Fip1 is a multivalent interaction scaffold for processing factors in human mRNA 3' end biogenesis

Lena M. Muckenfuss, Anabel C. Migenda Herranz, Franziska M. Boneberg, Marcello Clerici, Martin Jinek*

[1]Department of Biochemistry, University of Zurich, Zurich, 8057, Switzerland

* To whom correspondence should be addressed. Tel: +41 44 635 5572, Email: jinek@bioc.uzh.ch

ABSTRACT

3' end formation of most eukaryotic mRNAs is dependent on the assembly of a ~1.5 megadalton multiprotein complex, that catalyzes the coupled reaction of pre-mRNA cleavage and polyadenylation. In mammals, the cleavage and polyadenylation specificity factor (CPSF) constitutes the core of the 3' end processing machinery onto which the remaining factors, including cleavage stimulation factor (CstF) and poly(A) polymerase (PAP), assemble. These interactions are mediated by Fip1, a CPSF subunit characterized by high degree of intrinsic disorder. Here, we report two crystal structures revealing the interactions of human Fip1 (hFip1) with CPSF30 and CstF77. We demonstrate that CPSF contains two copies of hFip1, each binding to the zinc finger (ZF) domains 4 and 5 of CPSF30. Using polyadenylation assays we show that the two hFip1 copies are functionally redundant in recruiting one copy of PAP, thereby increasing the processivity of RNA polyadenylation. We further show that the interaction between hFip1 and CstF77 is mediated via a short motif in the N-terminal "acidic" region of hFip1. In turn, CstF77 competitively inhibits CPSF-dependent PAP recruitment and 3' polyadenylation. Taken together, these results provide a structural basis for the multivalent scaffolding and regulatory functions of hFip1 in 3' end processing.

INTRODUCTION

3' end polyadenylation is a fundamental process in eukaryotic messenger RNA (mRNA) biogenesis, essential for the maturation of non-histone precursor mRNAs (pre-mRNAs) prior to their export into the cytoplasm. Poly(A) tails possess key functions in mRNA metabolism, governing mRNA export, translational efficiency, and stability (Nicholson and Pasquinelli, 2019; Passmore and Coller, 2022). Furthermore, alternative mRNA polyadenylation constitutes a key mechanism of gene expression control through dynamic regulation of polyadenylation site selection in pre-mRNA transcripts (Di Giammartino et al., 2011; Tian and Manley, 2016). Accordingly, defects in polyadenylation are linked to human diseases such as cancer, β-thalessemia, diabetes, or systemic lupus (Hollerer et al., 2014; Gruber and Zavolan, 2019; Dharmalingam et al., 2022).

mRNA 3' end biogenesis occurs by a two-step mechanism comprising endonucleolytic cleavage of the pre-mRNA transcript by the cleavage and polyadenylation specificity factor (CPSF) complex and subsequent polyadenylation of the free 3' end by the poly(A) polymerase (PAP). In human cells, the process is dependent on the controlled assembly of several protein factors on the pre-mRNA, including CPSF, RBBP6, cleavage stimulation factor (CstF), as well as mammalian cleavage factors I and II (CF Im and CFIIm, respectively), and PAP (Zhao et al., 1999; Xiang et al., 2014; Kumar et al., 2019; Boreikaite et al., 2022; Schmidt et al., 2022). Most of these protein factors are highly conserved between mammals and yeast, underlining the fundamental nature of this process (Xiang et al., 2014). The cleavage site, typically within a CA dinucleotide, is defined by the polyadenylation signal (PAS), a conserved hexanucleotide motif (predominantly AAUAAA) located approximately 10-30 nucleotides upstream (Proudfoot and Brownlee, 1976; Proudfoot, 2011).

The PAS is specifically recognized by CPSF (Chan et al., 2014; Schönemann et al., 2014; Clerici et al., 2018; Sun et al., 2018), which consists of two functional modules: the mammalian polyadenylation specificity factor (mPSF) comprising subunits CPSF160, WDR33, CPSF30 and hFip1 (for human factor interacting with poly(A) polymerase 1) (Bienroth et al., 1991; Murthy and Manley, 1992; Kaufmann et al., 2004; Shi et al., 2009), and the mammalian cleavage factor (mCF) containing the endonuclease CPSF73 (Mandel et al., 2006), CPSF100 as well as Symplekin (Sullivan et al., 2009). RBBP6 associates with mCF and is essential for pre-mRNA cleavage (Di Giammartino et al., 2014; Boreikaite et al., 2022; Schmidt et al., 2022). Within mPSF, the CPSF160-WDR33 subcomplex forms a rigid scaffold (Clerici et al., 2017) that interacts with CPSF30 (Clerici et al., 2018; Sun et al., 2018;

Zhang et al., 2019) and the CPSF100 subunit of mCF (Zhang et al., 2019). CPSF30 contains five C3H1-type zinc finger (ZF) domains and a C-terminal zinc knuckle domain which is absent in yeast homologue Yth1 (Barabino et al., 1997) and not required for mPSF complex assembly (Clerici et al., 2017). The ZF1 domain is necessary and sufficient for binding to the CPSF160-WDR33 heterodimer, while ZF2 and ZF3 together with WDR33 mediate recognition of the AAUAAA PAS hexamer motif (Clerici et al., 2018; Sun et al., 2018). ZF4 and ZF5 domains interact with hFip1 (Barabino et al., 2000; Hamilton and Tong, 2020). In previously determined cryo-EM structures of the yeast CPF and human mPSF complexes, ZF4 and ZF5 remained unresolved, indicating conformational flexibility with respect to the rigid mPSF core. Recently, a crystal structure of human CPSF30 ZF4-5 domains in complex with hFip1 has been determined (Hamilton and Tong, 2020) and complementary NMR studies of the yeast Fip1 homolog (Kumar et al., 2021) have shed light on the molecular details of the CPSF30-Fip1 interaction and revealed considerable structural dynamics of Fip1 in the context of the 3' processing machinery.

Mammalian CstF is a dimer of trimers comprising CstF77, CstF64 and CstF50 subunits (Takagaki et al., 1990; Yang et al., 2018). It is recruited to the pre-mRNA by U- and G/U-rich sequences located downstream of the cleavage site (Takagaki and Manley, 1997) that are recognized by CstF64 (Takagaki et al., 1992; MacDonald et al., 1994). Through stabilization of CPSF on the pre-mRNA, CstF plays an important role in PAS recognition and is essential for pre-mRNA cleavage (Takagaki et al., 1990; Boreikaite et al., 2022; Schmidt et al., 2022). Dimerization of CstF is mediated by the CstF77 HAT (half-a-tetratricopeptide repeat) domain homodimer (Bai et al., 2007), and further stabilized by CstF50 (Yang et al., 2018). The CstF77 homodimer has an arch-like shape and interacts asymmetrically with CPSF, contacting the CPSF160-WDR33 mPSF scaffold via only one side of the arch (Zhang et al., 2019).

Fip1 interacts with PAP and tethers it to CPSF bound near the nascent 3' end of the cleaved pre-mRNA, which is required for its processive polyadenylation (Preker et al., 1995; Helmling et al., 2001; Kaufmann et al., 2004; Meinke et al., 2008; Ezeokonkwo et al., 2011). Besides CPSF30 and PAP, biochemical and cellular studies have implicated Fip1 in interactions with other proteins including CPSF160, CstF77 (Preker et al., 1995; Kaufmann et al., 2004), WDR33 (Ohnacker et al., 2000; Clerici et al., 2017), Symplekin (Ghazy et al., 2009), and CF Im (Venkataraman et al., 2005). However, the molecular details of these interactions have not yet been revealed.

Here, we report structural and biochemical analysis of the interactions of hFip1 with CPSF30, PAP and CstF77 within the human 3' polyadenylation machinery. While confirming previous structural data (Hamilton and Tong, 2020), we notably show that mPSF contains two hFip1 copies, yet recruits only one PAP molecule at a time. The presence of two PAP binding sites in mPSF contributes to the processivity of 3' polyadenylation. Furthermore, we show that hFip1 interacts with CstF77 through a conserved helix in its N-terminal "acidic" region and reveal that CstF77 competes with PAP for hFip1-binding, which attenuates polyadenylation efficiency. These results deepen our understanding of hFip1 as a key interaction partner for 3' end processing factors, facilitating or regulating their spatiotemporal assembly on the pre-mRNA, and establish a framework for further mechanistic studies of hFip1 interactions and CstF-mediated regulation of mRNA 3' end biogenesis.

RESULTS

Structural basis for the human hFip1-CPSF30 interaction

The ZF4 and ZF5 domains of CPSF30 are necessary and sufficient for the interaction with the conserved central domain of hFip1 (hereafter referred to as hFip1^{CD}) (Clerici et al., 2017; Hamilton and Tong, 2020). Yet these domains could not be resolved in previously determined cryo-EM reconstructions of the human mPSF (Clerici et al., 2018; Sun et al., 2018), indicating that they are likely flexibly tethered to the mPSF core. To gain insights into the CPSF30-hFip1 interaction, we determined a crystal structure of a CPSF30 fragment spanning ZF4 and ZF5 domains (CPSF30$^{ZF4\text{-}ZF5}$, residues 118-178) in complex with hFip1^{CD} (residues 138-180, hFip1 isoform 4) at a resolution of 2.2 Å (**Figure 1B**). The structure reveals that hFip1^{CD} binds CPSF30 in a 2:1 stoichiometry, with one hFip1^{CD} molecule (hFip1^{CD}a) binding predominantly to ZF4 and the other (hFip1^{CD}b) to ZF5. Structural superpositions reveal that hFip1^{CD}a and hFip1^{CD}b bind to the same surfaces of ZF4 and ZF5 domains with a root-mean-square deviation of 0.87 Å over 54 aligned residues (**Figure 1C**). Moreover, superpositions with CPSF30 ZF2 and ZF3 domains reveal that the interaction surfaces on ZF4 and ZF5 are located on the opposite side of the ZF fold relative to the PAS RNA binding surfaces of ZF2 and ZF3 (**Figure 1C**). ZF2 and ZF3 interactions with the RNA are mainly mediated by π-π stacking of aromatic side chains with nucleobases and supplemented by protein mainchain hydrogen bond interactions (Clerici et al., 2018; Sun et al., 2018). Although the aromatic residues are conserved in ZF4 and ZF5 (**Figure 1 – figure supplement 1A**), RNA binding is likely precluded by the presence of proline residues at key mainchain hydrogen binding positions. The hFip1^{CD}a-CPSF30 interaction surface (803 Å^2) is almost twice as large as the hFip1^{CD}b-CPSF30 interface (478 Å^2) because hFip1^{CD}a binds at the ZF4-ZF5 junction and has additional contacts with ZF5. hFip1^{CD}a and hFip1^{CD}b also contact each other directly (215 Å^2). ZF4 interaction with hFip1^{CD}a is mediated by a hydrophobic interface centered on Phe127^{CPSF30} and Phe131^{CPSF30} and supported by additional salt-bridge contact involving Arg144^{CPSF30} and Asp155^{hFip1} (**Figure 1D**). In turn, the interaction of ZF5 with hFip1^{CD}b is mainly mediated by Tyr151^{CPSF30} and Phe155^{CPSF30} (**Figure 1E**). As Fip1 is conformationally dynamic in isolation (Meinke et al., 2008; Ezeokonkwo et al., 2011; Kumar et al., 2021), CPSF30 binding results in structural ordering of the CD region. Interactions with both ZF4 and ZF5 are mediated by a hydrophobic patch in hFip1^{CD} comprising the aromatic side chains of Trp150^{hFip1}, Phe161^{hFip1} and Trp170^{hFip1} (**Figure 1D,E**).

To validate our structural observations, we initially mutated ZF4 and ZF5 interaction surface residues in CPSF30$^{ZF4\text{-}ZF5}$ and tested the interactions of the mutant proteins with hFip1^{CD} in a pull-down assay (**Figure 1 – figure supplement 2A**). Individual substitutions of Tyr127^{CPSF30}, Tyr151^{CPSF30} or Phe155^{CPSF30} with glutamate resulted in substantial reduction of hFip1^{CD} binding, while simultaneous mutation of both ZF4 and ZF5 residues resulted in loss of hFip1 binding, in agreement with our structural observations. In hFip1^{CD}, glutamate substitution of aromatic residues in the hydrophobic interaction patch either substantially reduced (Trp150^{hFip1}, Trp170^{hFip1}) or completely disrupted (Phe161^{hFip1}) the hFip1^{CD}-CPSF30$^{ZF4\text{-}ZF5}$ interaction (**Figure 1 – figure supplement 2B**). We subsequently performed size exclusion chromatography coupled to multi-angle static light scattering (SEC-MALS) to analyze the stoichiometry of hFip1^{CD}-CPSF30$^{ZF4\text{-}ZF5}$ complexes. hFip1^{CD} and wild-type CPSF30$^{ZF4\text{-}ZF5}$ formed a 2:1 complex. In contrast, CPSF30$^{ZF4\text{-}ZF5}$ proteins containing Y127E^{CPSF30} or F155E^{CPSF30} mutations formed a 1:1 complex with hFip1^{CD}, while simultaneous mutation of both residues resulted in complete loss of binding (**Figure 1F**). Together, these results confirm that human CPSF30 has two independently functional hFip1 binding sites, one on ZF4 and the other on ZF5, each recruiting one copy of hFip1.

