for pre-mRNA cleavage

(A) Top panel: Cartoon of the domain structure of RBBP6. Domain boundaries (amino acid numbers) are indicated at the top and deletion boundaries at the bottom. Bottom panel: SDS polyacrylamide gel analysis of RBBP6 preparations.

(B) Titration of RBBP6 deletion variants in cleavage assays. 'Full-length' was MBP-RBBP6 with the MBP tag cleaved off; 'full-length plus CF I' was RBBP6 from a co-expression with CF I. A quantification is shown at the bottom.

(C) Cleavage activity of $RBBP6_{1-780}$ carrying point mutations.

(D) $RBBP6_{1-780}$ was titrated in a filter-binding assay with SV40 late RNA. One experiment of two is shown (compare Table S1).

(E) RNA binding activity co-purifies with RBBP6. Top panel, UV profile and RNA binding activity of the final Resource Q column of an $RBBP6_{1-780}$ purification. Bottom panel, SDS-PAGE of the same fractions. The main RBBP6 band is marked.

Fig. 4: RBBP6 contacts the ß-lactamase domain of CPSF73

(A) Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that RBBP6 directly interacts with the CPSF complex and, albeit weaker, the cleavage module.

(B) Segmentation of the filtered cryo-EM map of the eight-subunit CPSF-RBBP6 complex. The proximal and distal lobes of the cleavage module are indicated.

(C) Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that RBBP6 directly interacts with CPSF73 and not CPSF100.

(D) Cross-links between $CPSF73_{1-458}$ and $RBBP6_{1-81}$ mapped onto the computationally predicted model of the complex. Intra- and intermolecular cross-links are in black and blue, respectively, with the thickness of the line indicating their score (thicker = higher score), and the respective residues indicated as spheres. A cross-link to an unmodeled region is shown in white (placed to the closest visible Cα atom). The table lists measured distances for cross-links visualized in the left panel.

(E) Computationally predicted structural model of the $CPSF73_{1-458}$-$RBBP6_{1-81}$ complex with close-up of the putative binding interface. Approximate location of the CPSF73 active site is indicated by a red dashed circle. Possible ionic interactions are indicated with black dotted lines. Labeled positively charged RBBP6 residues were changed to aspartic acid.

(F) Coomassie-stained SDS-PAGE analysis of pull-down experiment showing that reverse-charged mutations in the predicted RBBP6-CPSF73 interface disrupt the interaction of RBBP6 with the cleavage module and CPSF complex.

(G) $RBBP6_{1-335}$, either wild-type or containing the same reverse-charge mutations as in (F), was titrated into cleavage reactions.

Fig. 5: Composition of CstF and the cleavage module active in pre-mRNA cleavage

(A) Left, different versions of CstF and its subunits were analyzed by SDS PAGE and Coomassie staining. Right, CstF (64/τ) with a FLAG tag on CstF64τ and, as a control, CstF (64/64) without a FLAG tag were used in a FLAG pull-down experiment. I, input; FT, flow-through; E1, E2, FLAG peptide eluates; B, material remaining on the beads eluted with SDS

sample buffer. Proteins were detected by Coomassie staining. The identity of the band running at the CstF64τ position in CstF (64/64) is unknown; based on western blots, it is neither 64τ nor a fragment of CstF77.

(B) Both CstF64 and 64τ are active in pre-mRNA cleavage, and CstF50 is essential. Proteins shown in (A) and their combinations were tested in pre-mRNA cleavage assays.

(C) Both CM3 and CM4 function in pre-mRNA cleavage. The two protein complexes, each with wild-type (+) or point-mutated (-) CPSF73, were tested in cleavage assays. The SV40 late Δ RNA was tested as a negative control.

Fig. 6: Pre-mRNA cleavage depends on ATP

(A) Creatine phosphate is not required for cleavage. Pre-mRNA cleavage reactions were carried out with nuclear extract (NXT), a mixture of purified processing factors (PM) or the same mixture supplemented with creatine kinase (5 µg/ml). In addition to the presence of 0.5 mM 3'-dATP, individual reactions were supplemented with inorganic phosphate (pH 8.0), creatine, a mixture of both, phosphocreatine or phosphoserine as indicated, each at 20 mM. A representative experiment of n = 5 is shown.

(B) ATP is essential for cleavage, but cleavage of phosphoanhydride bonds is not. Reactions contained the nucleotides indicated, each at 0.5 mM, and either WT PAP or PAP D115A. With WT PAP, cleavage in the presence of CTP, GTP or ITP was not visible, presumably due to heterogeneous limited extension of the cleavage product. A representative experiment of n = 3 is shown.

(C) A cleavage reaction lacking CF I is still ATP-dependent. Cleavage reactions were carried out with WT PAP +/- CF I as indicated. 3'-dATP was titrated between 3.9 and 1000 µM.

(D) Mutations in the active site of poly(A) polymerase do not affect the ATP dependence of RNA cleavage. Reactions contained WT PAP or mutants as indicated. 3'-dATP was titrated between 2 and 1000 µM in geometric progression. A quantification (n = 3, highest ATP concentration omitted) is shown on the right. Data were fitted to the Michaelis-Menten equation.

(E) Mutant hClp1 can be purified in a stable complex with hPCF11. The Coomassie-stained SDS PAGE shows the peak fraction from each Mono Q column. Based on western analysis of comparable wild-type preparations, most additional bands are breakdown products of hPcf11.
(F) Mutations in the ATP binding site of hClp1 affect the RNA 5' kinase activity of CF II. CF II preparations shown in (**E**) were tested in RNA kinase assays with ATP or ITP. The top panel shows representative time courses. The bottom shows a quantification averaged from n = 3.
(G) Mutations in the ATP binding site of hClp1 affect ATP dependence of RNA cleavage. CF II preparations shown in (**E**) were tested in cleavage assays with mutant PAP. ATP was titrated from 0.12 µM (WT) or 1.95 µM (mutants) to 500 µM. Top panel, representative experiment. Bottom panel, average of n = 3. Curves were fitted to the Michaelis-Menten equation.
(H) A mutation in the ATP binding site of hClp1 changes the nucleotide specificity of cleavage. The protein preparations indicated were tested as in (**G**) except that ITP was titrated. Top panel, representative experiment. Bottom panel, average of n = 3. Curves were fitted to the Michaelis-Menten equation.

Fig. 7: Model of the reconstituted pre-mRNA cleavage complex

RBBP6 and ATP are proposed to stabilize the activated conformation of the complex. CF I as a non-essential factor is not shown as part of the core complex, but it stimulates processing by interaction with hFip1 (Zhu et al., 2018).

STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by Elmar Wahle at ewahle@biochemtech.uni-halle.de and/or Elena Conti at conti@biochem.mpg.de

Transfer of all MultiBAC constructs is restricted by an MTA between ETH Zürich and Martin-Luther University Halle-Wittenberg.

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
E. coli DH10B	Invitrogen (Thermo Fisher)	Cat# 18297-010 (2011)
E. coli DH10B EMBACy	ETH, Zurich	
E. coli DH10B BAC-mCherry	This work	
E. coli BL21(DE3) Codon plus-RIL	Stratagene (Agilent)	
E. coli BL21 BW23473	ETH, Zurich	
Cell lines		
Sf21	Gibco	
High5	Invitrogen	
HEK293T	ATCC	
Chemicals		
Hela cell nuclear extract	Ipracell, Belgium	
NiNTA Agarose	Qiagen	
Prepacked Chromatography columns for protein purification	GE Healthcare (Cytiva)	
Glutathion-Sepharose	GE Healthcare (Cytiva)	Cat#17075601
Heparin Sepharose	GE Healthcare (Cytiva)	Cat#17046701
DEAE-Sepharose CL6B	GE Healthcare (Cytiva)	Cat#17071001
StrepTrap HP column	Cytiva	Cat#28907548
MonoQ 5/50 GL column	Cytiva	Cat#17516601
Superose6 prep grade	GE Healthcare (Cytiva)	Cat#17048903
Superose 6i 10/300	Cytiva	Cat#29091596
Superdex Peptide 3.2/300	GE Healthcare (Cytiva)	Cat#29036231
Superdex 200 HR10/30	GE Healthcare (Cytiva)	Cat#17108801
Superdex 200i 3.2/300	Cytiva	Cat#28990946
Dynabeads Protein G beads	Invitrogen	Cat#10004D
Anti-Flag M2 affinity gel	Sigma Aldrich	Cat#A2220
Flag-Peptide	Sigma Aldrich	Cat#F3290
Leupeptin, Pepstatin. PMSF	Roche Diagnostics	
AEBSF	AppliChem	Cat#A1421
cOmplete, EDTA-free protease inhibitor cocktail	Roche	Cat#05056489001
Radioactive ^{32}P-labeled ribonucleotides	Hartmann Analytics	
3'-dATP (Cordycepin triphosphat)	Sigma Aldrich	Cat# C9137
ITP, ATP analogs	Jena Biocscience	

m2 7,3'-OGP3G (ARCA Cap analog)	Jena Biocscience	Cat#NU-855-5
HEPES, DTT	Gerbu Biotechnik GmbH	Cat#1009, Cat#1008
PVA (30-70 kD), PEG 6000	Merck	
TMAO	Sigma Aldrich	Cat#92277
Glycogen	Serva	Cat#39766
Glycerol	Roth	
tRNA E. coli MRE 600	Roche Diagnostics	
Proteins		
bovine His-PABPN1	(Kühn et al., 2009)	
bovine-His-PAP full-length	(Kühn et al., 2009)	
bovine-His-PAP513	(Kerwitz et al., 2003)	
CF I	(Schäfer et al., 2018)	
CF II and variants	(Schäfer et al., 2018) and this work	
CstF	(Schäfer et al., 2018)	
Additional recombinant proteins	See Supplemental Table 3	See Supplemental Tables 3/4
RNase Inhibitor, murine	New England Biolabs	Cat#M0314L
Restriction enzymes	New England Biolabs	
Cre Recombinase	New England Biolabs	Cat#M0298S
Q5 DNA Polymerase	New England Biolabs	Cat#M0491L
Pwo DNA Polymerase	Peqlab/VWR	Cat#PEQL01-5010
Gibson Assembly Master Mix	New England Biolabs	Cat#E2611L
Phosphatases	New England Biolabs	
Trypsin (sequencing grade)	Promega	Cat# V5111
SP6 RNA Polymerase	Roche Diagnostics, NEB	#10810274001, Cat#M0314S
Oligonucleotides		
primer	Invitrogen	
Primer sequences	This work	See Supplemental Table 6
Recombinant DNA		
BigBAC vectors	Jan-Michael Peters	Addgene Cat#1000000088
MultiBac vectors	ETH Zurich	
CM3	This work	See Supplemental Table 5
CM4	This work	See Supplemental Table 5
mPSF	This work	See Supplemental Table 5
CF I	(Schäfer et al., 2018) and this work	See Supplemental Table 5
CF II and variants	(Schäfer et al., 2018) and this work	See Supplemental Table 5
CstF (64/64)	(Schäfer et al., 2018) and this work	See Supplemental Table 5
CstF (τ/τ)	This work	See Supplemental Table 5
CstF (64/τ)	This work	See Supplemental Table 5
CstF Δ50	This work	See Supplemental Table 5
CstFτ Δ50	This work	See Supplemental Table 5
CstF50	This work	See Supplemental Table 5
mCTD	This work	See Supplemental Table 5
pLIB-Rbbp6 and variants	This work	See Supplemental Table 5
PP1 complex	This work	See Supplemental Table 5
PP1α	This work	See Supplemental Table 5

bovine His-PAP in pET19b	(Kühn et al., 2009)	pUK-His PAP wt
bovine His-PAP513 in pET19b	(Kühn et al., 2009)	pUK His-PAP513
bovine-His-PAP D113A or D115A in pET28	This work	pGS-His-PAP D113A pGS-HisPAP D115A
bovine-His-PAP513 D113A or D115A in pET28	This work	pGS-His-PAP513 D113A pGS-HisPAP513 D115A
hSsu72 in pET-SUMO	This work	pET-SUMOadapt-hSsu72
Rbbp6 and variants in pET30	This work	pET30-HF-Rbbp6
GST-mPol II CTD in pET21	This work	pET21-GST-mCTD
Antibodies against		
CTD Ser-2-P (CTD7 3E10)	Dirk Eick; Helmholtz Zentrum München	
anti-FLAG M2	Sigma Aldrich	Cat#F3165
Software		
ImageQuant 5.0	MolecularDynamics/GE HealthCare	
Merox v. 2.0.1.4	(Gotze et al., 2015)	
SigmaPlot 12.5	Systat Software GmbH	
MSConvert	(Chambers et al., 2012)	
xiView	(Graham et al., 2019)	
MaxQuant 1.6.	(Cox and Mann, 2008)	
SnapGene 4.2	Insightful Science	
SerialEM	(Schorb et al., 2019)	
Focus	(Biyani et al., 2017)	
Relion v3.1	(Zivanov et al., 2018)	
Cryosparc v3.1	(Punjani et al., 2017)	
UCSF ChimeraX v1.2.5	UCSF; (Pettersen et al., 2021)	
PyMOL v2.4.2	PyMOL Molecular Graphics System, Schrödinger, LLC	
Instrumentation		
Orbitrap Q-Exactive Plus	Thermo Fisher	
Ultimate 3000 RSLCnano-HPLC system	Thermo Fisher	