Functional redundancy of hFip1-CPSF30 interactions in human CPSF

To probe the functional significance of the dual CPSF30-hFip1 interaction interfaces in the context of human CPSF, we co-expressed wild-type or mutant CPSF30 together with FLAG epitope-tagged CPSF160, WDR33 and hFip1 in baculovirus-infected insect cells, and performed tandem affinity purifications during which purified recombinant PAP was added *in trans* after the first purification step. hFip1 co-purified with mPSF containing wild-type CPSF30 and co-precipitated PAP (**Figure 1G**). Expression of CPSF30 ZF4 or ZF5 mutants (Y127E or Y151E, respectively) resulted in reduced recovery of both hFip1 and PAP (**Figure 1G**), consistent with the reduced stoichiometry of the CPSF30-hFip1 interaction observed *in vitro* (**Figure 1F**). In turn, expression of a CPSF30 construct containing mutations in both the ZF4 and ZF5 binding sites resulted in the loss of hFip1 from mPSF, which was thus unable to interact with PAP (**Figure 1G**). Together, these results indicate that both hFip1 binding sites in CPSF30 contribute to the integrity of mPSF *in vivo* and both are capable of recruiting hFip1 and consequently PAP. Notably, the expression levels of mPSF mutant complexes incapable of binding hFip1 (Y127E/Y151^{CPSF30}) were substantially reduced, consistent with the role of hFip1 in stabilizing the CPSF30 zinc finger fold (Kumar et al., 2021).

We next assessed the requirement of the hFip1-CPSF30 interactions for RNA 3' polyadenylation using an *in vitro* polyadenylation assay. Incubation of a model RNA substrate with purified wild-type mPSF (**Figure 2 – figure supplement 1A**) and PAP resulted in processive polyadenylation, which was dependent on the presence of ATP in the solution and an AAUAAA hexameric PAS in the RNA (**Figure 2A**). The efficiency and processivity of 3' polyadenylation were reduced upon incubation of the substrate with mPSF complexes containing CPSF30 ZF4 or ZF5 mutants capable of binding only one copy of hFip1 (**Figure 2A**). No RNA polyadenylation was observed upon incubation with mPSF containing the CPSF30 ZF4/ZF5 double mutant (**Figure 2A**), consistent with the loss of hFip1 (**Figure 1G**). The loss of polyadenylation could not be rescued by the addition of recombinant hFip1 *in trans*. Collectively, these observations indicate that both hFip1 binding sites in CPSF30 contribute to the processivity of RNA 3' polyadenylation, suggesting that the presence of two hFip1 copies, and thus two PAP recruitment sites, in mPSF is required for high efficiency of 3' polyadenylation. However, neither hFip1 binding site is strictly necessary for RNA 3' polyadenylation, suggesting their functional redundancy.

PAP recruitment occurs via hFip1 N-terminal region

In *S. cerevisiae*, a poorly conserved peptide motif in the N-terminal region of Fip1 directly interacts with the poly(A) polymerase Pap1 (Meinke et al., 2008). Similarly, the N-terminal region of human hFip1, upstream of the CD, is required for PAP interaction (Kaufmann et al., 2004) but the precise PAP interaction site in human hFip1 has not been identified. To this end, we tested the interaction of green fluorescent protein (GFP)-tagged PAP with purified mPSF complexes containing truncated hFip1 fragments in an *in vitro* pull-down experiment. PAP was detectably, albeit weakly, co-precipitated by mPFS containing a hFip1 fragment spanning both the N-terminal and CD regions (residues 1-195) as well as by mPFS containing an N-terminally truncated hFip1 (residues 36-195) (**Figure 2B**). However, further truncation of hFip1 resulted in the loss of PAP binding, indicating that a region spanning residues 36-80 in human hFip1 is required for PAP interaction (**Figure 2B**). An additional pull-down experiment using recombinant PAP and glutathione-S-transferase (GST)-fused hFip1 fragments revealed that although the hFip1 region comprising residues 36-80 was required for PAP interaction, it was not sufficient (**Figure 2 – figure supplement 1B**). This suggests that additional parts of hFip1 contribute to PAP binding.

We subsequently tested the activity of mPSF complexes containing N- or C-terminally truncated hFip1 in the polyadenylation assay. In agreement with the interaction data, mPSF complexes containing hFip1 fragments spanning residues 1-195 or 36-190 were able to support efficient RNA 3' polyadenylation (**Figure 2C**), whereas mPSF complexes containing hFip1 fragments comprising residues 80-195 or 130-195 were not. Together, these results indicate that hFip1 residues 36-80 are required for the recruitment of PAP to effect mPSF-dependent 3' polyadenylation. Interestingly, we also observed that polyadenylation levels were reduced with mPSF containing full-length hFip1 (residues 1-378, isoform 4), as compared to mPSF containing C-terminally truncated hFip1 (residues 1-195), suggesting that the C-terminal region of hFip1, which is proline-rich and predicted to be intrinsically disordered, has an inhibitory effect on mPSF-dependent 3' polyadenylation.

CPSF recruits only one copy of poly(A) polymerase

Prior studies have indicated that a complex comprising CPSF30 ZF4 and ZF5 domains and two hFip1 molecules is capable of simultaneous interaction with two PAP molecules *in vitro* (Hamilton and Tong, 2020). To determine whether this also occurs in the context of mPSF, we analyzed the mPSF-PAP interaction by SEC-MALS. Despite only weakly interacting in pull-down analysis, at high PAP concentrations (40 μM), mPSF and PAP formed a stable complex that could be purified by SEC. Analysis of this complex using SEC-MALS revealed an apparent molecular mass of 347 kDa, consistent with the molecular mass expected for a complex containing two hFip1 molecules and one PAP (337 kDa) (**Figure 2D**). Addition of excess PAP to the pre-purified mPSF-PAP sample did not lead to stable formation of a 1:2 complex. These results indicate that mPSF is capable of stable association with only one PAP molecule at a time, despite the presence of two copies of hFip1 in the complex.

The N-terminal region of hFip1 interacts with CstF77

In analogy with the yeast polyadenylation machinery, human Fip1 has previously been shown to interact with CstF77 (Preker et al., 1995; Kaufmann et al., 2004). To validate these observations and identify the interaction determinants in hFip1, we performed a pull-down experiment with GST-tagged hFip1 fragments and maltose binding protein (MBP)-tagged fragment of CstF77 comprising the HAT domain (residues 21-549). The very N-terminal region of hFip1 spanning residues 1-35 was necessary and sufficient for the interaction with the CstF77 HAT domain (**Figure 3A**). Notably, this region is dispensable for the interaction of hFip1 with PAP and for RNA 3' polyadenylation (**Figure 2B,C**).

To shed light on the hFip1-CstF77 interaction, we subsequently reconstituted a complex comprising the hFip1$^{1-35}$ fragment with a truncated construct of the CstF77 HAT domain (residues 241-549) and determined its X-ray crystallographic structure at a resolution of 2.7 Å. The structure reveals that hFip1 binds to a conserved positively-charged patch located on the convex surface of the CstF77 HAT domain arch (**Figure 3B, Figure 3 – figure supplement 1A,B**). Within the hFip1$^{1-35}$ fragment, only the evolutionarily conserved residues 20-27 were ordered, adopting an alpha-helical conformation (**Fig 3C,D**). Interaction of hFip1$^{1-35}$ with CstF77 involves salt bridge contacts of Glu22^{hFip1} and Glu23^{hFip1} with Arg402^{CstF77}, and hydrophobic contacts involving Leu26^{hFip1} and Tyr27^{hFip1} with Phe398^{CstF77}, Val428^{CstF77}, Ile432^{CstF77} and Leu435^{CstF77}. Additionally, the Tyr27^{hFip1} side chain interacts with Arg395^{CstF77} via a π-π stacking. Corroborating these structural observations, simultaneous alanine substitutions of Glu22^{hFip1} and Glu23^{hFip1}, or Trp25^{hFip1}, Leu26^{hFip1} and Tyr27^{hFip1}, respectively, disrupted the hFip1$^{1-35}$-CstF77$^{21-549}$ interaction in a pull-down experiment, whereas alanine substitution of Trp25^{hFip1} alone did not have an effect (**Figure 3E**). In turn, mutation of the positively charged interaction interface in CstF77 (Arg395, Arg402, and Lys431 mutated to alanines) abolished the interaction with hFip1$^{1-35}$ (**Figure 3E**).

A previously determined cryo-EM reconstruction of the human mPSF-CstF77 complex revealed that the interaction of the CstF77 HAT domain dimer with mPSF is primarily mediated by extensive contacts with WDR33 and CPSF160 (Zhang et al., 2019). Upon close inspection, the cryo-EM map from this dataset (EMDB entry EMD-20861) exhibits residual densities on both CstF77 protomers that could be attributed to the binding of two hFip1 molecules via their N-terminal regions (**Figure 3H**). This observation indicates that CstF77 is capable of binding two hFip1 copies when bound to mPSF. We subsequently tested the contribution of hFip1$^{1-35}$ to the mPSF-CstF77 interaction in a pull-down experiment using MBP-tagged CstF77 and mPSF complexes containing truncated hFip1 fragments. Although all mPSF complexes were capable of binding CstF77, reduced levels of CstF77 co-precipitation were observed with mPSF containing N-terminally truncated hFip1 that lacked the CstF77 interacting region (**Figure 3 – figure supplement 2A**). Taken together, these results suggest that direct interactions between hFip1 and CstF77 contribute to the assembly of the CPSF-CstF complex during mRNA 3' end biogenesis.

CstF77 inhibits polyadenylation by competition for hFip1

As CstF77 and PAP bind to non-overlapping, yet adjacent, sites in hFip1, CstF77 binding could nevertheless preclude PAP recruitment due to steric hindrance. To probe this, we carried out a pull-down experiment with GST-tagged hFip1 and mixtures of MBP-tagged CstF77 and GFP-tagged PAP at varying molar ratios. In the presence of excess CstF77, PAP binding was considerably reduced, indicating that CstF77 competes with PAP for binding to hFip1 (**Figure 4A**). Consistent with this, CPSF-dependent RNA 3' polyadenylation was substantially reduced in the presence CstF77, suggesting that CstF77 inhibits 3' polyadenylation via interaction with hFip1 (**Figure 4B**). Consistently, the inhibitory effect of CstF77 was reduced either when mPSF lacked the N-terminal CstF77 interaction site in hFip1 (**Figure 4B**) or when CstF77 ($CstF^{mut}$) was incapable of interaction with the N-terminal region of hFip1 (**Figure 4 – figure supplement 1A)**. In both cases, addition of excess CstF77 led to a reduction of RNA 3' polyadenylation rate, although not to the same extent (**Figure 4B, Figure 4 – figure supplement 1A**). Together, these results suggest that CstF77 inhibits RNA 3' polyadenylation in both hFip1-dependent and independent manners.

DISCUSSION

Despite extensive efforts to obtain structural insights into the molecular organization and regulation of the eukaryotic mRNA 3' end processing machinery, high-resolution structural information has so far only been obtained for stable sub-assemblies composed of structurally rigid subunits (Casañal et al., 2017; Clerici et al., 2017, 2018; Sun et al., 2018; Hill et al., 2019; Zhang et al., 2019; Hamilton and Tong, 2020; Kumar et al., 2021). Although hFip1 is an integral component of the CSPF complex, specifically its mPSF module, it has not been structurally visualized in this context owing to its intrinsically disordered nature (Meinke et al., 2008).

In our study, we reveal the molecular basis for the interactions of human hFip1 with both CPSF30, PAP and CstF77. While confirming the 2:1 binding stoichiometry of the hFip1:CPSF30 interaction in isolation (Hamilton and Tong, 2020; Kumar et al., 2021), we expand this finding to the CPSF complex, confirming that its mPSF module assembles with two hFip1 copies in cells and demonstrating that both the ZF4 and ZF5 domains in CPSF30 are capable of binding hFip1 independently. Using polyadenylation assays we show that the two hFip1 copies are functionally redundant in recruiting PAP to the mPSF, which increases the processivity of RNA 3' polyadenylation. As recruitment of PAP to the 3' end of the cleaved pre-mRNA is prerequisite for its processivity (Ezeokonkwo et al., 2011), while PAP only weakly associates with the mPSF, the presence of two hFip1 copies thus likely increases the 3' polyadenylation efficiency polyadenylation by increasing the local PAP concentration. While recent studies of human CPSF30 and its yeast homologue Yth1 reported higher binding affinity for the Fip1:ZF4 interaction as compared to Fip1:ZF5 (Hamilton and Tong, 2020; Kumar et al., 2021), we show that polyadenylation efficiency is reduced equally independent of which hFip1 interaction site (ZF4 or ZF5) is impaired. This indicates that PAP recruitment by mPSF is the limiting factor in 3' polyadenylation.