Expression clones

Cloning mostly followed standard ligation-dependent procedures using enzymes from NEB. Reverse transcription of human ORFs was done using purified total RNA from HEK293 cells as described (Schönemann et al., 2014). New flanking restrictions sites were introduced by PCR using proof-reading polymerases (Pwo, Peqlab; Q5, NEB) and appropriate DNA primers (Invitrogen). Point mutations were introduced by the QuickChange protocol (Agilent) or as described by (Picard et al., 1994). All PCR products were verified by DNA sequencing (Eurofins). In some cases, ORFs were subcloned into additional vectors for fusion with N-terminal peptide tags or additional restriction sites before transfer into the MultiBac system.

Alternatively, peptide tags were introduced into plasmids by ligation of phosphorylated oligonucleotides with compatible overhangs.

Plasmids and protocols for the MultiBac system were those described in (Bieniossek et al., 2008; Fitzgerald et al., 2006). Plasmids and protocols for the BigBac system are described in (Weissmann and Peters, 2018; Weissmann et al., 2016) (Addgene). Individual ORFs were integrated into plasmids either by restriction digestion and ligation or by Gibson assembly (NEB). If more than two expression cassettes were combined in plasmids, this was performed either using DNA ligase and restrictions sites in the multiplication modules of the MultiBac plasmids or plasmids were fused by loxP-dependent Cre recombinase (NEB) (Bieniossek et al., 2008). The transfer vector for the PP1alpha complex was generated by Gibson assembly of four gene cassettes and the SwaI-opened pBIG1 vector according to (Weissmann and Peters, 2018) with minor variations.

All cDNAs and proteins used in this study are listed in **Supplemental Table 3**. Affinity tags and other additional amino acid sequences are reported in **Supplemental Table 4**. Oligonucleotides used for mutagenesis and reverse transcription are listed in **Supplemental Table 6**. Detailed information about the strategy used for cloning of individual open reading frames as well as DNA oligonucleotides used will be provided on request.

Cell culture and virus propagation

Suspension cultures of Sf21 cells (Invitrogen) were grown at 27.5 °C on a rotary shaker in ExCell 420 serum-free medium (Sigma Aldrich) and diluted to 0.5 x 10^6 cells/ml every 2 – 3 days. Cells were transfected with bacmids either with Cellfectin II (Invitrogen), and the supernatant (virus P1) was harvested after 3 days, or with Fugene HD (Promega), and the supernatant was harvested after 5 days. For virus propagation, cells were infected with 0.02 to 0.1 % culture volume of P1 and cultivated for up to 7 days. For protein expression, cells were infected at 1.5 x 10^6 cells/ml with 1 % culture volume of the propagated virus and harvested 68 to 72 h later. Viruses for protein production are listed in **Supplemental Table 5.**

Protein purification

All proteins used for the reconstitution reaction were human except where noted. Procedures for the expression and purification of the following proteins have been described: CF I (Schäfer et al., 2018), full-length bovine PAP and bovine PABPN1 (Kühn et al., 2009). Bovine PAP_{1-513} was purified according to (Martin and Keller, 1996) except that a Resource Q column was used for the second step. CF II was purified as described (Schäfer et al., 2018) except that the final gel filtration column was omitted. All protein purifications described below were carried out under refrigeration. The pH of buffers was adjusted at room temperature in 1 M stock solutions. For Ni-NTA chromatography, the pH was re-adjusted after refrigeration in the final buffer. Ni-NTA beads were from Qiagen, anti-FLAG M2 agarose from Sigma, and all other column materials and pre-packed columns from GE Healthcare. Sf21 cells were harvested by centrifugation, resuspended in the buffer specified for each protein, and lysed by sonification with a Branson Sonifier, typically with 5 x 20 bursts at medium setting. Bacteria were lysed with a Constant Systems (Daventry, UK) device. Lysates were centrifuged for 30 min at 20,000 g.

Purification of mPSF was modified from (Schönemann et al., 2014). Cells from a 4 l culture were resuspended in 120 ml mPSF buffer (50 mM TEA pH 8.0, 10% sucrose, 0,5 mM DTT, 0.2 mM AEBSF, 2 µg/ml leupeptin, 2 µg/ml pepstatin) containing 250 mM KCl and lysed as described above. The cleared lysate was adjusted to pH 8.0 by the addition of KOH and loaded on a 30 ml DEAE-Sepharose CL6-B column equilibrated in mPSF buffer plus 250 mM KCl. The flowthrough was agitated overnight with 0.3 ml Ni-NTA-agarose equilibrated in mPSF buffer plus 250 mM KCl, 5 mM imidazole. Beads were washed with 3 x 3 ml mPSF buffer plus 250 mM KCl, 10 mM imidazole, and protein was eluted with 4 x 3 ml of the same buffer plus 500 mM imidazole. The first three fractions were pooled, dialysed for 2 x 1 h against mPSF buffer plus 250 mM KCl and bound to 1 ml anti-FLAG beads for 2 h with agitation. Beads were washed with 10 ml mPSF buffer plus 250 mM KCl, and protein was eluted with 3 x 0.5 – 1 ml of the same buffer containing 0.15 mg/ml FLAG peptide. Fractions were pooled and dialysed

for 2 x 1.5 h against mPSF buffer plus 250 mM KCl containing 25 mM Hepes-KOH, pH 8.0, instead of TEA.

CstF was purified from 800 ml overexpressing Sf21 cells by a modification of (Schäfer et al., 2018). Cells were resuspended in 50 ml CstF buffer (50 mM TEA pH 8.0, 10% sucrose, 1 mM DTT, 1 mM EDTA, 1 mM PMSF) plus 200 mM KCl, 10 mM imidazole, and lysed as described above. The cleared lysate was agitated with 3 ml Ni-NTA-agarose. Beads were washed for 15 minutes with 10 column volumes of CstF buffer plus 200 mM KCl, 25 mM imidazole. Protein was eluted for 2 x 20 - 30 minutes with 15 ml CstF buffer plus 200 mM KCl, 250 mM imidazole. Elution fractions were pooled, diluted to 75 mM KCl with CstF buffer, and loaded on a 1 ml Mono Q column. The column was eluted with gradients of 20 column volumes from 75 to 500 mM KCl and of 5 column volumes to 1 M KCl, both in CstF buffer.

CstFΔ50 complexes were expressed in 400 ml Sf21 cells. Cells were resuspended in CstFΔ buffer (50 mM Tris pH 8.0, 10 mM imidazole, 10% glycerol, 0.02% NP-40, 1 mM DTT, 1 mM PMSF, 1 µg/ml leupeptin and 1 µg/ml pepstatin) plus 300 mM KCl and 10 mM imidazole. Cleared lysates were agitated with 300 µl Ni-NTA beads. Beads were washed three times with 10 column volumes of CstFΔ buffer plus 300 mM KCl and 10 mM imidazole. Protein was eluted with five times 2 column volumes CstFΔ buffer plus 200 mM KCl and 500 mM imidazole. Elution fractions were pooled, dialysed against CstFΔ buffer plus 75 mM KCl, and loaded on a 1 ml Resource Q column. The column was washed with the same buffer. Proteins were eluted with gradients of 15 column volumes from 75 to 500 mM KCl and 5 column volumes to 1 M KCl, all in CstFΔ buffer.

Sf21 cells expressing CM3 or CM4 were resuspended in CM buffer A (50 mM Tris-HCl pH 7.5, 10 % sucrose, 5 µM $ZnCl_2$, 5 mM DTT, 1 mM PMSF, 1 µg/ml leupeptin and 1 µg/ml pepstatin) plus 75 mM KCl and lysed as described above. The cleared lysate was loaded on a 30 ml DEAE-Sepharose column. The column was washed with 2 column volumes CM buffer A plus 75 mM KCl, and proteins were eluted with a three column volume gradient to 600 mM KCl and a final step of 1 M KCl, both in CM buffer A. Fractions containing cleavage module were pooled, dialysed against CM buffer B (20 mM Hepes-KOH, pH 7.9, 10 % sucrose, 5 µM

$ZnCl_2$ and 2.5 mM DTT) plus 75 mM KCl and loaded on a 6 ml Resource S column. The flow-through was loaded on two combined 5 ml HiTrap Heparin-Sepharose columns in CM buffer B plus 75 mM KCl. Proteins were eluted with a fifteen column volume gradient to 400 mM KCl and a final step of 1 M KCl in CM buffer B. Fractions containing cleavage module were pooled and precipitated for 30 min on ice with ammonium sulphate at 50% saturation. The precipitate was collected by centrifugation (20 min, 20,000 g), dissolved in CM buffer B plus 200 mM KCl with a reduced DTT content (1 mM), and proteins were separated on a 49 ml Superose 6 column in the same buffer.

Sf21 cells expressing CDC73 were resuspended in CDC73 buffer (50 mM potassium phosphate, pH 7.5, 10 % glycerol, 1 mM DTT, 1 mM PMSF, 1 µg/ml leupeptin and 1 µg/ml pepstatin) plus 300 mM KCl, 10 mM imidazole, and lysed as described above. The cleared lysate was agitated for 2 h with 0.5 ml Ni-NTA agarose. Beads were washed twice with 10 column volumes CDC73 buffer plus 300 mM KCl, 10 mM imidazole, and the protein was eluted with CDC73 buffer plus 300 mM KCl, 500 mM imidazole. 950 µl of the main elution fraction was diluted with CDC73 buffer to 100 mM KCl and loaded on a 1 ml Resource Q column. The column was washed with 2 column volumes CDC73 buffer plus 100 mM KCl and eluted with a 15 column volume gradient to 1 M KCl in CDC73 buffer.

E. coli BL21-CodonPlus expressing his-tagged SUMO-Ssu72 were grown in SB-medium at 37°C. Expression was induced with 500 µM IPTG at OD_{600} = 1 and incubation continued overnight at 16°C. Cells were harvested, resuspended in 500 mM NaCl, 100 mM Tris-HCl, pH 8.0, and frozen in liquid nitrogen. Purification on a Ni-NTA column was carried out as described (Zhang et al., 2011). Elution fractions containing the Ssu72 fusion protein were pooled, mixed with his-tagged SUMO-protease, and dialyzed against 100 mM NaCl, 20 mM Tris-HCl, pH 8.0, 10 mM β-mercaptoethanol. The dialysate was agitated for 2 h with Ni-NTA agarose. The flow-through containing SUMO-free Ssu72 was collected.