Although the yeast Fip1-Pap1 interaction has been extensively characterized biochemically and structurally (Meinke et al., 2008), the yeast Pap1 interaction motif is poorly conserved in human Fip1 (**Figure 4 – figure supplement 2A**), only partially mapping to residues 80-86 (hFip1 isoform 4). We show that an additional N-terminal segment in human Fip1 spanning residues 36-80 is required but not sufficient for PAP binding, underscoring the differences between human and yeast Fip1-PAP interactions and explaining the low degree of conservation in the respective interacting regions (Helmling et al., 2001; Meinke et al., 2008). Notably, our biophysical analysis of the human mPSF-PAP

interaction reveals that despite the presence of two hFip1 copies, only one copy of PAP is stably recruited by mPSF, in contrast to the observation of two PAP copies stably bound by the CPSF30-hFip1 subcomplex in isolation (Hamilton et al., 2020)), and the detection of two Pap1 copies in the polymerase module of yeast CPF (Casañal et al., 2017). It is not clear why the binding of a second PAP molecule by mPSF is precluded, even though our results imply that PAP can be recruited via either hFip1 molecule. We speculate that this might be due to molecular crowding or steric hindrance when mPSF is bound to a substrate RNA, particularly considering that the two Fip1 molecules make asymmetric interactions with mPSF. Notwithstanding, these findings suggest that mPSF contains two hFip1 interaction modules to ensure efficient PAP recruitment. Furthermore, the presence of two hFip1 copies might be required for mPSF integrity and its interactions with CstF.

The interaction between CPSF and CstF has previously been shown to involve direct contacts between the CstF77 homodimer and an extensive interface provided by the CPSF160 and WDR33 subunits of CPSF (Zhang et al., 2019), yet CstF also interacts with CPSF via hFip1 (Kaufmann et al., 2004). Our crystal structure of the hFip1-CstF77 subcomplex reveals that a hFip1 binds via conserved motif within the N-terminal “acidic” region to the convex arch of the CstF77 HAT domain on both protomers in the CstF77 homodimer, resulting in a 2:2 stoichiometry. By reanalysis of previously reported cryo-EM data (Zhang et al., 2019), we reveal that this interaction mode is preserved in the context of the mPSF-CstF complex. Strikingly, CstF77 inhibits 3’ polyadenylation in a manner both dependent and independent of its interaction with hFip1. Accordingly, the hFip1-PAP and hFip1-CstF77 interactions appear to be competitive, possibly as a result of the proximity of the PAP and CstF77 interaction sites within the hFip1 N-terminus. As CstF is strictly required for CPSF73-dependent pre-mRNA cleavage while PAP might not be (Boreikaite et al., 2022), these results imply that the CPSF-CstF interaction is disrupted or undergoes a conformational rearrangement after cleavage to enable PAP recruitment to the cleaved pre-mRNA and subsequent 3’ polyadenylation.

To perform the coupled reaction steps of cleavage and polyadenylation, the 3’ end processing machinery likely undergoes a sequence of conformational and compositional rearrangements as polyadenylation site recognition by the mPSF module of CPSF and activation by RBBP6 triggers CstF-dependent cleavage by the mCF, after which the nascent 3’ end needs to be made accessible to PAP for subsequent poly(A) synthesis (Boreikaite et al., 2022; Schmidt et al., 2022). Based on our structural and biochemical findings, we propose a model in which hFip1 acts as a coordinator of the two steps of

3' end processing. Initially, the two hFip1 molecules present in mPSF facilitate the assembly of CPSF and CstF on the pre-mRNA via the interactions of their N-terminal motifs with CstF77 (**Figure 4C**). In part, these interactions also preclude PAP recruitment until the pre-mRNA has been cleaved and a free 3' end has been generated. Upon endonucleolytic cleavage of the pre-mRNA by CPSF73, a conformational rearrangement, possibly driven by the dissociation of the downstream cleavage product and concomitant displacement of CstF, reduces sterical constraints around the nascent 3' end, which enables hFip1 to associate with PAP to initiate processive 3' polyadenylation of the cleaved pre-mRNA (**Figure 4D**). The conformational and compositional transitions required for accessing the nascent 3' end are orchestrated by hFip1 and facilitated by its flexible attachment to mPSF via CPSF30, as well as by its intrinsic conformational dynamics (Meinke et al., 2008; Ezeokonkwo et al., 2011; Kumar et al., 2021). The presence of two hFip1 molecules in the 3' end processing complex promotes efficient PAP recruitment and contributes to the processivity of 3' end polyadenylation. This model is supported by recent findings reported by Boreikaite et al., which demonstrated that the presence of PAP is not required for endonucleolytic cleavage by mCF, but is contradicted by the study of Schmidt et al., which reported that PAP is required for pre-mRNA cleavage, necessitating further studies focused on the functional role of PAP in pre-mRNA cleavage.

In sum, these results advance our understanding of hFip1 as a multivalent interaction scaffold for 3' end processing factors and unravel a novel aspect of polyadenylation regulation by CstF. Through interspacing binding sites for processing factors with intrinsically disordered, low-complexity sequences hFip1 can achieve the required degree of conformational freedom to accommodate the remodeling of the 3' end processing machinery and ensure correct spatiotemporal regulation of the processing factors at the nascent mRNA 3' end. The molecular basis of these transitions, however, awaits further structural and biophysical investigations.

MATERIALS AND METHODS

Protein expression and purification

Cloning for expression in E. coli

Constructs encoding for CPSF30 isoform 3 (Uniprot O95639-3), hFip1 isoform 4 (Uniprot Q6UN15-4), poly(A) polymerase alpha (Uniprot P51003-1), and CstF77 (Uniprot Q12996-1) were cloned into ligation-independent cloning (LIC) expression vectors 1B (gift from Scott Gradia, Addgene plasmid #29653), 1M (Addgene plasmid #29656), 2G-T (Addgene plasmid #29707), 2GFP-T (Addgene plasmid #29716), and co-transformation vector 13S-A (Addgene plasmid 48323), respectively. DNA encoding for hFip1$^{130\text{-}195}$ was first cloned into 2G-T, PCR amplified starting from the GST-tag and inserted into 13S-A using LIC cloning. Point mutations in CPSF30, hFip1, PAP, and CstF77 were introduced by obtaining linear DNA fragments (GeneArt Strings, Thermo Fisher) encoding for the desired construct with LIC overhangs and cloned into the respective expression vectors according to Supplementary Table 1.

Cloning for expression in Sf9 cells

DNA encoding human CPSF160 (Uniprot Q10570), WDR33 (Uniprot Q9C0J8-1), CPSF30 isoform 3, and hFip1 isoform 4 were cloned into MacroBac Series 438 cloning system vectors (Gradia et al., 2017) according to Supplementary Table 1. Subcloning of three- or four-subunit mPSF complexes into a single baculovirus transfer plasmid was performed following the MacroBac protocol (Gradia et al., 2017). For FLAG-tagged mPSF complexes, subcloning was performed using the biGBac protocol (Weissmann et al., 2016).

Expression and purification of CPSF30 and hFip1 for SEC-MALS

His_6-MBP-TEV-CPSF30$^{1\text{-}243}$ wt and mutants (Y127E, F155E, Y127E/F155E) were expressed overnight in *E. coli* BL21 star (DE3) cells and His_6-GFP-TEV-hFip1$^{1\text{-}195}$ in *E. coli* Rosetta2 (DE3) cells, respectively, at 18 °C by addition of IPTG to a final concentration of 0.5 mM at OD_{600} of about 0.6-0.8. Cells were resuspended in buffer A supplemented with 0.5 mM TCEP, 1 µM Pepstatin, and 400 µM AEBSF protease inhibitor followed by lysis via sonication. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and clarified lysate was purified on Ni-NTA agarose resin (QIAGEN) eluted with buffer A supplemented with 0.5 mM TCEP and 200 mM imidazole. The protein was further purified by size exclusion chromatography on a Superdex 75 (Cytiva) column, eluting with buffer A supplemented with

1 mM DTT. Eluting peak fractions were concentrated in centrifugal filter (Amicon Ultra-15, MWCO 30 kDa, Merck Millipore), flash frozen, and stored at -80 °C.

PAP expression and purification

His_6-MBP-TEV-PAP^{1-504} was expressed in *E. coli* Rosetta2 (DE3) cells overnight at 18 °C by induction with 0.5 mM IPTG at OD_{600} of about 0.6-0.8. Cells were lysed by sonication in 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.5 mM TCEP supplemented with 0.1 µM Pepstatin and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and cleared lysate was subjected to Ni-NTA resin (QIAGEN), washed and protein eluted with buffer B (20 mM Tris-HCl pH 8.0, 200 mM NaCl) supplemented with 200 mM imidazole. Eluted protein was further purified on MBPTrap HP (Cytiva), eluting in buffer B supplemented with 10 mM maltose. The eluted protein fractions were injected onto a Superdex 200 column (Cytiva) equilibrated in buffer B supplemented with 1 mM DTT. Tag was cleaved off the protein with His_6-MBP-TEV protease, and the cleavged tags including protease removed from protein sample using a MBPTrap HP (Cytiva). For use in pull-down analysis, the tag was not cleaved from His_6-MBP-TEV-PAP^{1-504} wt and mutant (R395A, R402A, K431A) after size exclusion chromatography. Purified protein was concentrated in centrifugal filter (Amicon Ultra-15, MWCO 50 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C.

Expression of mPSF complexes

For expression of mPSF complexes in *Sf9* cells (Thermo Fisher Scientific, cat. no. 11496015; cell line was authenticated and tested for mycoplasma contamination by manufacturer, no further validation was done by the authors), recombinant baculoviruses were generated according to the Bac-to-Bac Baculovirus expression system (Invitrogen). 2 ml of P3 virus were used to infect 1 L of *Sf9* insect cells at a density of 1.1×10^6 ml^{-1}. Cells were harvested 72 h post infection.

Purification of CPSF complexes for polyadenylation assays and pull-down analysis

Cells were resuspended in buffer C (25 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% Glycerol, 0.5 mM TCEP) supplemented with 5 mM imidazole, 0.05% Tween-20, and cOmplete Protease-Inhibitor-Cocktail (Roche). Cells were lysed by sonication, cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and the clarified lysate was purified on Ni-NTA resin (QIAGEN) eluting in buffer C supplemented with 200 mM imidazole. The eluted protein was incubated with Strep-Tactin Sepharose (IBA Lifesciences) beads, washed with ten column volumes of buffer B, and eluted with buffer C supplemented with 5 mM

Desthiobiotin. Strep-Tactin purified mPSF complexes were concentrated in centrifugal filter (Amicon Ultra-15, MWCO 300 kDa, Merck Millipore) to approximately 0.5 $mg \cdot ml^{-1}$. To account for impurities, mPSF complex concentrations were assessed on SDS PAGE and adjusted accordingly (Figure S1C), aliquoted, flash frozen, and stored at -80 °C. For use in pull-down analysis, CPSF complexes were used directly after Strep-Tactin purification.

Purification of CPSF-PAP complex for SEC-MALS analysis

mPSF complexes comprising CPSF160-WDR33$^{1-410}$-CPSF30-hFip1$^{1-198}$ for SEC-MALS analysis were produced as described above with subsequent tag removal by incubation with His_6-TEV protease. mPSF (assuming to comprise two hFip1) was supplemented with untagged PAP$^{1-504}$ in 2.5-fold molar excess and 1.2-fold molar excess of 27 nt SV40 PAS-containing mRNA with cryptic polyA tail (CUGCAAUAAACAACUUAACAACAAAAA). The complex was purified on a Superose 6 column (Cytiva) in 20 mM HEPES pH 8.0, 150 mM KCl, 0.5 mM TCEP. mPSF:PAP complex was concentrated in centrifugal filter (Amicon Ultra-15, MWCO 100 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C.

Expression and purification of GST-hFip1 proteins for pull-down analysis

His_6-GST-TEV-hFip1$^{1-195}$ and His_6-GST-TEV-hFip1$^{1-35}$ were expressed overnight in *E. coli* BL21 star (DE3) cells at 18 °C by induction with 0.5 mM IPTG at OD_{600} of about 0.6-0.8. Cells were lysed by sonication in buffer A supplemented with 5 mM imidazole, 1 µM Peptsatin A, and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and clarified lysate was purified on Ni-NTA resin (QIAGEN) eluting in buffer A supplemented with 200 mM imidazole in gravity flow. Eluted protein was loaded on a HiTrap Q FF (Cytiva) anion exchange chromatography column and eluted with a linear gradient from 200 mM to 1 M NaCl over 15 CV in 25 mM Tris-HCl pH 7.5 and 1 mM DTT. Eluting peak fractions were further purified on a Superdex 75 (Cytiva) column equilibrated in 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 1 mM DTT. Protein was concentrated in centrifugal filter (Amicon Ultra-15, MWCO 30 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C. His_6-GST-TEV-hFip1$^{36-195}$ was expressed in *E. coli* BL21 star (DE3) cells overnight at 18 °C by induction with 0.5 mM IPTG at OD_{600} of about 0.6-0.8. Cells were lysed by sonication in buffer A supplemented with 1 mM DTT, 1 µM Peptsatin A, and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C). Clarified lysate was subjected to a GSTrap Fast Flow (Cytiva) column and eluted in buffer A supplemented with 1 mM DTT and 10 mM GSH. His_6-GST-

TEV-hFip1$^{1-35}$ mutants (E22A, E23A; W25A, L26A, Y27A; W25A) were expressed and purified analogously to His$_6$-GST-TEV-hFip1$^{36-195}$ but with buffers containing 500 mM NaCl. All proteins were further purified on a Superdex 200 (Cytiva) column equilibrated in buffer A supplemented with 1 mM DTT. Protein was concentrated in centrifugal filter (Amicon Ultra-15, MWCO 30 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C.

Expression and purification of His$_6$-CstF77 proteins for pull-down analysis

His$_6$-TEV-CstF77$^{21-549}$ wt and mutant (R395A, R402A, K431A) were expressed overnight in *E. coli* BL21 star (DE3) cells at 18 °C by induction with 0.5 mM IPTG at OD$_{600}$ of about 0.6-0.8. Cells were lysed by sonication in 40 mM Tris-HCl pH 7.5, 500 mM NaCl, 5 mM imidazole, supplemented with 1 µM Peptsatin A and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and clarified lysate was purified on Ni-NTA resin (QIAGEN) eluting with buffer A supplemented with 250 mM imidazole in gravity flow. Salt concentration and pH of protein sample were reduced to 60 mM NaCl and pH 7.0 by dilution and purified on a HiTrap SP FF (Cytiva) cation exchange chromatography column. Protein was eluted from column with a linear gradient from 60 mM to 1 M NaCl over 10 CV in 20 mM Tris-HCl pH 7.0 and 1 mM DTT. Eluting peak fractions were further purified on a Superdex 200 (Cytiva) column equilibrated in 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM DTT. Protein was concentrated in centrifugal filter (Amicon Ultra-15, MWCO 100 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C.

Expression and purification of MBP-CstF77 proteins for pull-down analysis

For use in pull-down analysis, His$_6$-MBP-TEV-CstF77$^{21-549}$ wt and mutant (R395A, R402A, K431A) were expressed and purified analogous to CstF77 for co-crystallization with hFip1, omitting tag cleavage with His$_6$-TEV protease prior to size exclusion chromatography.

Expression and purification of GFP-PAP for pull-down analysis

His$_6$-GFP-TEV-PAP$^{1-504}$ was expressed overnight in *E. coli* Rosetta2 (DE3) cells at 18 °C by induction with 0.5 mM IPTG at OD$_{600}$ of about 0.6-0.8. Cells were lysed by high-pressure cell disruption at 25 kpsi in buffer B supplemented with 5 mM imidazole, 0.5 mM TCEP, 0.1 µM Pepstatin, and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and clarified lysate was subjected to Ni-NTA resin (QIAGEN), washed and protein eluted with buffer B supplemented with 250 mM imidazole. Eluted protein was further purified on Superdex 200 column (Cytiva) equilibrated in

20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5 mM TCEP. Purified protein was concentrated in centrifugal filter (Amicon Ultra-15, MWCO 50 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C.

CPSF30-hFip1 complex preparation

Plasmids encoding for His_6-TEV-CPSF30$^{118-178}$ and GST-TEV-hFip1$^{130-195}$ were co-transformed and proteins were expressed overnight in *E. coli* BL21 star (DE3) cells at 18 °C by addition of IPTG to a final concentration of 0.5 mM at OD_{600} of about 0.6-0.8. Cells were resuspended in buffer A (25 mM Tris-HCl pH 7.5, 200 mM NaCl) supplemented with 1 µM Pepstatin and 400 µM AEBSF, and lysed by sonication. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and protein was purified on Glutathione Sepharose 4 Fast Flow resin (Cytiva), eluting with buffer A supplemented with 10 mM reduced L-Glutathione (GSH). After overnight cleavage with His_6-TEV protease, the protein was further purified by size exclusion chromatography on a Superdex 75 (Cytiva) column with a GSTrap Fast Flow (Cytiva) column in tandem to capture any residual GST tags, eluting with 25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT. Eluting peak fractions were concentrated in centrifugal filter (Amicon Ultra-15, MWCO 10 kDa, Merck Millipore) to 6.14 $mg \cdot ml^{-1}$, flash frozen, and stored at -80 °C.

CstF77-hFip1 complex preparation

His_6-MBP-TEV-CstF77$^{241-549}$ was expressed in *E. coli* BL21 star (DE3) cells at 18 °C overnight by addition of IPTG to a final concentration of 0.5 mM at OD_{600} of about 0.6-0.8. Cells were resuspended in buffer containing buffer A supplemented with 1 mM DTT, 1 µM Pepstatin and 400 µM AEBSF, and lysed by sonication. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and protein was purified on amylose resin (NEB) including a high salt wash with buffer containing 25 mM Tris-HCl pH 7.5, 500 mM NaCl, and 1 mM DTT prior to elution with buffer A supplemented with 1 mM DTT and 10 mM maltose. After digestion with His_6-TEV protease, the tags and protease were removed from the protein by passage through a Ni-NTA superflow cartridge (QIAGEN). The protein was further purified by size exclusion chromatography on a Superdex 200 Increase (Cytiva) column, eluting with 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM TCEP. His_6-GST-TEV-hFip1$^{1-35}$ was expressed in *E. coli* BL21-AI (Invitrogen) cells overnight at 18 °C by induction with 0.2% arabinose at OD_{600} of 0.8. Cells were lysed by high-pressure cell disruption at 25 kpsi in buffer A supplemented with 1 mM DTT, 1 µM Pepstatin and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C) and cleared lysate was subjected to a GSTrap Fast Flow (Cytiva) column, washed with 25 mM Tris-HCl pH 7.5, 500 mM NaCl, and 1 mM DTT prior to elution in buffer A supplemented with 1 mM DTT and

10 mM GSH. Affinity tag was cleaved from protein using His_6-MBP-TEV protease while dialyzing into buffer A supplemented with 1 mM DTT and $hFip1^{1-35}$ was further purified by size exclusion chromatography on a Superdex 75 (Cytiva) column into 25 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT. The absolute mass of $hFip1^{1-35}$ (4.1 kDa) was confirmed with ESI-MS analysis. Peak fractions of both CstF77 and hFip1 were pooled individually, concentrated, flash frozen, and stored at -80 °C.

CPSF30-hFip1 complex crystallization and structure determination

CPSF30:hFip1 complex was crystallized at 20 °C using the hanging drop vapour diffusion method by mixing 0.5 µl of protein at 6.14 $mg \cdot ml^{-1}$ with 0.5 µl of reservoir solution containing either 1.8 M $(NH_4)SO_4$, 0.1 M Bis-Tris pH 6.5 (native data set) or 1.626 M $(NH_4)SO_4$, 0.1 M Bis-Tris pH 6.5 (zinc SAD data set). Crystals were transferred into reservoir solution supplemented with 20% (v/v) Glycerol for cryo-protection prior to flash-cooling by plunging into liquid nitrogen. X-ray diffraction data were recorded at beam line X06DA (PXIII) at Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) on an PILATUS 2M-F (Dectris) detector, at a wavelength of 1.28095 Å using an oscillation range of 0.1° and an exposure time of 0.1 s per image while rotating the crystal through 360°. Detailed data collection statistics are listed in Table 1. Diffraction data were processed with XDS (Kabsch, 2010) in space group $P2_1$, with four complex copies in the asymmetric unit and the presence of pseudmerohedral twinning. Twin law h, -k, -h-l was determined using phenix.xtriage (Zwart et al., 2005) comprising a twin fraction of approximately 48%. Exploiting the presence of zinc ions bound to CPSF30, phase determination was performed by single-wavelength anomalous diffraction (SAD) followed by phasing and density modification with autoSHARP (Vonrhein et al., 2007). A homology model based on CPSF30 ZF2 (PDB ID: 6FUW) was fitted into the electron density in Coot (Emsley and Cowtan, 2004), followed by automated model building using phenix.autobuild (Terwilliger et al., 2007). The structure was completed by iterative cycles of manual model-building in Coot and refinement with phenix.refine (Adams et al., 2010). Molecular models were visualized using PyMOL (Schrödinger LLC, 2021).

CstF77-hFip1 complex structure elucidation

A 1.5-fold molar excess of hFip1 was added per CstF77 molecule (corresponding to a 3-fold molar excess to a dimer of CstF77) and concentrated to 13.7 $mg \cdot ml^{-1}$ (A280=23.36) using a centrifugal filter (Amicon Ultra-0.5, MWCO 3 kDa, Merck Millipore) prior to crystallization. The CstF77-hFip1 complex was crystallized using the sitting drop vapor diffusion method by mixing 0.1 µl protein with 0.1 µl

reservoir solution containing 0.1 M Bicine pH 9.0, 10% w/v PEG 20k, 2% v/v Dioxane. Crystals were cryo-protected by transfer into reservoir solution supplemented with 24% (v/v) Glycerol prior to flash-cooling with liquid nitrogen. X-ray diffraction data were recorded at beam line X06SA (PXI) at Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) on an EIGER 16M (Dectris) detector, using an oscillation range of 0.2° and an exposure time of 0.1 s per image while rotating the crystal through 360°. Detailed data collection statistics are listed in Table 1. Diffraction data were processed with Autoproc (Vonrhein et al., 2011) in space group $P6_122$. The structure was solved using residues 241-549 of murine CstF77 (PDB ID: 2OOE) as search model for phasing with molecular replacement (MR) in phenix.phaser (McCoy et al., 2007). A total of two CstF77 molecules could be placed into the electron density, corresponding to a dimer. After rigid-body refinement of the molecular replacement solution, the structure was completed by iterative cycles of manual model-building in Coot, including the placement of the hFip1 peptides into the electron density unoccupied by CstF77, and refinement with phenix.refine (Adams et al., 2010). Molecular models were visualized using PyMOL (Schrödinger LLC, 2021).

Pull-down assays

Pull-down analysis of mPSF-PAP interaction

Strep-Tactin purified mPSF complexes were incubated with 30 µl Anti-FLAG M2 magnetic beads (Sigma-Aldrich) equilibrated in FLAG wash buffer (25 mM Tris pH 8.0, 200 mM NaCl, 0.1% Tween-20) and gently agitated at 4 °C for 2 h. The beads were washed three times with 0.5 ml of FLAG wash buffer and the bound protein was eluted with 1X SDS-PAGE loading buffer supplemented with 100 $\mu g \cdot ml^{-1}$ 3X FLAG peptide (Sigma-Aldrich) on ice. FLAG elutions were analyzed by SDS-PAGE on 4-20% Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad) without prior heating to preserve the GFP fluorescence. GFP fluorescence was visualized on a Typhoon FLA 9500 laser scanner (Cytiva) at 473 nm and subsequently stained with Coomassie brilliant blue R250. For equal mPSF complex concentrations to compare the corresponding GFP-hFip1 and GFP-PAP fluorescences, loading volumes were adjusted according to CPSF160 band intensities. Beads control loading volume correspons to the maximum mPSF sample loading volume.

Pull-down analysis of hFip1-CstF77 interaction

For pull-down analysis with purified hFip1 and CstF77 proteins (wt and mutants), 10 µg of purified His_6-GST-hFip1 protein was immobilized on 15 µl Glutathione Sepharose 4 Fast Flow beads (Cytiva) and washed three times with 0.5 ml pull-down wash buffer (20 mM Tris pH 7.5, 200 mM NaCl, 0.05% Tween-20, 0.5 mM TCEP). His_6-MBP-CstF77 protein was added to the immobilized protein at 4-fold molar excess and incubated gently agitating at 4 °C for 1 h followed by washing three times with 0.5 ml of pull-down wash buffer. The bound protein was eluted at room temperature by adding 1X SDS-PAGE loading buffer and analyzed by SDS-PAGE on 4-20% Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad) stained with Coomassie brilliant blue R250.