Sf21 cells expressing RBBP6 or its variants were resuspended in RBBP6 buffer (50 mM Tris/HCl pH 8.0, 10 % sucrose, 1 mM DTT, 10 µM $ZnCl_2$,1 mM PMSF, 1 µg/ml pepstatin, 1 µg/ml leupeptin) plus 200 mM KCl, 10 mM imidazole and lysed as described above. Cleared

lysates were bound for 2 – 4 h to Ni-NTA-Agarose beads equilibrated in the same buffer. The beads were washed with the same buffer, and protein was eluted with the same buffer containing 250 mM imidazole. For further purification of the deletion variants, relevant elution fractions were pooled, dialysed against RBBP6 buffer containing 50 mM KCl, centrifuged and loaded on a Resource Q column. The column was eluted with three consecutive gradients: 2 column volumes to 200 mM KCl, 10 column volumes to 400 mM KCl and 3 column volumes to 1 M KCl, all in RBBP6 buffer. Full-length constructs tagged with MBP, MBP-SNAP, or GFP as well as the His-FLAG-tagged ΔRING mutant were pooled after Ni-NTA purification, diluted to a final concentration of 150 mM KCl with RBBP6 buffer, and loaded on a Resource Q column. The column was eluted with a ten column volume gradient up to 400 mM KCl and then three column volumes up to 1 M KCl. MBP and GFP tags were cleaved with PreScission protease and the MBP-SNAP tag with TEV-protease. Proteases were used at a 1:50 ratio for two hours at room temperature. *E. coli* BL21-CodonPlus cells expressing shortened RBBP6 variants were harvested, resuspended in RBBP6 buffer plus 200 mM KCl and lysed. Purification was as described for proteins expressed in Sf21 cells except that 10 % glycerol replaced sucrose.

Sf21 cells expressing XRN2 were resuspended in XRN2 buffer (50 mM Tris/HCl pH 7.5, 10 % sucrose, 3 mM $MgCl_2$, 1 mM DTT, 10 µM $ZnCl_2$, 1 mM PMSF, 1 µg/ml pepstatin 1 µg/ml leupeptin) containing 200 mM KCl, 10 mM imidazole, and lysed as described above. The protein was purified on a Ni-NTA column as described for RBBP6. Elution fractions containing XRN2 were pooled, dialysed against XRN2 buffer plus 75 mM KCl, and loaded on a Resource Q column. The protein was eluted with two gradients of 15 column volumes to 500 mM KCl and 3 column volumes to 1 M KCl in the same buffer.

Purification of the PP1α complex from Sf21 cells was carried out as described for full-length Rbbp6 except that Tris-HCl was used at pH 7.5 and wash buffer for the Ni-NTA column contained 20 mM imidazole.

The RNA polymerase II CTD with an N-terminal GST fusion was expressed in *E. coli* BL21-CodonPlus or Sf21 cells. Bacteria were harvested, resuspended in 100 mM KCl, 50 mM Tris-HCl pH 7.5, 10% glycerol, 1 mM DTT, 0.5 mM PMSF, 0.8 µg/ml pepstatin, 0.5 µg/ml

leupeptin and lysed. Sf21 cells were resuspended in the same buffer and lysed. In both cases, cleared lysates were bound to equilibrated GSH-Sepharose. Elution was carried out in lysis buffer plus 10 mM reduced glutathione. Fractions containing the desired protein were loaded on a Superdex 200 HR10/30 size exclusion column. The column was run in lysis buffer containing 5 mM DTT and lacking glutathione. For dephosphorylation, 2 µg of a fraction from the final Superdex column was incubated with 2.5 units Quick CIP (NEB) at 37 °C. Aliquots were withdrawn at different time points and analysed by SDS polyacrylamide gel electrophoresis and Coomassie staining or by western blotting with an antibody specific for Ser2P (CTD7 3E10) (Chapman et al., 2007).

The concentration of $RBBP6_{1-340}$ was determined in a spectrophotometer at 280 nm. Concentrations of all other proteins were determined by densitometry (ImageQuant 5.0; Molecular Dynamics) of Coomassie R250-stained bands in SDS polyacrylamide gels and comparison to BSA standards.

Mammalian cell protein expression and purification

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). Three stable cell lines were generated expressing either RBBP6 (RBBP6 residues 1-335 N-terminally fused to a TwinStrep tag and a 3C protease cleavage site), the cleavage module (TwinStrep-3C-Symplekin, CPSF100, CPSF73, CstF64), or CPSF (TwinStrep-3C-Symplekin, CPSF100, CPSF73, CstF64, CPSF160, WDR33 residues 1-413, CPSF30). Additionally, RBBP6 variants carrying an N-terminal TwinStrep-3C tag and a C-terminal FLAG tag (RBBP6 residues 1-335, wildtype and mutant forms) were transiently expressed using polyethyleneimine.

For protein expression, cultures were adjusted to a density of $1x10^6$ cells per mL in FreeStyle 293 expression medium (Gibco, Thermo Fisher Scientific). The cells were induced with doxycycline and harvested 48 h after induction. For all purifications, lysis buffers were supplemented with EDTA-free complete Protease Inhibitor Cocktail (Roche), DNase I and

Benzonase before cells were lysed with a glass dounce homogenizer and cleared by centrifugation.

All RBBP6 variants were purified in 1x DPBS supplemented with 2 mM DTT and 10 µM $ZnCl_2$. Cleared lysate was applied onto a StrepTrap HP column (Cytiva), washed with 20 column volumes of buffer and eluted using 5 mM desthiobiotin in wash buffer. The N-terminal TwinStrep tag was removed by addition of 3C protease 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.

The four-subunit cleavage module was purified in 1x DPBS supplemented with 2 mM DTT, 10 µM $ZnCl_2$ using a StrepTrap HP column. After washing, bound proteins were eluted with 5 mM desthiobiotin in wash 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.

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. After extensive washing, bound proteins were eluted with 5 mM desthiobiotin in wash buffer. Protein preparations used for cryo-EM were concentrated and flash frozen in liquid nitrogen. All other protein preparations were applied onto 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.

RNAs

The following RNAs have been described: cleavage substrate SV40 late and its Δ derivative (Schäfer et al., 2018); L3 and its Δ derivative with a point mutation in the AAUAAA signal (Humphrey et al., 1987); ‘pre-cleaved’ RNA L3pre and its Δ derivative (Christofori and Keller,

1989). Mutations in the UGUA motifs of L3 were introduced by means of synthetic L3 DNA fragments corresponding to bp 4 - 369 in pSP64-L3 (Humphrey et al., 1987), counting from the transcription start point of transcription. One such fragment contained G to C base substitutions in the UGUA-motifs at vector positions 144 and 155 (double mutant; GENWIZ), a second fragment contained an additional mutation at position 196 (triple mutant; Eurofins). Synthetic fragments were digested with BamHI/SacII and used to replace the corresponding fragment in pSP64-L3. The original L3pre RNA is lacking the upstream UGUA upstream motifs, containing vector sequences instead. Therefore, L3pre-v2 was generated to include the UGUA motifs: The original plasmids pSP64-L3 and pSP64-L3Δ as well as the derivatives carrying the UcUA mutations were used as templates for PCRs (forward primer: TCATATTGTCGTTAGAACGCG, reverse primer: TACAAATAAAAACATTTGCCT, reverse primer for UcUA triple mutant: TAgAAATA-AAAACATTTGCCT) using Q5 DNA polymerase (NEB) to generate SP6 promotor-containing L3pre DNA fragments that end at the position of 3'-cleavage (after bp 198, counting from the transcription start; this is one nucleotide 3' of the linearisation site in the original L3pre construct). The amplified DNA was purified, confirmed by sequencing and then used for run-off-transcription of L3pre V2 variants. RNA was synthesized as described (Schäfer et al., 2018) in the presence of reverse cap analog (ARCA; NEB, Jena Biosciences). A 5'-phosphorylated RNA of 51 nt corresponding to the expected downstream cleavage fragment of the SV40 late substrate (Sheets et al., 1987) was synthesized by Biomers.net GmbH, Ulm, Germany. The RNA was 3' end-labeled with RNA ligase and [^{32}P]-pCp.

Processing assays

Reaction conditions for pre-mRNA cleavage underwent some evolution during this project. Under the most recent standard conditions, cleavage assays contained 10 µL 2x cleavage buffer (40 mM HEPES-KOH, pH 8.0, 4 mM DTT, 2 mM $MgCl_2$, 1 mM ATP, 1 M trimethylamine oxide [TMAO]), 1 µg *E.coli* tRNA and up to 8 µL protein mix. The protein mix for each reaction contained 500 fmol each of CFI, CFII, CSTF, bPAP D115A, CM3; 60 fmol mPSF; 2500 fmol RBBP6 1-340 or other RBBP6 variants as indicated; and 4 U murine RNase

inhibitor (NEB). The pH of TMAO was adjusted to ~ 7.5 in a 6 M stock solution by the addition of HCl. Proteins were pre-diluted to 1 µM in CDB200 (50 mM HEPES-KOH, pH 8.0, 20 % glycerol, 1 mM DTT, 0.5 mM EDTA, 200 mM KCL, 0.1 % Tween-20, 0.2 mg/ml RNase-free BSA [Merck]). The total reaction volume was made up to 20 µL with CDB200. In earlier assays, WT PAP was used in combination with 3'-dATP to prevent polyadenylation; 3'-dATP was also used in reactions with nuclear extract (non-dialyzed nuclear extract obtained from Ipracell, Mons, Belgium; 4 µl/reaction). Also, 4% polyvinyl alcohol (PVA; 30 – 70 kDa) was initially used instead of TMAO and, with an older preparation, mPSF at 500 fmol/assay. Such deviations from standard conditions are noted in figure legends. Mixtures were assembled on ice in the absence of cleavage substrate, and reactions were started by the addition of 50 fmol substrate RNA and transfer to 30°C. After 1 h, they were stopped by the addition of 50 µl 2x proteinase K buffer (200 mM Tris-HCl, pH 7.9, 300 mM NaCl, 25 mM EDTA, 2% SDS), 10 µg each of glycogen and proteinase K and H_2O to a total volume of 100 µl. After 30 min at 37°C, RNA was ethanol-precipitated and analyzed by 6% polyacrylamide-urea gel electrophoresis and phosphoimaging. For the calculation of cleavage efficiency, substrate RNA, 5' cleavage fragment and background of each lane were quantified with ImageQuant 5.0. The background values were subtracted from the remaining substrate RNA and 5' fragment, and the amount of 5' fragment was corrected for the loss of radioactivity of the 3' fragment. Percent cleavage was calculated by the division of the corrected 5' fragment through the sum of 5' fragment and uncleaved substrate RNA. For ATP titrations, data for a single time point were fitted to the Michaelis-Menten equation (SigmaPlot 12.5).

For RNase H/oligo(dT) digestion, purified RNA was dissolved in 3 µl RNase H reaction mix (1x RNase H buffer [NEB], 40 pmol oligo$(dT)_{12}$, 4U murine RNase inhibitor [NEB], 1U RNAse H [NEB]) and incubated for 30 min at 37°C. Reactions were stopped by addition of 3 µl formamide loading buffer, and RNAs were separated on denaturating polyacrylamide gel.

For the analysis of the 3' cleavage fragment by primer extension, a 3x standard cleavage reaction was set up except that unlabeled SV40 late RNA was used and the amounts of this RNA and mPSF were doubled. RNAs were purified by phenol - chloroform extraction

and ethanol precipitation, dissolved in 3 µl H_2O, and 1 µl 0.5 µM primer (5' ^{32}P- labelled and gel-purified; CCCCCTGAACCTGAAACATA] and 1 µl dNTPs (10 mM each) were added. Mixtures were incubated for 5 min at 65°C and slowly cooled to room-temperature. 5 µl RT mastermix (2x reaction buffer, 20 mM DTT, 4U murine RNase inhibitor [NEB], 30 U ProtoScript II Reverse Transcriptase [NEB]) were added, and reactions were incubated for 1h at 42°C and then stopped by addition of 10 µl formamide loading buffer. RNAs were separated on a denaturing 10% polyacrylamide gel.

Polyadenylation reactions were done as described (Schönemann et al., 2014).

RNA binding assays

Nitrocellulose filter binding was done in a 40 µl volume essentially as described (Schäfer et al., 2018) except that the buffer was 20 mM Hepes-KOH, pH 7.9, 100 mM NaCl, 1 mM $MgCl_2$, 0.5 mM ATP, 0.04 % Tween 20, 2 mM DTT. Proteins were pre-diluted in 50 mM Hepes-KOH, pH 7.9, 100 mM NaCl, 10 % glycerol, 0.5 mM EDTA, 0.05 % Tween 20, 0.2 mg/ml BSA, 1 mM DTT. Affinity measurements were carried out with 0.1 nM SV40 late RNA. Higher RNA concentrations were used to follow RNA binding activity across column fractions. For K_D determination, the binding equilibrium was expressed as a quadratic equation and data were fitted in Microsoft Excel Solver.

RNA kinase assays

Assays were performed according to (Schäfer et al., 2018) except that NP40 in the reaction buffer was replaced by Tween-20 and each reaction contained 1 µM C_{14} RNA (3'-labeled by ligation to [5'-^{32}P]-pCp and gel-purified), 1 nM CFII and 0.5 mM nucleotide. The reactions were stopped by the addition of 1 volume 8 M Urea and analyzed on a denaturing 9% polyacrylamide gel.

Protein-protein interaction assays

Pull-down assays using purified proteins were performed using magnetic Protein G Dynabeads (Invitrogen, Thermo Fisher Scientific) coupled with anti-FLAG M2 antibody (Sigma F1365). Anti-FLAG antibody was added to beads equilibrated in wash buffer (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$, 0.01% NP-40 substitute). After incubation for 30 min rotating head-over at 4˚C, beads were washed three times and used immediately. 25 pmol FLAG tagged RBBP6 was mixed with an equal amount of putative binding partner (final concentration 1 µM each), and the reaction was incubated for 60 min at 4˚C in 20 mM HEPES-KOH pH 7.9, 55 mM NaCl, 1 mM $MgCl_2$, 20 mM creatine phosphate and 0.4 mM DTT. Equilibrated beads were added and incubated 60 min rotating head-over at 4˚C. The beads were washed three times with 20x resin volume of wash buffer before bound proteins were eluted with 0.2 mg/mL 3x FLAG peptide diluted in wash buffer.

For pull-down assays using HEK cell lysates, $5x10^6$ cells per condition were transiently transfected using polyethyleneimine, induced immediately with doxycycline and harvested after 48 h. Flash frozen cells were thawed and lysed in wash buffer containing 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 (Invitrogen, Thermo Fisher Scientific) and a KingFisher pull-down system (Thermo Fisher Scientific) operated at room temperature. After binding for 30 min, beads were washed four times before bound proteins were eluted in SDS-containing sample buffer.

Cryo-EM sample preparation, data collection and data processing

Before grid preparation, RBBP6 was run over a Superdex 200i 3.2/300 column (Cytiva) equilibrated in 20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$. The peak fraction was mixed with an equal amount of CPSF (final concentration at 1.5 µM) in final buffer containing 20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 20 mM KCl, 5 mM $MgCl_2$ and the reaction was incubated for 40 min at 4˚C. 0.04 % (v/v) n-octyl-ß-D-glucoside was added directly before 4 µL sample was 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 Scientific) operated at 4˚C and 100 % humidity.

Cryo-EM data were collected on a FEI Titan Krios microscope (Thermo Fisher Scientific) 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 a beam-tilt based acquisition scheme in SerialEM (Schorb et al., 2019), the sample was imaged with a total exposure of 63 $e^-/Å^2$ evenly spread over 9.5 s and 63 frames. The target defocus ranged between -0.7 and -2.8 µm.

Movies were pre-processed on-the-fly using Focus (Biyani et al., 2017) while automatically discarding images of poor quality. 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 (Punjani et al., 2017) for further processing in 3D (**Suppl Fig. 4C**).

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

Cross-linking and mass spectrometry

For cross-linking mass spectrometry with BS3, 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 sec at high intensity. For reduction and alkylation of the proteins, 40 mM 2-cloroacetamide (CAA, Sigma-Aldrich) and 10 mM tris(2-carboxyethyl)phosphine (TCEP; Thermo Fisher Scientific) 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 8 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 Scientific) 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 Scientific) (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 minutes, followed by a linear increase to 60% over 5 minutes, a linear increase to 95% over the next 5 minutes, and finally maintenance at 95% for another 5 minutes. 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 3 x 10^6 and 10^5, 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 crosslinked peptide pairs, a database search was performed against a FASTA containing the sequences of the proteins under investigation. DSS was set as a crosslinker. 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.

For crosslinking of the pre-mRNA cleavage complex with DSBU, PAP and RBBP6 were dialyzed against 25 mM HEPES-KOH, pH 8.0, 200 mM KCl, 10 % sucrose, 1 mM DTT to remove Tris buffer; all other proteins were present in buffers lacking amines. Reactions with or without PVA were set up. Each reaction contained 250 µL 2x cleavage buffer (see above), 6.25 pmol SV40 late RNA, 25 µg *E.coli* tRNA, 50 µl CDB200, 12.5 pmol each of CFI, CFII, CstF, bPAP, CM3, mPSF and either 62.5 pmol $RBBP6_{1-340}$ or 31.25 pmol $RBBP6_{1-780}$. The reaction volume was increased to 500µl with 25 mM HEPES-KOH, pH 8.0, 200 mM KCl, 10 % sucrose, 1 mM DTT. After pre-incubation for 30 minutes at 30°C, 1.5 µl DSBU (Thermo Fisher; 50 mM in DMSO) was added and the incubation continued for another 30 minutes at 25°C. The reaction was stopped by the addition of 25 µl 1 M Tris-HCl, pH 8.0. Proteins were precipitated by the addition of TCA to 10% and pelleted by centrifugation. Pellets were washed with 80 % ice-cold acetone, air-dried, and dissolved in 25 µl 8 M urea in 0.4 M ammonium bicarbonate. Sulfide bonds were reduced for 20 min at 50°C with 7.5 mM DTT, and alkylation was performed at 25°C for 20 min with 14 mM chloroacetamide. The volume was made up to 200 µl with H_2O, and 2.5 µg trypsin (sequencing grade modified trypsin, Promega) was added. After incubation over night at 37°C, the reaction was stopped by the addition of 20 µl 10 % TFA. Peptides were first purified by StageTips (Rappsilber et al., 2007). Per reaction 10 StageTips with 5 Empore C_{18} extraction discs (Sigma-Aldrich) were prepared and washed with 20 µl methanol, then with buffer B (60% acetonitrile, 0.1% TFA) and finally buffer A (0.1% TFA). Equal volumes of the crosslink reactions were loaded on the StageTips. The StageTips were washed with 20 µl buffer A and eluted with 5 µl buffer B. The eluate was loaded on a Superdex Peptide 3.2/300 column (GE Healthcare) running at 25 µL/min in 3% acetonitrile, 0.1% TFA. Fractions of 50µl were collected and analyzed by nano-HPLC/nano-ESI-MS/MS using an Ultimate 3000 RSLC nano-HPLC system coupled to an Orbitrap Q-Exactive Plus

mass spectrometer equipped with NanoFlex source (both Thermo Fisher Scientific). Samples were injected on a trap column (PepMap RP-C18, 300 µm × 5 mm, 5 µm, 100 Å, Thermo Fisher Scientific) with 0.1 % TFA at a flow rate of 30 µL/min. After 15 min, peptides were eluted via a 200 cm µPAC™ separation column (Pharmafluidics). For separation, 240 min gradients were applied (3-6.5 % B over 25 min at 600 nL/min followed by 6.5-33 % B over 205 min at 300 nL/min; A: 0.1 % formic acid in water, B: 0.08 % formic acid in acetonitrile). Data were acquired in data-dependent MS/MS mode (one MS survey scan, followed by 10 MS/MS scans of the 10 most abundant signals). MS survey scans (*m/z* range 375-1799) were recorded with *R* = 140000 at *m/z* 200, the target of the automated gain control (AGC) was set to $3x10^6$ and the maximum injection time was 100 ms. For MS/MS scans, triply to seven times protonated ions were selected for fragmentation by HCD with stepped normalized collision energies (NCEs: 27, 30, and 33 V). The following parameters were applied for MS/MS scans: *R* = 17.500, AGC: 2×10^5 and maximum injection time: 250 ms; dynamic exclusion was set to 60 s. The raw files were converted to mgf files by MsConvert (Chambers et al., 2012). Crosslinks were analyzed by Merox v. 2.0.1.4 (Gotze et al., 2015) and visualized with the help of xView (Graham et al., 2019).

References

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase IIC-terminal domain couples transcription and 3 ' end processing. Molecular Cell *13*, 67-76.

Alcott, C.E., Yalamanchili, H.K., Ji, P., van der Heijden, M.E., Saltzman, A., Elrod, N., Lin, A., Leng, M., Bhatt, B., Hao, S.*, et al.* (2020). Partial loss of CFIm25 causes learning deficits and aberrant neuronal alternative polyadenylation. Elife *9*.

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

Berg, J.M., Tymoczko, J.L., Gatto, G.J., and Stryer, L. (2019). Biochemistry, 9th edn (New Yor, NY: W. H. Freeman and Company).

Bieniossek, C., Richmond, T.J., and Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. Curr Protoc Protein Sci *51*, 5.20.21 - 25.20.26.

Bienroth, S., Keller, W., and Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. EMBO J *12*, 585-594.

Biyani, N., Righetto, R.D., McLeod, R., Caujolle-Bert, D., Castano-Diez, D., Goldie, K.N., and Stahlberg, H. (2017). Focus: The interface between data collection and data processing in cryo-EM. J Struct Biol *198*, 124-133.

Brown, K.M., and Gilmartin, G.M. (2003). A mechanism for the regulation of pre-mRNA 3 ' processing by human cleavage factor I-m. Molecular Cell *12*, 1467-1476.

Casanal, A., Kumar, A., Hill, C.H., Easter, A.D., Emsley, P., Degliesposti, G., Gordiyenko, Y., Santhanam, B., Wolf, J., Wiederhold, K.*, et al.* (2017). Architecture of eukaryotic mRNA 3 '-end processing machinery. Science *358*, 1056-1059.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J.*, et al.* (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol *30*, 918-920.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science *318*, 1780-1782.

Chen, C.Y., and Shyu, A.B. (2011). Mechanisms of deadenylation-dependent decay. Wiley Interdiscip Rev RNA *2*, 167-183.

Christofori, G., and Keller, W. (1989). Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. Mol Cell Biol *9*, 193-203.

Clerici, M., Faini, M., Aebersold, R., and Jinek, M. (2017). Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. Elife *6*, e33111.

Clerici, M., Faini, M., Muckenfuss, L.M., Aebersold, R., and Jinek, M. (2018). Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. Nat Struct Mol Biol *25*, 135-138.

Coseno, M., Martin, G., Berger, C., Gilmartin, G., Keller, W., and Doublie, S. (2008). Crystal structure of the 25 kDa subunit of human cleavage factor I-m. Nucleic Acids Res *36*, 3474-3483.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol *26*, 1367-1372.