Pull-down analysis of hFip1-CstF77-PAP interaction

For competetive pull-down analysis of both CstF77 and PAP with $hFip1^{1-195}$, 5 µg of purified His_6-GST-$hFip1^{1-195}$ protein was immobilized on 15 µl Glutathione Sepharose 4 Fast Flow beads (Cytiva) equilibrated in pull-down wash buffer, gently agitated at 4 °C for 1 h, and washed three times with 0.5 ml pull-down wash buffer. His_6-MBP-CstF77 and His_6-GFP-PAP were incubated with the bait, either individually or combined (1:1) at 4-fold molar excess, as well as adding a 32-fold molar excess of one protein while keeping the other at 4-fold molar excess, resulting in a *de facto* 8-fold excess of one protein over the other (8:1, 1:8).

In-vitro polyadenylation assays

Reaction conditions for pre-mRNA polyadenylation were adjusted for the individual need of each assay and evolved over the course of the project. To account for potential impurities and to ensure equal mPSF complex concentrations, CPSF160 band intensities were assessed on SDS-PAGE (Figure S1C) and concentrations adjusted accordingly. All polyadenylation reactions were performed in polyadenylation buffer (25 mM Tris pH 7.5, 10% Glycerol, 50 mM KCl, 2 mM $MgCl_2$, 0.05% Tween-20, 1 mM DTT) with 20 nM 5' Cy5-labelled 27-nucleotide mRNA with cryptic poly(A) (five A's) tail as template. All proteins were first diluted in polyadenylation buffer. Protein-RNA mixes with a total volume of 36 µl were prepared on ice, pre-heated at 37 °C for 1 min and reaction was started by the addition of pre-heated 12 µl ATP at 37 °C. Reaction mix for polyadenylation assay with CPSF30 ZF4/ZF5 mutants (Figure 2A) contained 80 nM mPSF complexes, 1.46 µM PAP, and a final concentration of 4 µM ATP. Reaction mix for polyadenylation assay with hFip1 truncations (Figure 2C) contained 40 nM mPSF complexes, 120 nM PAP, and a final concentration of 500 µM ATP. Reaction mix for polyadenylation

assay with CstF77 at 80 nM (denoted 0.5), 160 nM (denoted + or 1), 320 nM, 640 nM, or 1280 nM (denoted 2, 4, 8, respectively) (Figure 4B) contained 80 nM mPSF complexes, 160 nM PAP, and a final concentration of 500 µM ATP. Time points were taken at indicated times (1 min, 10 min) and polyadenylation stopped by the addition of EDTA with final concentration of 166 mM and incubation with 20 µg Proteinase K at 37 °C for 10 mins. The reactions were mixed with 2X denaturing PAGE loading dye (90% Formamide, 5% Glycerol, 25 mM EDTA, Bromphenol blue), incubated at 95 °C for 10 minutes and analyzed on a 15% denaturing PAGE gel containing 8 M urea and 0.5X TBE. In-gel fluorescence of 5' Cy5-labelled RNA was visualized with Typhoon FLA 9500 laser scanner (Cytiva) at 635 nm.

SEC-MALS analysis

Size exclusion chromatography combined with multiangle light-scattering (SEC-MALS) was carried out on an HPLC system (Agilent LC1100, Agilent Technologies) coupled to an Optilab rEX refractometer and a miniDAWN three-angle light-scattering detector (Wyatt Technology). Data analysis was performed using the ASTRA software (version 7.3.2; Wyatt Technology).

SEC-MALS analysis of hFip1-CPSF30 complex

For unambigous determination of the stoichiometry of the respective hFip1-CPSF30 complexes, tagged proteins were used to increase the molecular weight difference between the 2:1 and 1:1 complexes of hFip1-CPSF30. Stoichiometry of the complexes was determined injecting 33 µg His_6-MBP-$CPSF30^{1-243}$ (wt and mutants) and 4-fold molar excess of His_6-GFP-$hFip1^{1-195}$ pre-mixed in a total injection volume of 100 µl. Proteins were separated on a Superdex 200 10/300 GL column (Cytiva) run at 0.5 ml/min at room temperature in 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 0.5 mM TCEP (pH was adjusted at room temperature).

SEC-MALS analysis of mPSF-PAP complex

Stoichiometry of the mPSF-PAP complex was determined injecting 50 µg of pre-purified mPSF-PAP comprising CPSF160-$WDR33^{1-410}$-CPSF30-2x$hFip1^{1-198}$, PAP^{1-504}, and 27-nt SV40 PAS-containing mRNA in a total injection volume of 100 µl. In a second run, pre-purified 50 µg of mPSF:PAP was spiked with additional 41.6 µg PAP^{1-504} (5-fold molar excess) in a total injection volume of 100 µl to test whether excess PAP can lead to a stable 1:2 complex of mPSF and PAP. Proteins were separated on a

587 Superose 6 10/300 GL column (Cytiva) run at 0.5 ml/min at room temperature in 20 mM HEPES pH
588 8.0, 150 mM KCl, 0.5 mM TCEP (pH was adjusted at room temperature).

589 ***Multiple sequence alignment***

590 The multiple sequence alignment of hFip1 orthologs was produced with MAFFT version 7 (Katoh et al.,
591 2018) and visualized using Jalview (Waterhouse et al., 2009). Input sequences are listed in
592 Supplementary Table 3.

593 ***Analysis of interaction interfaces***

594 Buried surface area of the hFip1-CPSF30 interaction interface was calculated using the PDBePISA
595 (Proteins, Interfaces, Structurs and Assemblies) tool (Krissinel and Henrick, 2007).

596 ***3D density map analysis***

597 Visualization and analysis of the 3D density map for CPSF160-WDR33-CPSF30-PAS RNA-CstF77
598 complex (EMD-20861) was performed with UCSF Chimera (Pettersen et al., 2004), developed by the
599 Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.
600 The 3D density map was segmented and color-coded based on the corresponding atomic model (PDB
601 ID: 6URO). The CstF77-hFip1 crystal structure from this study was superimposed onto the atomic
602 model of CstF77.

DATA AVAILABILITY

The atomic coordinates and structure factors for the crystallographic structures of the Fip1-CPSF30Fip1 and Fip1-CstF77 complexes have been deposited in the Protein Data Bank under accession codes 7ZYH and 7ZY4, respectively. All data generated or analyzed during this study are included in the manuscript and supporting files. Source data files for gel images in Figures 1, 2, 3, 4, and associated Figure Supplements are provided.

FUNDING

This work was supported by Boehringer Ingelheim Fonds PhD Fellowship and by the National Center for Competence in Research (NCCR) RNA & Disease, funded by the Swiss National Science Foundation.

ACKNOWLEDGEMENTS

We thank Birgit Dreyer for assistance with SEC-MALS, Beat Blattmann (University of Zurich Protein Crystallization Center) for performing crystallization screens, and Levi Kopp for assistance with protein crystallization. We thank Vincent Olieric, Takashi Tomizaki and Meitian Wang (Swiss Light Source, Paul Scherrer Institute) for assistance with crystallographic data collection. We are grateful to Stefanie Jonas and members of the Jinek laboratory for critical reading of the manuscript.

REFERENCES

Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* **66**:213–221. doi:10.1107/S0907444909052925

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**:W344–W350. doi:10.1093/nar/gkw408

Bai Y, Auperin TC, Chou C-YY, Chang G-GG, Manley JL, Tong L. 2007. Crystal Structure of Murine CstF-77: Dimeric Association and Implications for Polyadenylation of mRNA Precursors. *Mol Cell* **25**:863–875. doi:10.1016/J.MOLCEL.2007.01.034

Barabino SML, Hübner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997. The 30-kd subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are rna-binding zinc finger proteins. *Genes Dev* **11**:1703–1716. doi:10.1101/gad.11.13.1703

Barabino SML, Ohnacker M, Keller W. 2000. Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J* **19**:3778–3787. doi:10.1093/emboj/19.14.3778

Bienroth S, Wahle E, Suter-Crazzolara C, Keller W. 1991. Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J Biol Chem* **266**:19768–19776. doi:10.1016/s0021-9258(18)55058-4

Boreikaite V, Elliott TS, Chin JW, Passmore LA. 2022. RBBP6 activates the pre-mRNA 3' end processing machinery in humans. *Genes Dev* **36**:210–224. doi:10.1101/gad.349223.121

Casañal A, Kumar A, Hill CH, Easter AD, Emsley P, Degliesposti G, Gordiyenko Y, Santhanam B, Wolf J, Wiederhold K, Dornan GL, Skehel M, Robinson C V., Passmore LA. 2017. Architecture of eukaryotic mRNA 3′-end processing machinery. *Science (80-)* **358**:1056–1059. doi:10.1126/science.aao6535

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

Clerici M, Faini M, Aebersold R, Jinek M. 2017. Structural insights into the assembly and polyA signal recognition mechanism of the human CPSF complex. *Elife* **6**:e33111. doi:10.7554/eLife.33111

Clerici M, Faini M, Muckenfuss LM, Aebersold R, Jinek M. 2018. Structural basis of AAUAAA polyadenylation signal recognition by the human CPSF complex. *Nat Struct Mol Biol* **25**:135–138. doi:10.1038/s41594-017-0020-6

Dharmalingam P, Mahalingam R, Yalamanchili HK, Weng T, Karmouty-Quintana H, Guha A, Thandavarayan R. 2022. Emerging roles of alternative cleavage and polyadenylation (APA) in human disease. *J Cell Physiol* **237**:149–160. doi:10.1002/jcp.30549

Di Giammartino DC, Li W, Ogami K, Yashinskie JJ, Hoque M, Tian B, Manley JL. 2014. RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 39 UTRs. *Genes Dev* **28**:2248–2260. doi:10.1101/gad.245787.114

Di Giammartino DC, Nishida K, Manley JL. 2011. Mechanisms and Consequences of Alternative Polyadenylation. *Mol Cell* **43**:853–866. doi:10.1016/J.MOLCEL.2011.08.017

Emsley P, Cowtan K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* **60**:2126–2132. doi:10.1107/S0907444904019158

Ezeokonkwo C, Zhelkovsky A, Lee R, Bohm A, Moore CL. 2011. A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation. *RNA* **17**:652–64. doi:10.1261/rna.2273111

Ghazy MA, He X, Singh BN, Hampsey M, Moore C. 2009. The Essential N Terminus of the Pta1 Scaffold Protein Is Required for snoRNA Transcription Termination and Ssu72 Function but Is Dispensable

for Pre-mRNA 3′-End Processing. *Mol Cell Biol* **29**:2296–2307. doi:10.1128/MCB.01514-08

Gradia SD, Ishida JP, Tsai MS, Jeans C, Tainer JA, Fuss JO. 2017. MacroBac: New Technologies for Robust and Efficient Large-Scale Production of Recombinant Multiprotein ComplexesMethods in Enzymology. Academic Press. pp. 1–26. doi:10.1016/bs.mie.2017.03.008

Gruber AJ, Zavolan M. 2019. Alternative cleavage and polyadenylation in health and disease. *Nat Rev Genet*. doi:10.1038/s41576-019-0145-z

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

Helmling S, Zhelkovsky A, Moore CL. 2001. Fip1 Regulates the Activity of Poly(A) Polymerase through Multiple Interactions. *Mol Cell Biol* **21**:2026–2037. doi:10.1128/mcb.21.6.2026-2037.2001

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

Hollerer I, Grund K, Hentze MW, Kulozik AE. 2014. mRNA 3′end processing: A tale of the tail reaches the clinic. *EMBO Mol Med* **6**:16–26. doi:10.1002/emmm.201303300

Kabsch W. 2010. XDS. *Acta Crystallogr Sect D Biol Crystallogr* **66**:125–132. doi:10.1107/S0907444909047337

Katoh K, Rozewicki J, Yamada KD. 2018. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**:1160–1166. doi:10.1093/bib/bbx108

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

Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**:774–97. doi:10.1016/j.jmb.2007.05.022

Kumar A, Clerici M, Muckenfuss LM, Passmore LA, Jinek M. 2019. Mechanistic insights into mRNA 3′-end processing. *Curr Opin Struct Biol*. doi:10.1016/j.sbi.2019.08.001

Kumar A, Yu CWH, Rodríguez-Molina JB, Li X-H, Freund SMV, Passmore LA. 2021. Dynamics in Fip1 regulate eukaryotic mRNA 3′ end processing. *Genes Dev* **35**:1510–1526. doi:10.1101/gad.348671.121

MacDonald CC, Wilusz J, Shenk T. 1994. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol* **14**:6647–6654. doi:10.1128/mcb.14.10.6647-6654.1994

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. 2006. Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. *Nature* **444**:953–956. doi:10.1038/nature05363

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**:658–674. doi:10.1107/S0021889807021206

Meinke G, Ezeokonkwo C, Balbo P, Stafford W, Moore C, Bohm A. 2008. Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered proteid. *Biochemistry* **47**:6859–6869. doi:10.1021/bi800204k

Murthy KGK, Manley JL. 1992. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* **267**:14804–14811. doi:10.1016/s0021-9258(18)42111-4

Nicholson AL, Pasquinelli AE. 2019. Tales of Detailed Poly(A) Tails. *Trends Cell Biol* **29**:191–200. doi:10.1016/j.tcb.2018.11.002