Dass, B., McMahon, K.W., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., and MacDonald, C.C. (2001). The gene for a variant form of the polyadenylation protein CstF-64 is on chromosome 19 and is expressed in pachytene spermatocytes in mice. Journal of Biological Chemistry *276*, 8044-8050.

Davidson, L., Muniz, L., and 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. Gene Dev *28*, 342-356.

de Klerk, E., Venema, A., Anvar, S.Y., Goeman, J.J., Hu, O.H., Trollet, C., Dickson, G., den Dunnen, J.T., van der Maarel, S.M., Raz, V.*, et al.* (2012). Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res *40*, 9089-9101.

Di Giammartino, D.C., Li, W., Ogami, K., Yashinskie, J.J., Hoque, M., Tian, B., and 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.

Dikfidan, A., Loll, B., Zeymer, C., Magler, I., Clausen, T., and Meinhart, A. (2014). RNA specificity and regulation of catalysis in the eukaryotic polynucleotide kinase Clp1. Mol Cell *54*, 975-986.

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Gene Dev *34*, 132-145.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020). The Dynamics of Cytoplasmic mRNA Metabolism. Mol Cell *77*, 786-799 e710.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat Methods *3*, 1021-1032.

Ghazy, M.A., Gordon, J.M., Lee, S.D., Singh, B.N., Bohm, A., Hampsey, M., and Moore, C. (2012). The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. Nucleic Acids Res *40*, 1214-1225.

Gotze, M., Pettelkau, J., Fritzsche, R., Ihling, C.H., Schafer, M., and Sinz, A. (2015). Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis. J Am Soc Mass Spectr *26*, 83-97.

Graham, M.J., Combe, C., Kolbowski, L., and Rappsilber, J. (2019). xiView: a common platform for the downstream analysis of crosslinking mass spectrometry data. bioRxiv, https://doi.org/10.1101/561829.

Gruber, A.J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. Nat Rev Genet *20*, 599-614.

Haddad, R., Maurice, F., Viphakone, N., Voisinet-Hakil, F., Fribourg, S., and Minvielle-Sebastia, L. (2012). An essential role for Clp1 in assembly of polyadenylation complex CF IA and Pol II transcription termination. Nucleic Acids Res *40*, 1226-1239.

Hamilton, K., and Tong, L. (2020). Molecular mechanism for the interaction between human CPSF30 and hFip1. Gene Dev *34*, 1753-1761.

Hill, C.H., Boreikaite, V., Kumar, A., Casanal, A., Kubik, P., Degliesposti, G., Maslen, S., Mariani, A., von Loeffelholz, O., Girbig, M.*, et al.* (2019). Activation of the Endonuclease that Defines mRNA 3 ' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. Molecular Cell *73*, 1217-1231.

Hirose, Y., and Manley, J.L. (1997). Creatine phosphate, not ATP, is required for 3' end cleavage of mammalian pre-mRNA in vitro. Journal of Biological Chemistry *272*, 29636-29642.

Hirose, Y., and Manley, J.L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. Nature *395*, 93-96.

Holbein, S., Scola, S., Loll, B., Dichtl, B.S., Hubner, W., Meinhart, A., and Dichtl, B. (2011). The P-Loop Domain of Yeast Clp1 Mediates Interactions Between CF IA and CPF Factors in Pre-mRNA 3 ' End Formation. Plos One *6*, e29139.

Hosp, F., Scheltema, R.A., Eberl, H.C., Kulak, N.A., Keilhauer, E.C., Mayr, K., and Mann, M. (2015). A Double-Barrel Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) System to Quantify 96 Interactomes per Day. Mol Cell Proteomics *14*, 2030-2041.

Humphrey, T., Christofori, G., Lucijanic, V., and Keller, W. (1987). Cleavage and polyadenylation of messenger RNA precursors in vitro occurs with large and specific 3' processing complexes. EMBO J *6*, 4159-4168.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kuhn, U., Menzies, F.M., Vrielink, J.A.F.O., Bos, A.J., Drost, J., Rooijers, K.*, et al.* (2012). The Poly(A)-Binding Protein Nuclear 1 Suppresses Alternative Cleavage and Polyadenylation Sites. Cell *149*, 538-553.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A.*, et al.* (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.

Kamieniarz-Gdula, K., Gdula, M.R., Panser, K., Nojima, T., Monks, J., Wisniewski, J.R., Riepsaame, J., Brockdorff, N., Pauli, A., and Proudfoot, N.J. (2019). Selective Roles of Vertebrate PCF11 in Premature and Full-Length Transcript Termination. Molecular Cell *74*, 158-172.

Kappo, M.A., AB, E., Hassem, F., Atkinson, R.A., Faro, A., Muleya, V., Mulaudzi, T., Poole, J.O., McKenzie, J.M., Chibi, M.*, et al.* (2012). Solution Structure of RING Finger-like Domain

of Retinoblastoma-binding Protein-6 (RBBP6) Suggests It Functions as a U-box. Journal of Biological Chemistry *287*, 7146-7158.

Kerwitz, Y., Kühn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E., and 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.

Kim, S., Yamamoto, J., Chen, Y.X., Aida, M., Wada, T., Handa, H., and Yamaguchi, Y. (2010). Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. Genes Cells *15*, 1003-1013.

Klykov, O., Steigenberger, B., Pektas, S., Fasci, D., Heck, A.J.R., and Scheltema, R.A. (2018). Efficient and robust proteome-wide approaches for cross-linking mass spectrometry. Nat Protoc *13*, 2964-2990.

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A., and 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.

Kühn, U., Buschmann, J., and Wahle, E. (2017). The nuclear poly(A) binding protein of mammals, but not of fission yeast, participates in mRNA polyadenylation. Rna *23*, 473-482.

Kühn, U., Gündel, M., Knoth, A., Kerwitz, Y., Rüdel, S., and 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.

Kulak, N.A., Geyer, P.E., and Mann, M. (2017). Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. Mol Cell Proteomics *16*, 694-705.

Kumar, A., Clerici, M., Muckenfuss, L.M., Passmore, L.A., and Jinek, M. (2019). Mechanistic insights into mRNA 3 '-end processing. Curr Opin Struc Biol *59*, 143-150.

Lee, S.D., and 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.

Li, W.C., You, B., Hoque, M., Zheng, D.H., Luo, W.T., Ji, Z., Park, J.Y., Gunderson, S.I., Kalsotra, A., Manley, J.L.*, et al.* (2015). Systematic Profiling of Poly(A) plus Transcripts Modulated by Core 3 ' End Processing and Splicing Factors Reveals Regulatory Rules of Alternative Cleavage and Polyadenylation. Plos Genetics *11*, e1005166.

Li, Z., Michael, I.P., Zhou, D., Nagy, A., and Rini, J.M. (2013). Simple piggyBac transposon-based mammalian cell expression system for inducible protein production. Proc Natl Acad Sci U S A *110*, 5004-5009.

Lidschreiber, M., Easter, A.D., Battaglia, S., Rodriguez-Molina, J.B., Casanal, A., Carminati, M., Baejen, C., Grzechnik, P., Maier, K.C., Cramer, P.*, et al.* (2018). The APT complex is involved in non-coding RNA transcription and is distinct from CPF. Nucleic Acids Res *46*, 11528-11538.

Mandel, C.R., Kaneko, S., Zhang, H.L., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3 '-end-processing endonuclease. Nature *444*, 953-956.

Martin, G., Gruber, A.R., Keller, W., and 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.

Martin, G., Jeno, P., and 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 *8*, 2380-2391.

Martin, G., and Keller, W. (1996). Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and a catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. Embo J *15*, 2593-2603.

Martin, G., Keller, W., and Doublie, S. (2000). Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. EMBO J *19*, 4193-4203.

Marzluff, W.F., and Koreski, K.P. (2017). Birth and Death of Histone mRNAs. Trends Genet *33*, 745-759.

Masamha, C.P., Xia, Z., Yang, J.X., Albrecht, T.R., Li, M., Shyu, A.B., Li, W., and Wagner, E.J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature *510*, 412-+.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3 '-RNA-processing factors. Nature *430*, 223-226.

Mirdita, M., Ovchinnikov, S., and Steinegger, M. (2021). ColabFold - Making protein folding accessible to all. BioRxiv, doi.org/10.1101/2021.1108.1115.456425.

Moore, C.L., and Sharp, P.A. (1985). Accurate Cleavage and Polyadenylation of Exogenous Rna Substrate. Cell *41*, 845-855.

Muller, M.Q., Dreiocker, F., Ihling, C.H., Schafer, M., and Sinz, A. (2010). Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. Anal Chem *82*, 6958-6968.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C., and Greenblatt, J. (2003). Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. J Biol Chem *278*, 33000-33010.

Noble, C.G., Beuth, B., and Taylor, I.A. (2007). Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res *35*, 87-99.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci *30*, 70-82.

Picard, V., Ersdal-Badju, E., Lu, A., and Bock, S.C. (1994). A rapid and efficient one-tube PCR-based mutagenesis technique using Pfu DNA polymerase. Nucleic Acids Res *22*, 2587-2591.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends Genet *32*, 508-522.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Pugh, D.J.R., Ab, E., Faro, A., Lutya, P.T., Hoffmann, E., and 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*.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods *14*, 290-296.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nature Protocols *2*, 1896-1906.

Rodriguez-Molina, J.B., O'Reilly, F.J., Sheekey, E., Maslen, S., Skehel, J.M., Rappsilber, J., and Passmore, L.A. (2021). Mpe1 senses the polyadenylation signal in pre-mRNA to control cleavage and polyadenylation. BioRxiv, https://doi.org/10.1101/2021.1109.1102.458805

Rüegsegger, U., Beyer, K., and Keller, W. (1996). Purification and characterization of human cleavage factor I-m, involved in the 3' end processing of messenger RNA precursors. Journal of Biological Chemistry *271*, 6107-6113.

Ruegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor I-m is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Molecular Cell *1*, 243-253.

Ruepp, M.D., Schweingruber, C., Kleinschmidt, N., and Schumperli, D. (2011). Interactions of CstF-64, CstF-77, and symplekin: Implications on localization and function. Mol Biol Cell *22*, 91-104.

Ryan, K., Murthy, K.G.K., Kaneko, S., and Manley, J.L. (2002). Requirements of the RNA polymerase IIC-terminal domain for reconstituting pre-mRNA 3 ' cleavage. Mol Cell Biol *22*, 1684-1692.

Sasado, T., Kondoh, H., Furutani-Seiki, M., and Naruse, K. (2017). Mutation in cpsf6/CFIm68 (Cleavage and Polyadenylation Specificity Factor Subunit 6) causes short 3 ' UTRs and disturbs gene expression in developing embryos, as revealed by an analysis of primordial germ cell migration using the medaka mutant naruto. Plos One *12*, e0172467.

Schäfer, P., Tüting, C., Schönemann, L., Kühn, U., Treiber, T., Treiber, N., Ihling, C., Graber, A., Keller, W., Meister, G.*, et al.* (2018). Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. Rna *24*, 1721-1737.

Schönemann, L., Kühn, U., Martin, G., Schäfer, P., Gruber, A.R., Keller, W., Zavolan, M., and Wahle, E. (2014). Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by WDR33. Genes Dev *28*, 2381-2393.

Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y., and Mastronarde, D.N. (2019). Software tools for automated transmission electron microscopy. Nat Methods *16*, 471-477.

Sheets, M.D., Stephenson, P., and Wickens, M.P. (1987). Products of Invitro Cleavage and Polyadenylation of Simian Virus-40 Late Pre-Messenger-Rnas. Mol Cell Biol *7*, 1518-1529.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell *33*, 365-376.

Sullivan, K.D., Steiniger, M., and Marzluff, W.F. (2009). A Core Complex of CPSF73, CPSF100, and Symplekin May Form Two Different Cleavage Factors for Processing of Poly(A) and Histone mRNAs. Molecular Cell *34*, 322-332.

Sun, Y., Zhang, Y., Hamilton, K., Manley, J.L., Shi, Y., Walz, T., and Tong, L. (2018). Molecular basis for the recognition of the human AAUAAA polyadenylation signal. Proc Natl Acad Sci U S A *115*, E1419-E1428.

Sun, Y.D., Zhang, Y.X., Aik, W.S., Yang, X.C., Marzluff, W.F., Walz, T., Dominski, Z., and Tong, L. (2020). Structure of an active human histone pre-mRNA 3 '-end processing machinery. Science *367*, 700-+.

Takagaki, Y., and Manley, J.L. (1994). A Polyadenylation Factor Subunit Is the Human Homolog of the Drosophila Suppressor of Forked Protein. Nature *372*, 471-474.

Takagaki, Y., and Manley, J.L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol Cell Biol *20*, 1515-1525.

Takagaki, Y., Rhyner, L.C., and Manley, J.L. (1989). Four factors are required for 3'-end cleavage of pre-mRNAs. Genes Dev *3*, 1711-1724.

Takagaki, Y., Ryner, L.C., and Manley, J.L. (1988). Separation and Characterization of a Poly(a) Polymerase and a Cleavage Specificity Factor Required for Pre-Messenger Rna Polyadenylation. Cell *52*, 731-742.

Tian, B., and Manley, J.L. (2017). Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol *18*, 18-30.

Venkataraman, K., Brown, K.M., and Gilmartin, G.M. (2005). Analysis of a noncanonical poly(A) site reveals a trinartite mechanism for vertebrate poly(A) site recognition. Gene Dev *19*, 1315-1327.

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

Weissmann, F., and Peters, J.M. (2018). Expressing Multi-subunit Complexes Using biGBac. Methods Mol Biol *1764*, 329-343.

Weissmann, F., Petzold, G., VanderLinden, R., Huis In 't Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A., and Peters, J.M. (2016). biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. Proc Natl Acad Sci U S A *113*, E2564-2569.

Weitzer, S., Hanada, T., Penninger, J.M., and Martinez, J. (2015). CLP1 as a novel player in linking tRNA splicing to neurodegenerative disorders. Wiley Interdiscip Rev RNA *6*, 47-63.

Wickens, M., and Stephenson, P. (1984). Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3' end formation. Science *226*, 1045-1051.

Xiang, K.H., Tong, L., and Manley, J.L. (2014). Delineating the Structural Blueprint of the Pre-mRNA 3 '-End Processing Machinery. Mol Cell Biol *34*, 1894-1910.

Yang, Q., Coseno, M., Gilmartin, G.M., and Doublie, S. (2011). Crystal Structure of a Human Cleavage Factor CFI(m)25/CFI(m)68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. Structure *19*, 368-377.

Yang, Q., Gilmartin, G.M., and Doublie, S. (2010). Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3 ' processing. P Natl Acad Sci USA *107*, 10062-10067.

Yang, W., Hsu, P.L., Yang, F., Song, J.-E., and 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.

Yang, X.C., Sun, Y.D., Aik, W.S., Marzluff, W.F., Tong, L., and Dominski, Z. (2020). Studies with recombinant U7 snRNP demonstrate that CPSF73 is both an endonuclease and a 5 '-3 ' exonuclease. Rna *26*, 1345-1359.

Yao, C.G., Choi, E.A., Weng, L.J., Xie, X.H., Wan, J., Xing, Y., Moresco, J.J., Tu, P.G., Yates, J.R., and Shi, Y.S. (2013). Overlapping and distinct functions of CstF64 and CstF64 tau in mammalian mRNA 3 ' processing. Rna-a Publication of the Rna Society *19*, 1781-1790.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A *108*, 1531-1536.

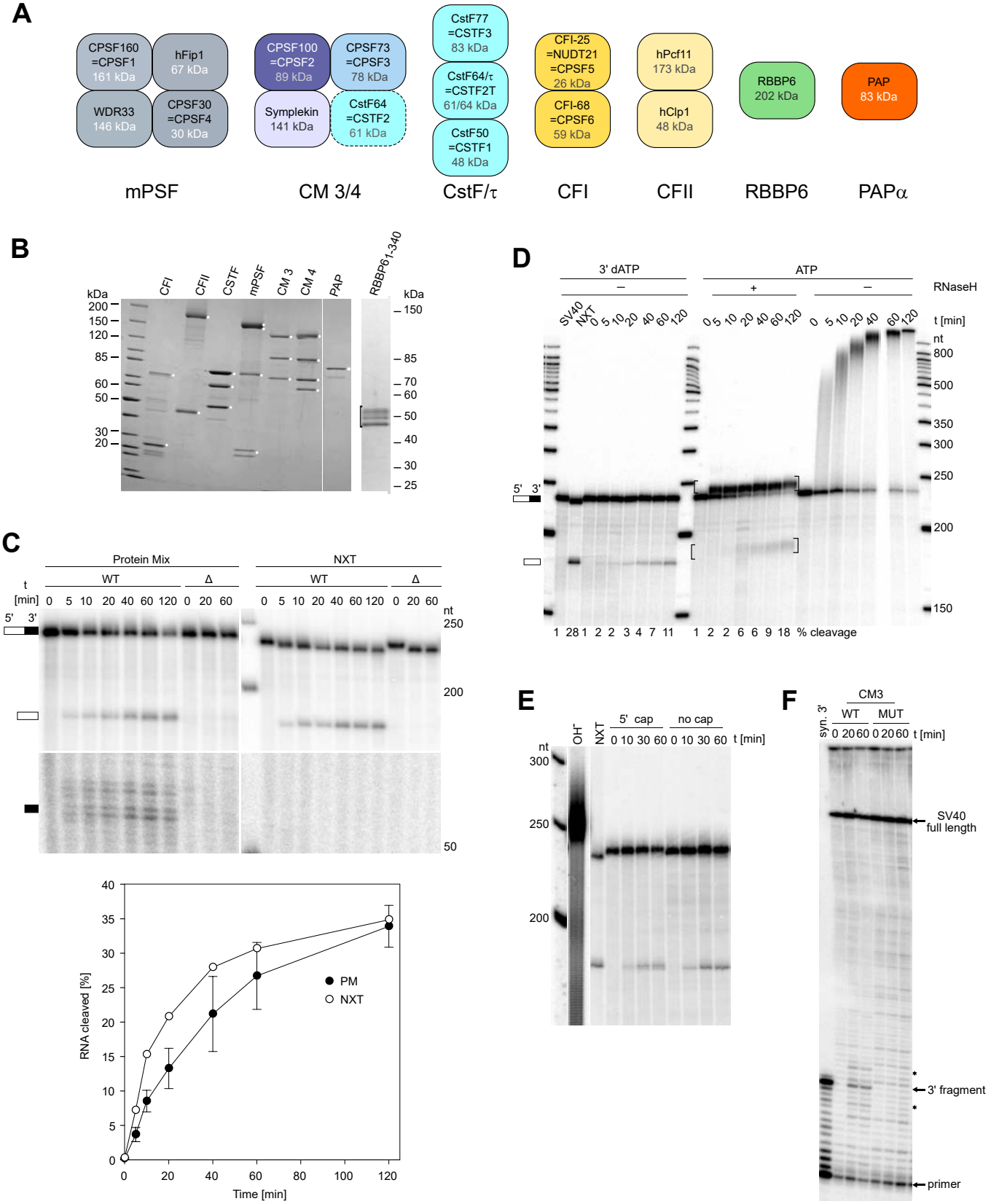
Zarkower, D., Stephenson, P., Sheets, M., and Wickens, M. (1986). The Aauaaa Sequence Is Required Both for Cleavage and for Polyadenylation of Simian-Virus 40 Pre-Messenger-Rna Invitro. Mol Cell Biol *6*, 2317-2323.

Zhang, Y., Zhang, M.M., and Zhang, Y. (2011). Crystal structure of Ssu72, an essential eukaryotic phosphatase specific for the C-terminal domain of RNA polymerase II, in complex with a transition state analogue. Biochem J *434*, 435-444.

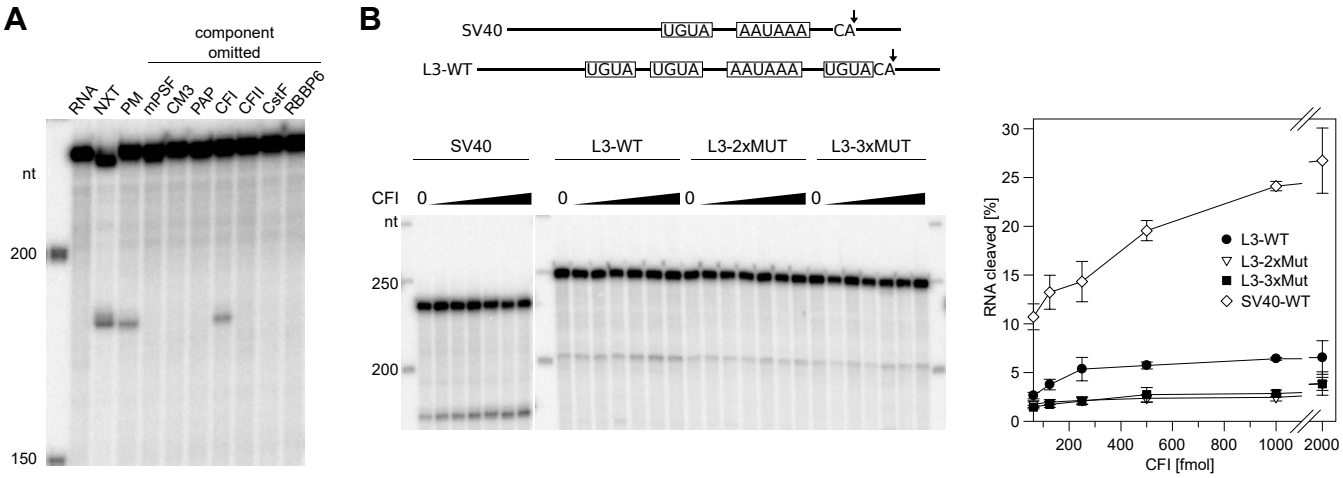
Zhang, Y.X., Sun, Y.D., Shi, Y.S., Walz, T., and Tong, L. (2020). Structural Insights into the Human Pre-mRNA 3 '-End Processing Machinery. Molecular Cell *77*, 800-809.

Zhu, Y., Wang, X.Y., Forouzmand, E., Jeong, J.S., Qiao, F., Sowd, G.A., Engelman, A.N., Xie, X.H., Hertel, K.J., and Shi, Y.S. (2018). Molecular Mechanisms for CFIm-Mediated Regulation of mRNA Alternative Polyadenylation. Molecular Cell *69*, 62-74.