713 Ohnacker M, Barabino SML, Preker PJ, Keller W. 2000. The WD-repeat protein Pfs2p bridges two
714 essential factors within the yeast pre-mRNA 3'-end-processing complex. *EMBO J* **19**:37–47.
715 doi:10.1093/emboj/19.1.37

716 Passmore LA, Coller J. 2022. Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression.
717 *Nat Rev Mol Cell Biol*. doi:10.1038/s41580-021-00417-y

718 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF
719 Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**:1605–
720 1612. doi:10.1002/jcc.20084

721 Preker PJ, Lingner J, Minvielle-Sebastia L, Keller W. 1995. The FIP1 gene encodes a component of a
722 yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase, Cell.
723 doi:10.1016/0092-8674(95)90391-7

724 Proudfoot NJ. 2011. Ending the message: Poly(A) signals then and now. *Genes Dev* **25**:1770–1782.
725 doi:10.1101/gad.17268411

726 Proudfoot NJ, Brownlee GG. 1976. 3′ Non-coding region sequences in eukaryotic messenger RNA.
727 *Nature* **263**:211–214. doi:10.1038/263211a0

728 Schmidt M, Kluge F, Sandmeir F, Kühn U, Schäfer P, Tüting C, Ihling C, Conti E, Wahle E. 2022.
729 Reconstitution of 3′ end processing of mammalian pre-mRNA reveals a central role of RBBP6.
730 *Genes Dev* **36**:195–209. doi:10.1101/gad.349217.121

731 Schönemann L, Kühn U, Martin G, Schäfer P, Gruber AR, Keller W, Zavolan M, Wahle E. 2014.
732 Reconstitution of CPSF active in polyadenylation: Recognition of the polyadenylation signal by
733 WDR33. *Genes Dev* **28**:2381–2393. doi:10.1101/gad.250985.114

734 Schrödinger LLC. 2021. The PyMOL Molecular Graphics System, Version 2.5.0.

735 Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL, Giammartino
736 DC Di, Taylor D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL, Di Giammartino DC, Taylor
737 D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL. 2009. Molecular Architecture of the
738 Human Pre-mRNA 3′ Processing Complex. *Mol Cell* **33**:365–376.
739 doi:10.1016/j.molcel.2008.12.028

740 Sullivan KD, Steiniger M, Marzluff WF. 2009. A Core Complex of CPSF73, CPSF100, and Symplekin
741 May Form Two Different Cleavage Factors for Processing of Poly(A) and Histone mRNAs. *Mol*
742 *Cell* **34**:322–332. doi:10.1016/j.molcel.2009.04.024

743 Sun Y, Zhang Y, Hamilton K, Manley JL, Shi Y, Walz T, Tong L. 2018. Molecular basis for the
744 recognition of the human AAUAAA polyadenylation signal. *Proc Natl Acad Sci U S A* **115**:E1419–
745 E1428. doi:10.1073/pnas.1718723115

746 Takagaki Y, MacDonald CC, Shenk T, Manley JL. 1992. The human 64-kDa polyadenylylation factor
747 contains a ribonucleoprotein- type RNA binding domain and unusual auxiliary motifs. *Proc Natl*
748 *Acad Sci U S A* **89**:1403–1407. doi:10.1073/pnas.89.4.1403

749 Takagaki Y, Manley JL. 1997. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol*
750 **17**:3907–3914. doi:10.1128/mcb.17.7.3907

751 Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T. 1990. A multisubunit factor, CstF, is
752 required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* **4**:2112–2120.
753 doi:10.1101/gad.4.12a.2112

754 Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung LW, Read RJ,
755 Adams PD. 2007. Iterative model building, structure refinement and density modification with the
756 PHENIX AutoBuild wizardActa Crystallographica Section D: Biological Crystallography.
757 International Union of Crystallography. pp. 61–69. doi:10.1107/S090744490705024X

758 Tian B, Manley JL. 2016. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol*.
759 doi:10.1038/nrm.2016.116

760 Venkataraman K, Brown KM, Gilmartin GM. 2005. Analysis of a noncanonical poly(A) site reveals a

761 tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19**:1315–1327.
762 doi:10.1101/gad.1298605

763 Vonrhein C, Blanc E, Roversi P, Bricogne G. 2007. Automated structure solution with autoSHARP.
764 *Methods Mol Biol* **364**:215–230. doi:10.1385/1-59745-266-1:215

765 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-A multiple
766 sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189–1191.
767 doi:10.1093/bioinformatics/btp033

768 Weissmann F, Petzold G, VanderLinden R, Huis In't Veld PJ, Brown NG, Lampert F, Westermann S,
769 Stark H, Schulman BA, Peters JM. 2016. BiGBac enables rapid gene assembly for the expression
770 of large multisubunit protein complexes. *Proc Natl Acad Sci U S A* **113**:E2564–E2569.
771 doi:10.1073/pnas.1604935113

772 Xiang K, Tong L, Manley JL. 2014. Delineating the structural blueprint of the pre-mRNA 3'-end
773 processing machinery. *Mol Cell Biol* **34**:1894–1910. doi:10.1128/MCB.00084-14

774 Yang W, Hsu PL, Yang F, Song JE, Varani G. 2018. Reconstitution of the CstF complex unveils a
775 regulatory rolefor CstF-50 in recognition of 3-end processing signals. *Nucleic Acids Res* **46**:493–
776 503. doi:10.1093/nar/gkx1177

777 Zhang Y, Sun Y, Shi Y, Walz T, Tong L. 2019. Structural Insights into the Human Pre-mRNA 3′-End
778 Processing Machinery. *Mol Cell* **77**:800-809.e6. doi:10.1016/j.molcel.2019.11.005

779 Zhao J, Hyman L, Moore C. 1999. Formation of mRNA 3' Ends in Eukaryotes: Mechanism, Regulation,
780 and Interrelationships with Other Steps in mRNA Synthesis. *Microbiol Mol Biol Rev* **63**:405–445.
781 doi:10.1128/MMBR.63.2.405-445.1999

782 Zwart, P. H., Grosse-Kunstleve, R. W., Adams PD. 2005. Xtriage and Fest: automatic assessment of
783 X-ray data and substructure structure factor estimation. *CCP4 Newsl* **43**:27–35.
784 doi:10.1.1.378.2281

785

FIGURES

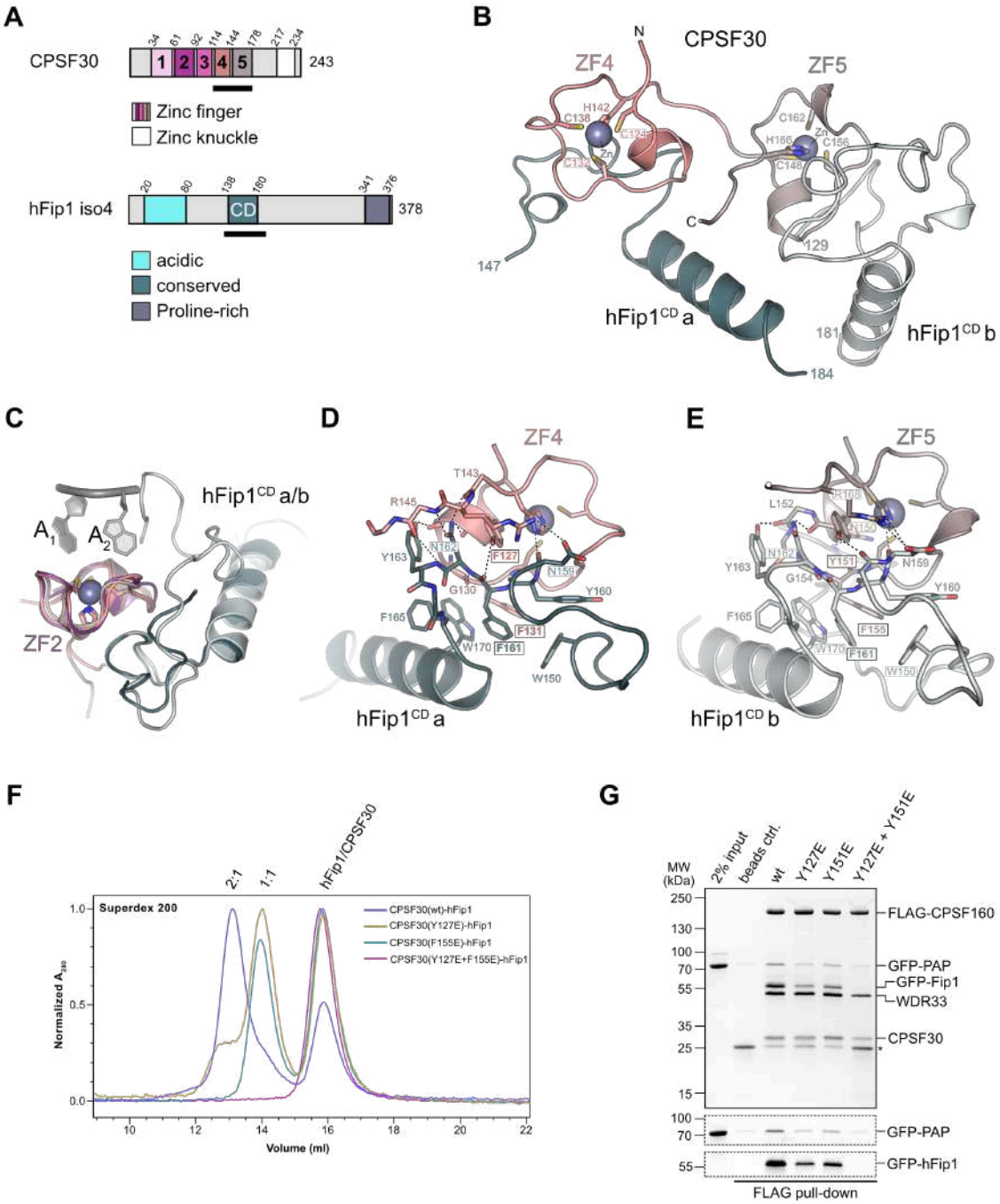


Figure 1: hFip1 interacts with CPSF30 with 2:1 stoichiometry. (**A**) Schematic representation of domain architecture of CPSF30 and hFip1. CPSF30 consists of five zinc finger (ZF) domains and a zinc knuckle domain. hFip1 isoform 4 comprises acidic, conserved, and proline-rich regions but lacks the RE/D region interacting with CF I_m, as well as the R-rich region, which has been shown to bind U-rich RNA in hFip1 isoform 1 (Kaufmann et al., 2004). (**B**) Cartoon representation of the crystal structure of $CPSF30^{ZF4\text{-}ZF5}$ in complex with two hFip1 fragments comprising the conserved domain (CD). (**C**) Superposition of CPSF30 ZF2 domain in complex with PAS RNA onto ZF4 and ZF5. (**D**) Detailed interaction interface of $hFip1^{CD}$ with CPSF30 ZF4. (**E**) Detailed interaction interface of $hFip1^{CD}$ with CPSF30 ZF5. (**F**) SEC-MALS chromatogram of MBP-$CPSF30^{ZF4\text{-}ZF5}$ selective hFip1 binding mutants for stoichiometry analysis with GFP-hFip1. (**G**) *In vitro* pull-down analysis of FLAG-epitope tagged mPSF comprising wild-type CPSF30 and its selective hFip1 binding mutants with GFP-PAP. Asterisk indicates anti-FLAG M2 antibody light chain. GFP-hFip1 and GFP-PAP are also visualized with in-gel GFP fluorescence (bottom).

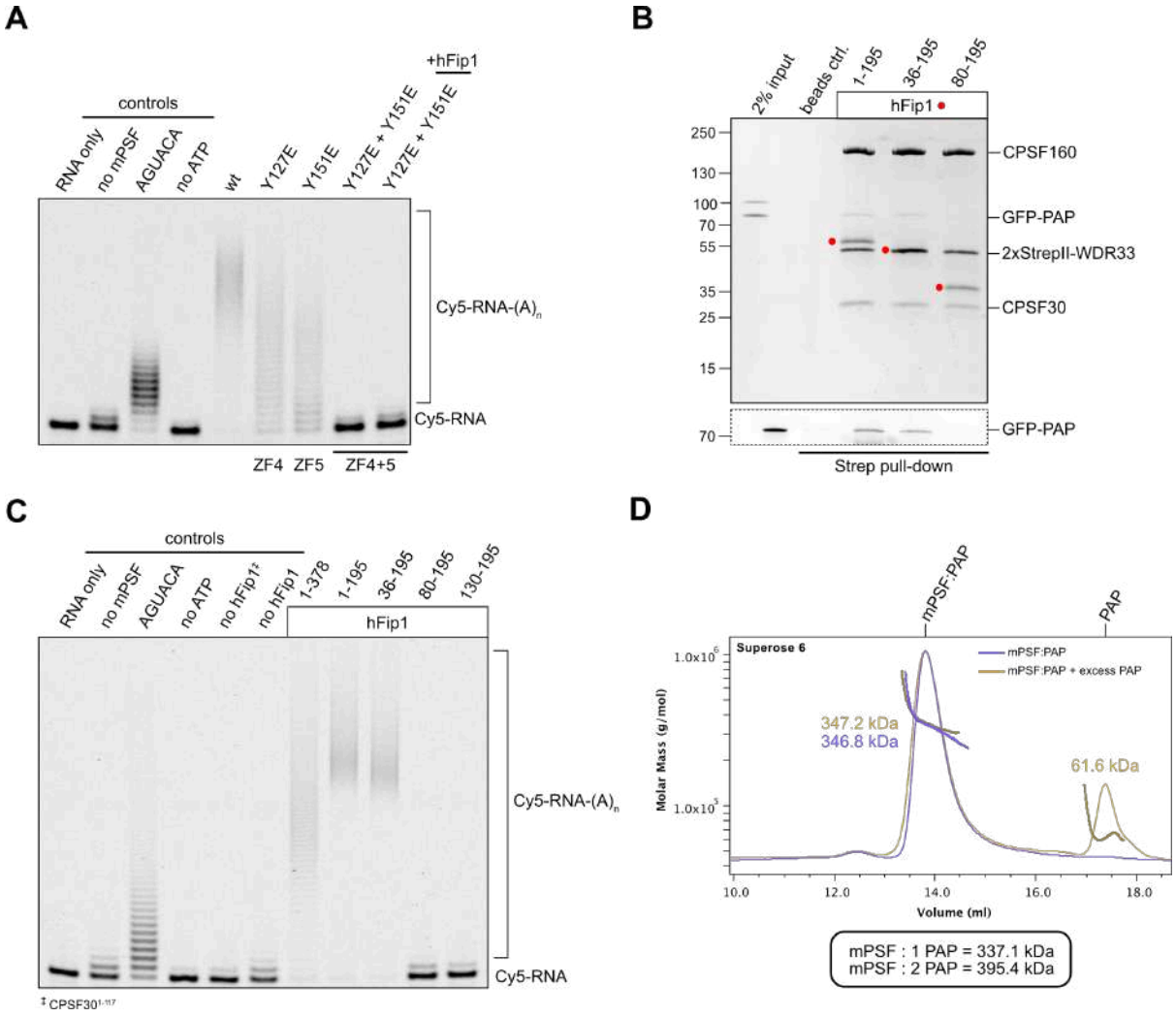


Figure 2: hFip1 directly recruits poly(A) polymerase. (**A**) Polyadenylation activity assay of mPSF complexes containing wild-type and mutant CPSF30 proteins using a Cy5-labelled PAS-containing RNA substrate. Polyadenylated RNA products are indicated as RNA-$(A)_n$. (**B**) Pull-down analysis of immobilized StrepII-tagged mPSF complexes comprising N-terminal truncations of hFip1 with GFP-PAP. GFP-PAP is visualized by in-gel GFP fluorescence (bottom). (**C**) Polyadenylation activity assay of mPSF complexes containing hFip1 truncations. (**D**) SEC-MALS analysis of reconstituted mPSF:PAP:RNA complexes and in the absence (purple) or presence of excess PAP (yellow). Theoretical molecular masses of 1:1 and 1:2 mPSF:PAP complexes are indicated.

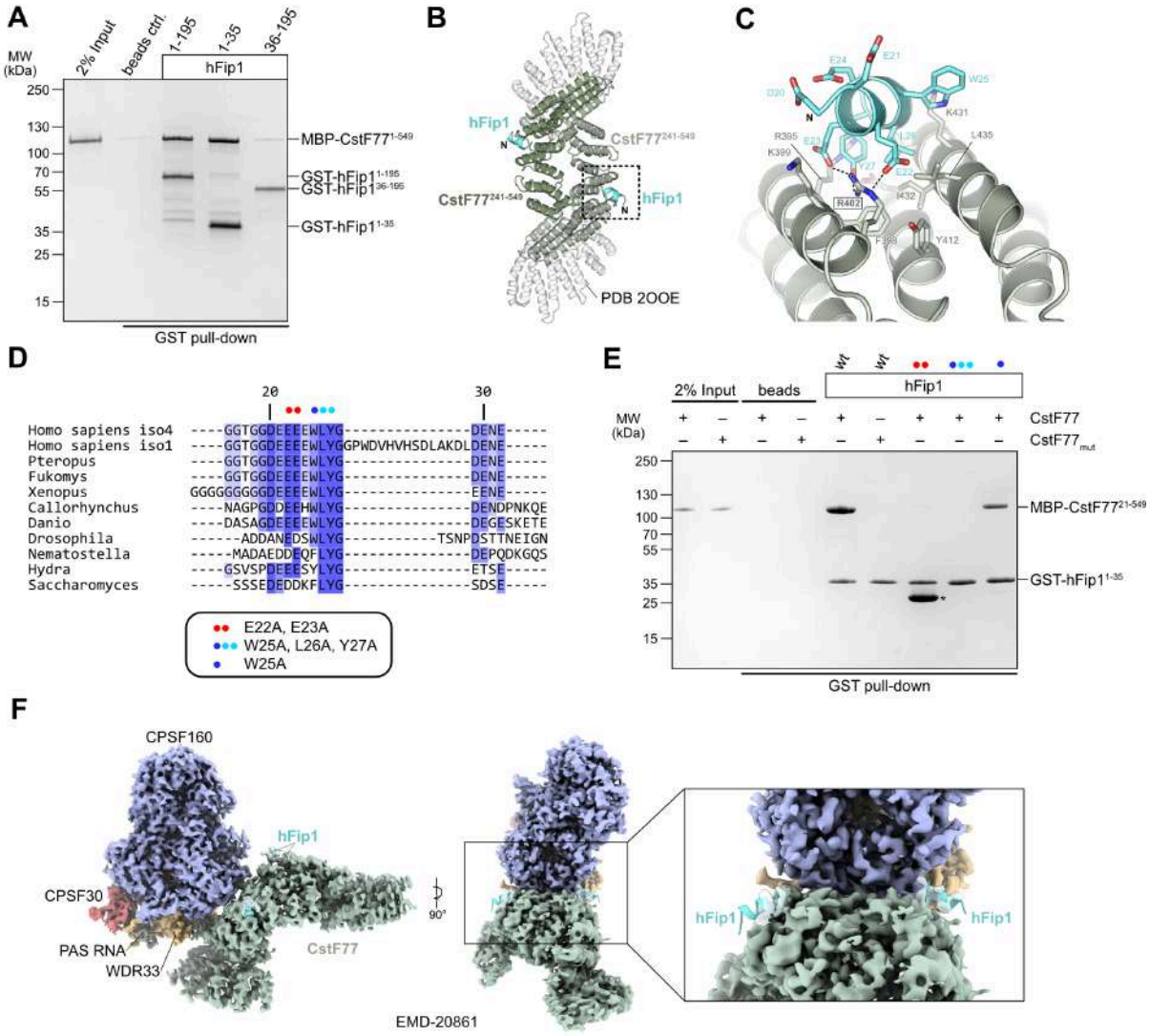


Figure 3: hFip1 interacts with CstF77 through a conserved motif within its N-terminal acidic domain. (**A**) Pull-down analysis of immobilized GST-hFip1 fragments with MBP-CstF77$^{21-549}$. (**B**) Cartoon representation of the crystal structure of the CstF77$^{241-549}$-hFip1$^{1-35}$ complex, superimposed onto the structure of murine CstF77 (white, PDB ID: 2OOE). (**C**) Zoomed-in view of the hFip1-CstF interaction interface. (**D**) Multiple sequence alignment of the N-terminal region of Fip1 orthologs. (**E**) Pull-down analysis of immobilized wild-type and mutant GST-hFip1$^{1-35}$ proteins with MBP-CstF77$^{21-549}$ and MBP-CstF77$_{mut}$ (R395A/R402A/K431A). Asterisk indicates contaminating free GST protein. (**F**) 3D cryo-EM density map (EMD-20861) of the human CPSF160-WDR33-CPSF30-PAS RNA-CstF77 complex (Zhang et al., 2019), displayed at contour level 0.015 and color-coded according to the corresponding atomic protein model (PDB ID 6URO). The hFip1-CstF77 crystal structure from this study was superimposed onto the atomic model of CstF77, and atomic model of hFip1 is shown (cyan). Inset shows a zoomed-in view of unassigned density that matches hFip1.

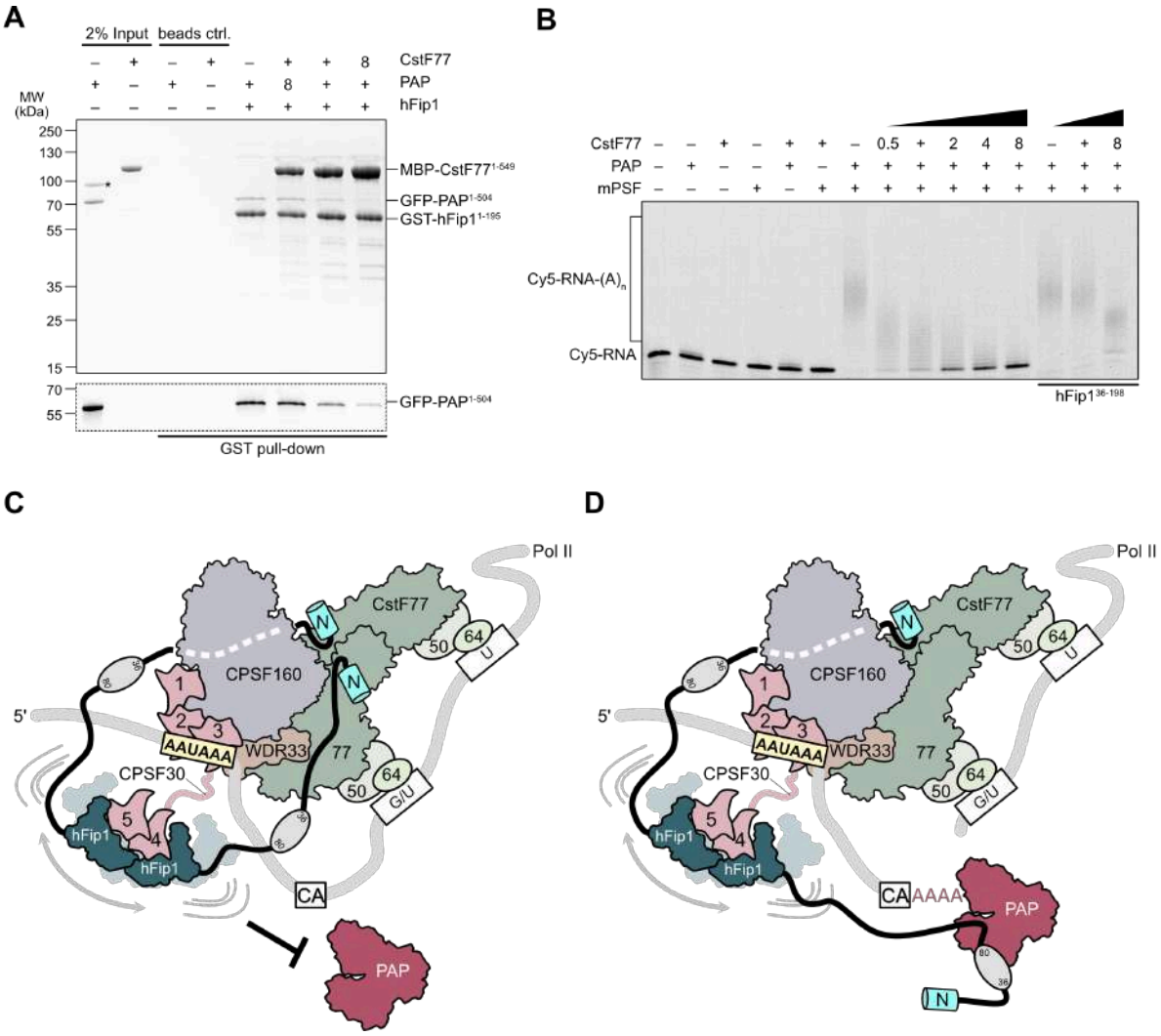
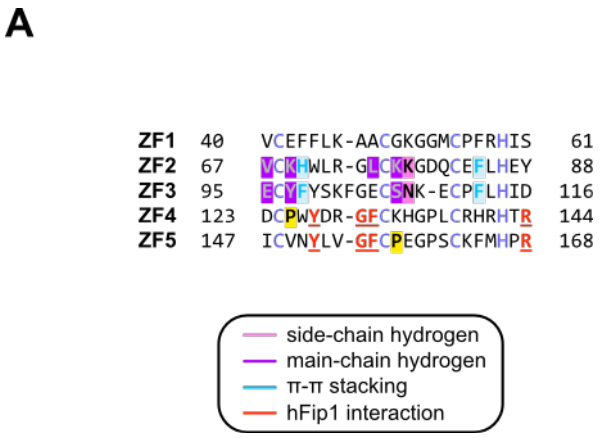