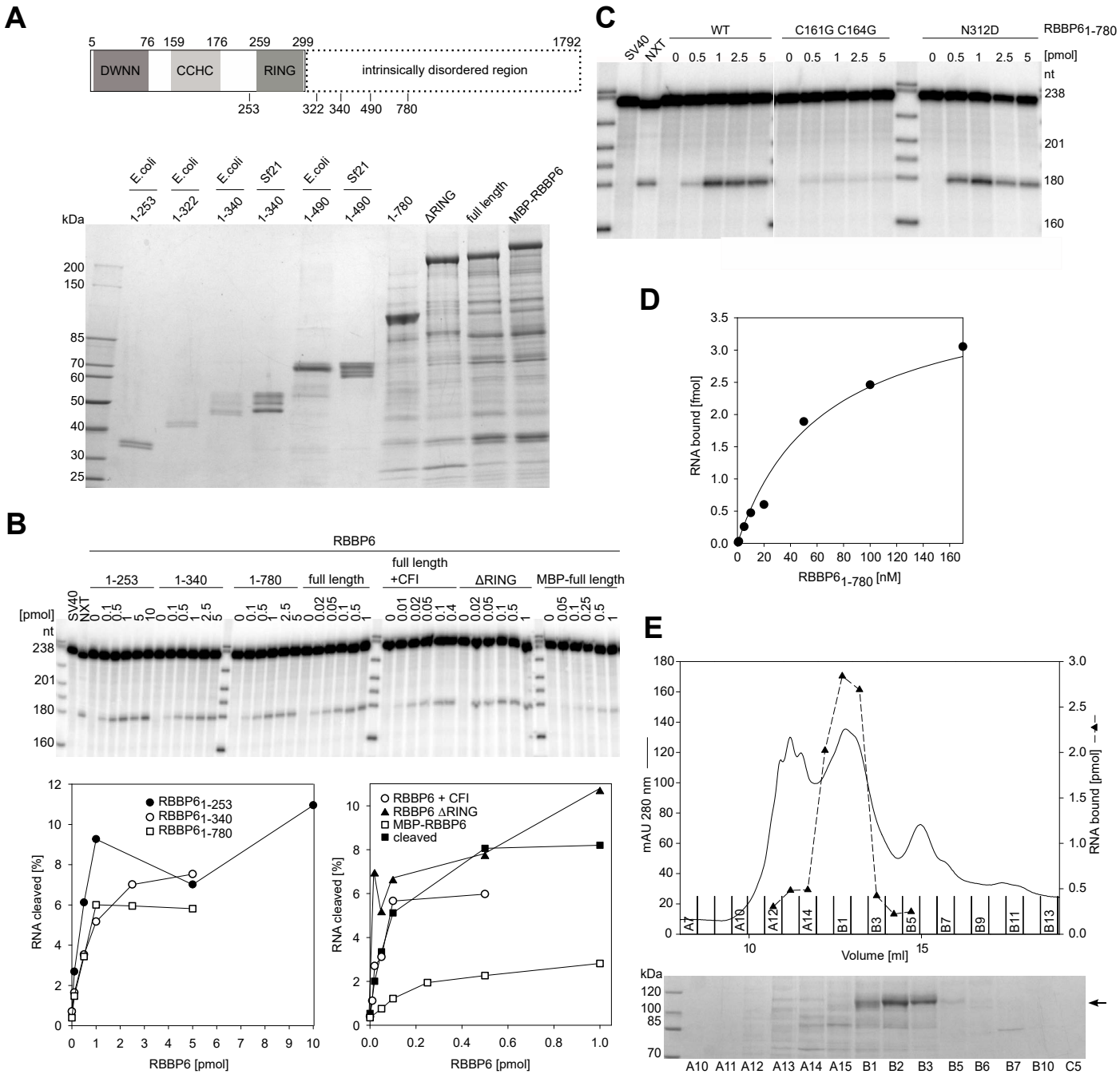
Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife *7*.



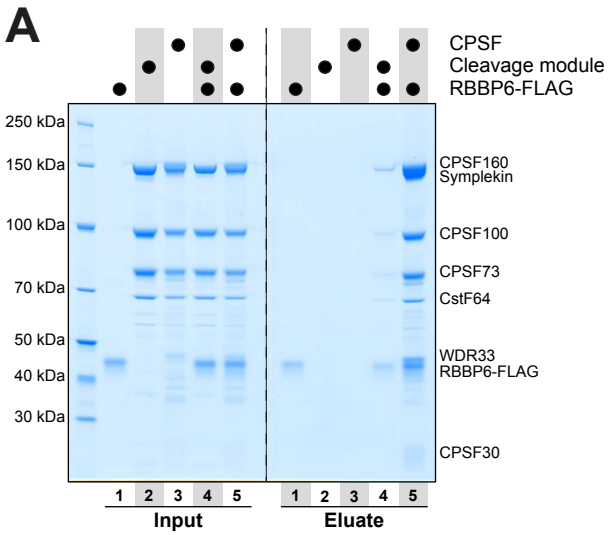
Schmidt et al. Figure 1

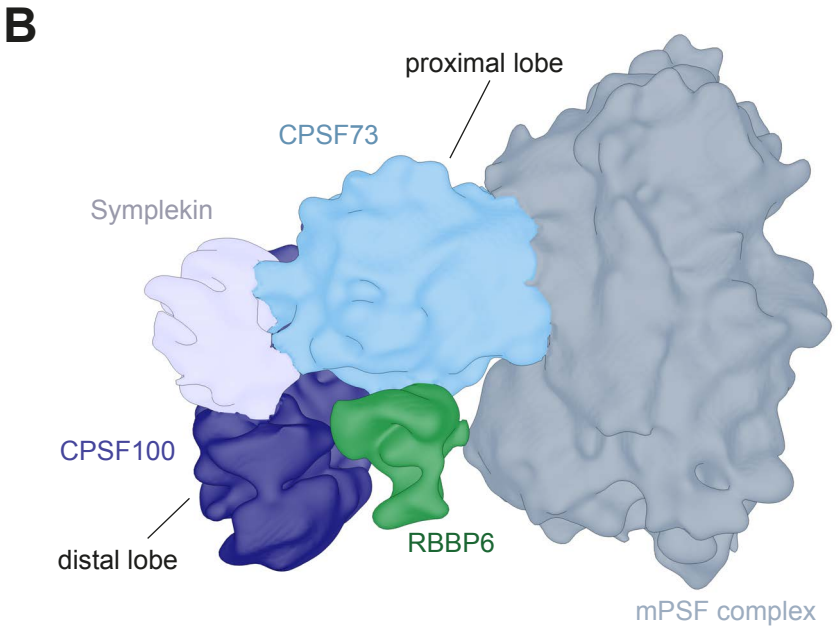


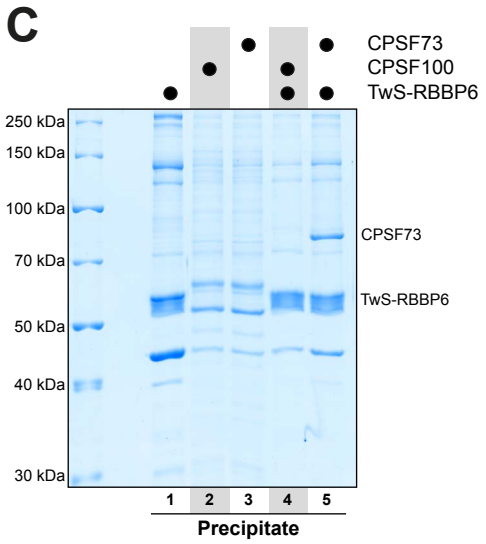
Schmidt et al. Figure 2

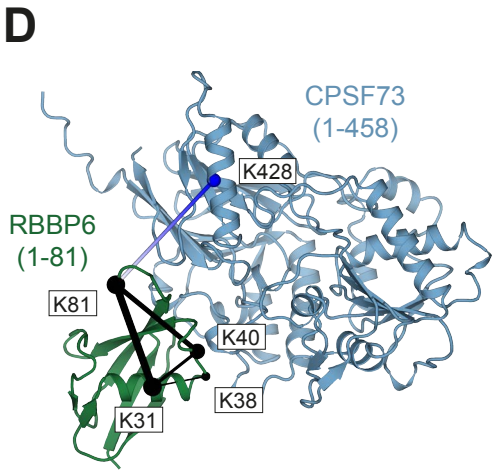


Schmidt et al. Figure 3

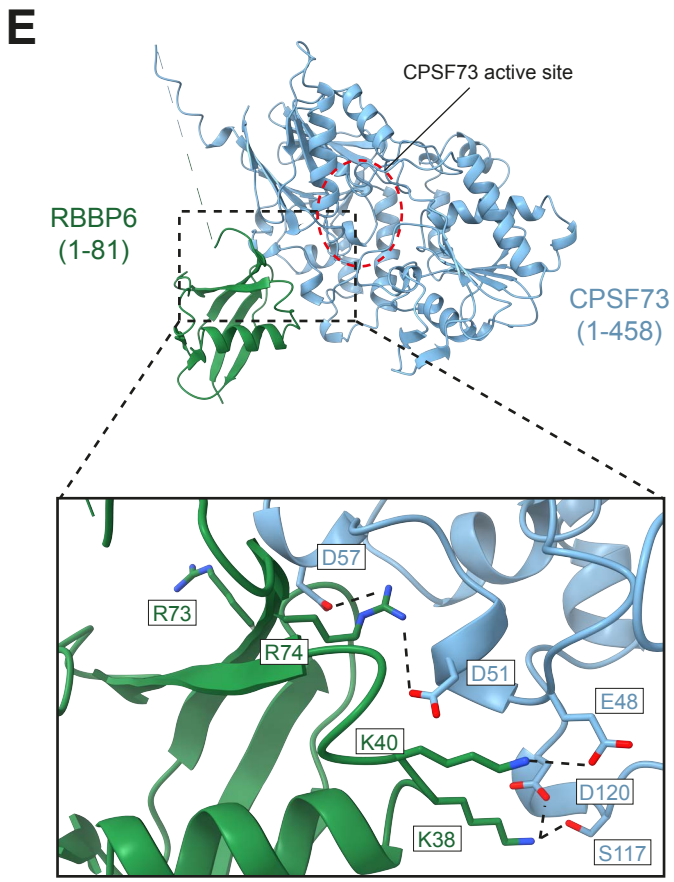


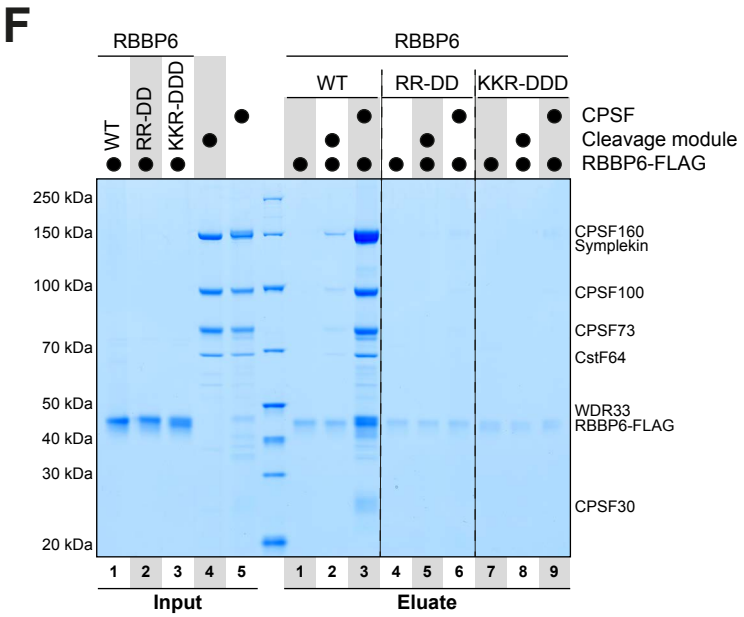


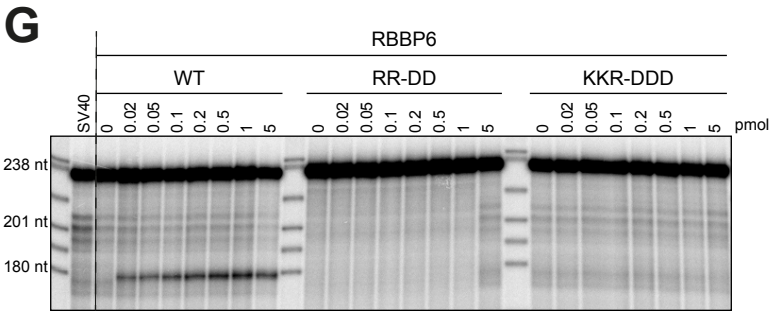




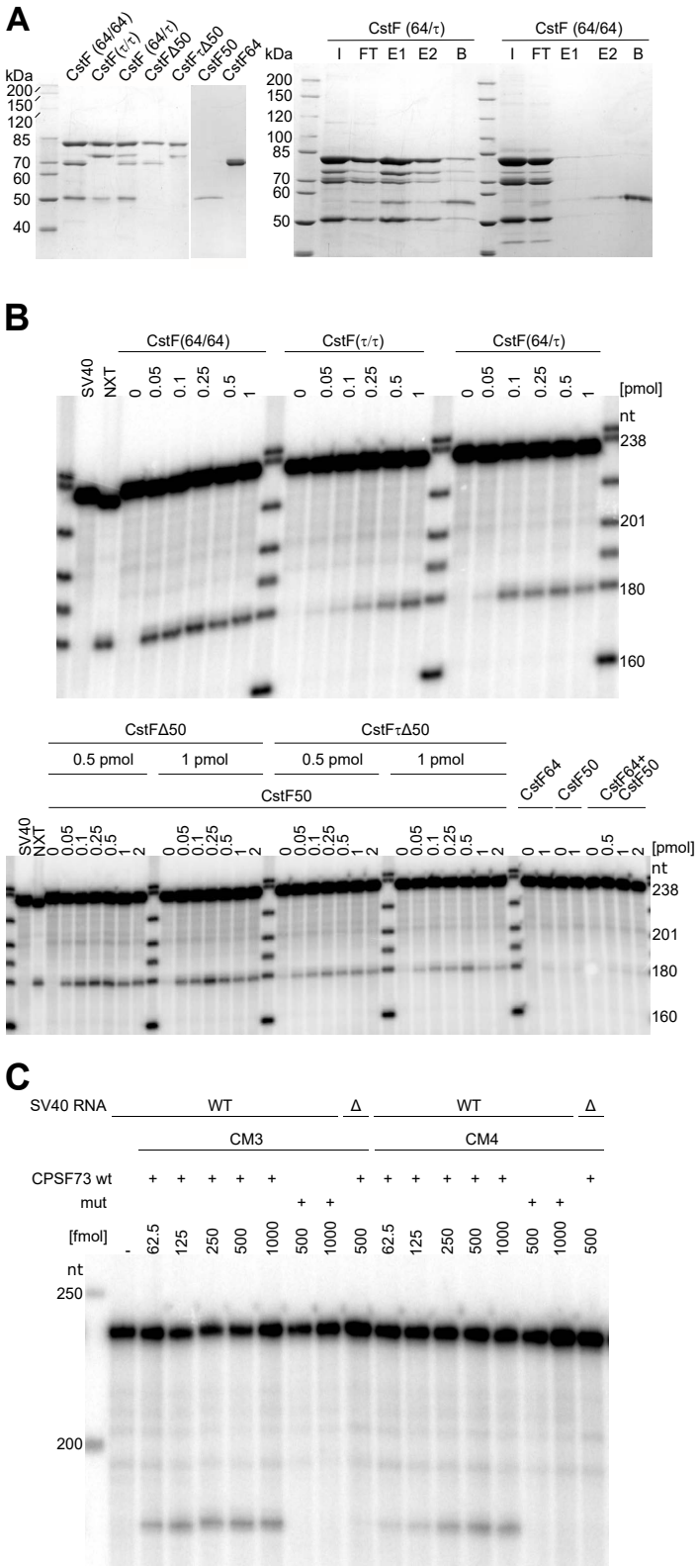
Cross-link	Cα-Cα (Å)	Type
K 85 --- K 428	> 25.4	Inter
K 81 --- K 428	25.4	Inter
K 81 --- K 31	18.8	Intra
K 81 --- K 40	18.5	Intra
K 31 --- K 40	10.0	Intra
K 31 --- K 38	10.0	Intra



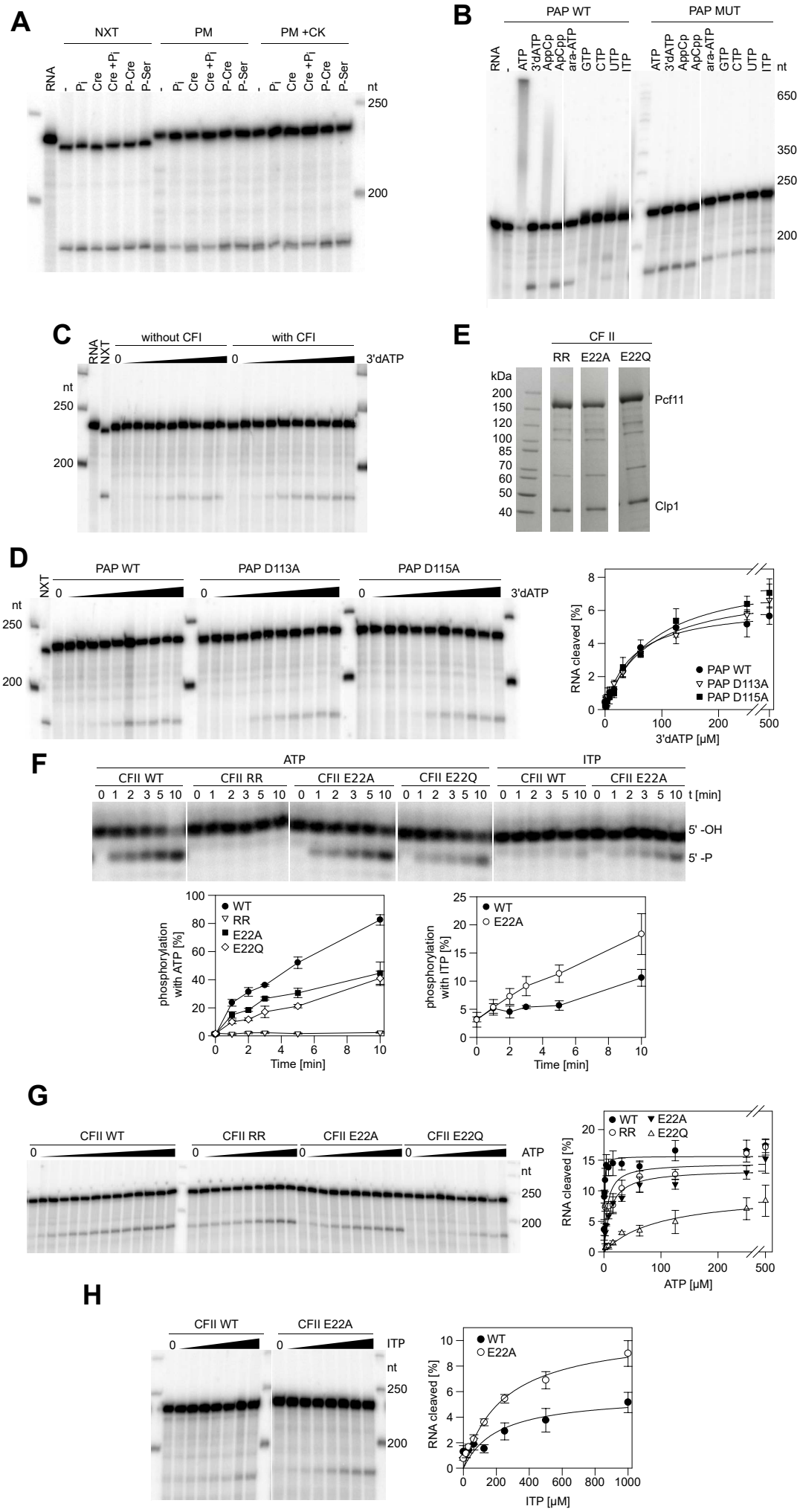




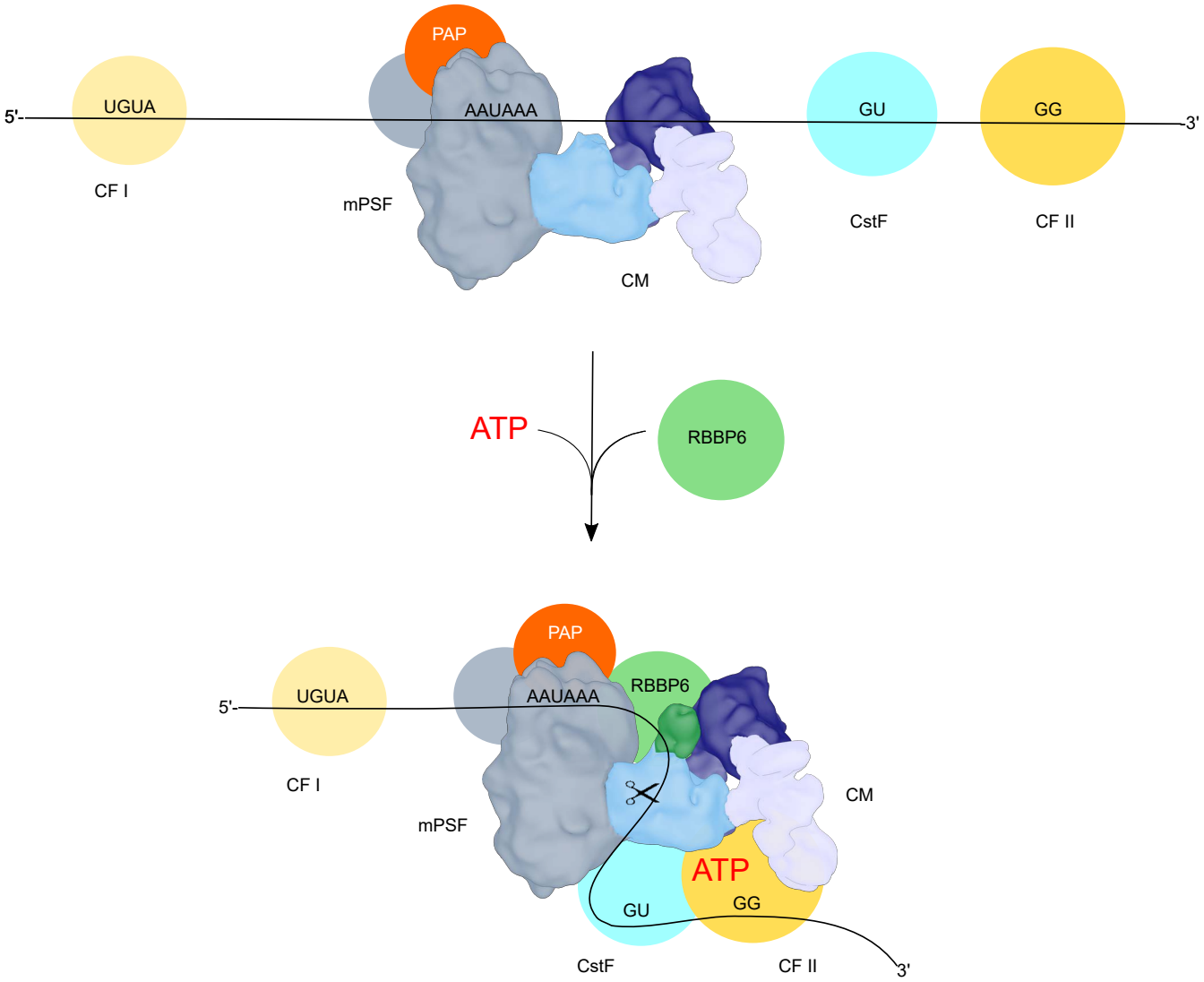
Schmidt et al. Figure 4



Schmidt et al. Figure 5



Schmidt et al. Figure 6



Schmidt et al. Figure 7