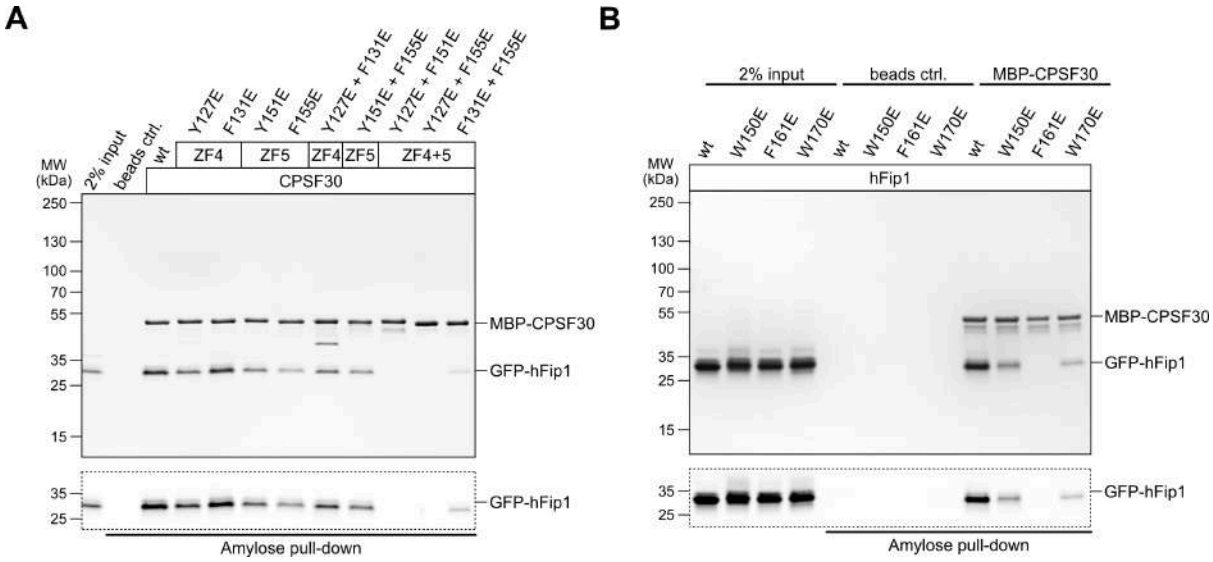
Figure 4: CstF77 competitively inhibits 3' polyadenylation. (**A**) Pull-down analysis of immobilized GST-hFip1$^{1-195}$ with varying molar ratios of GFP-PAP and MBP-CstF77$^{21-549}$. GFP-PAP is visualized by in-gel GFP fluorescence (bottom). Asterisk denotes contaminating protein. (**B**) Polyadenylation activity assay of mPSF complexes containing full-length hFip1 and N-terminally truncated hFip1 (hFip1$^{36-195}$) in the presence of varying molar ratios of CstF77. Polyadenylated RNA products are indicated as RNA-(A)$_n$. (**C**) Model of CPSF-mediated pre-mRNA cleavage and polyadenylation and CstF77-dependent inhibition of polyadenylation. The interactions between CstF77 and hFip1 competitively inhibit PAP recruitment (left). (**D**) Upon pre-mRNA cleavage, structural remodeling of the CPSF-CstF complex enables hFip1 to recruit PAP to the nascent 3' end of the mRNA and consequently stimulates polyadenylation.



828

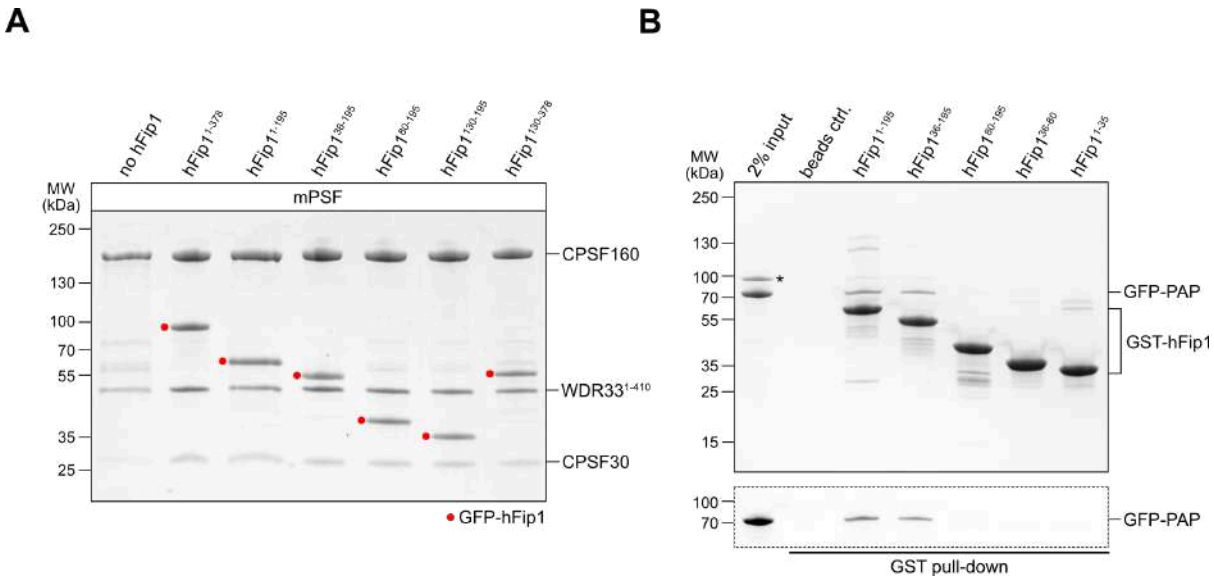
829 ***Figure 1 – figure supplement 1: Sequence alignment of CPSF30 zinc finger domains.*** (**A**) Sequence alignment
830 of human CPSF30 zinc finger domains. Residues responsible for RNA interactions (in ZF2/ZF3) or hFip1 interaction
831 (in ZF4/ZF5) are highlighted and the nature of their interaction color-coded. ZF4/ZF5 domains contain proline
832 residues (yellow) at positions corresponding to critical main-chain hydrogen bonding interactions in ZF2/ZF3.

833



834

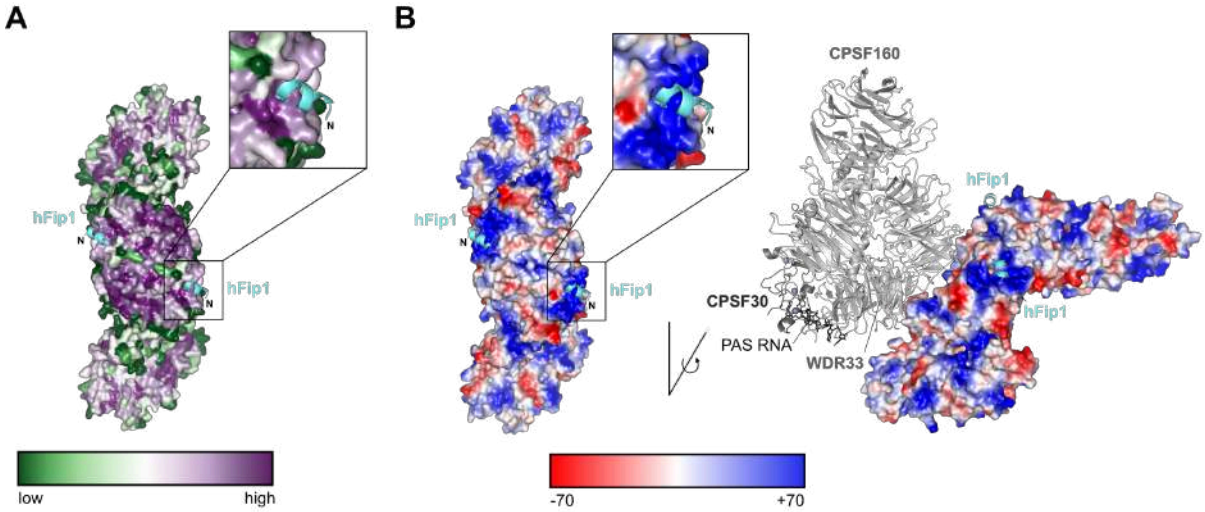
835 ***Figure 1 – figure supplement 2: Analysis of the hFip1-CPSF30 interaction using structure-guided point***
836 ***mutants.*** (**A**) Pull-down assay of immobilized MBP-tagged wild-type (wt) or mutant CPSF30 proteins with GFP-
837 hFip1. GFP-hFip1 is visualized by in-gel GFP fluorescence (bottom). Asterisk indicates contaminating free MBP
838 protein. (**B**) Pull-down assay of immobilized MBP-tagged wt CPSF30 and GFP-hFip1 mutants.



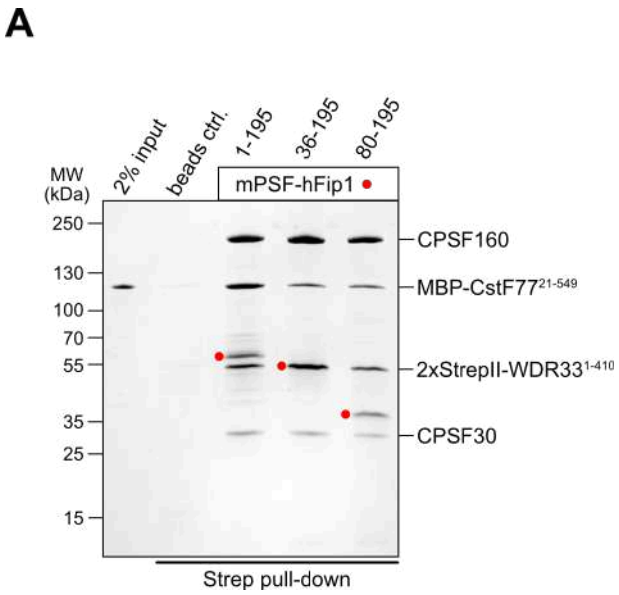
839

840 ***Figure 2 – figure supplement 1: Analysis of hFip1 regions required for PAP recruitment.*** (**A**) Purified mPSF
841 complexes containing hFip1 isoform 4 and truncations thereof (indicidated with red dots) used in polyadenylation
842 activity assay (Fig. 2C). (**B**) Pull-down assay of immobilized GST-tagged hFip1 fragments with GFP-PAP. GFP-
843 PAP is visualized by in-gel GFP fluorescence (bottom). Asterisk denotes protein contaminant.

844



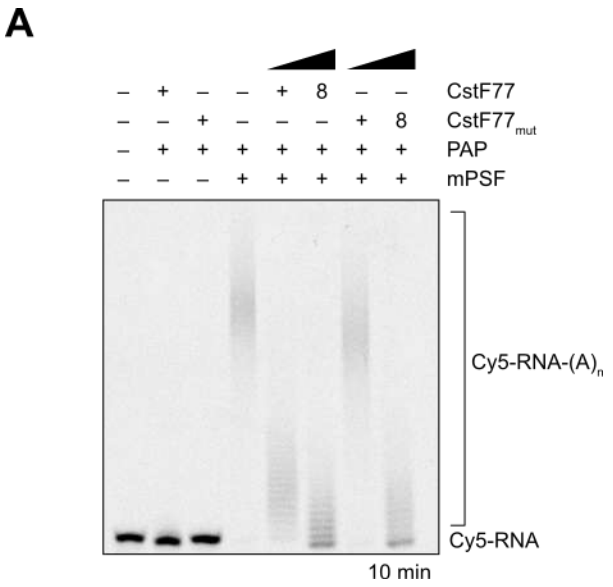
845
846 ***Figure 3 – figure supplement 1: hFip1 binds to a conserved positively-charged patch on CstF77.*** (**A**) Color-
847 coded representation of surface conservation of CstF77. Cartoon representation of bound hFip1$^{1-35}$ (cyan). Inset
848 shows a zoomed-in view of the sequence conservation of CstF77 surrounding at the Fip1 binding site. (**B**) Color-
849 coded electrostatic surface representation of CstF77 HAT homodimer alone (left) or bound to mPSF (right), both
850 with hFip1 (cyan) shown in cartoon representation. Inset shows a zoomed-in view of the CstF77 surface
851 electrostatics at the hFip1 binding site.



852

853 ***Figure 3 – figure supplement 2: The N-terminal region of hFip1 contributes to mPSF-CstF77 interaction.*** (**A**)
854 Pull-down analysis of immobilized StrepII-tagged mPSF complexes containing N-terminally truncated hFip1
855 proteins with MBP-CstF77 HAT domain.

856



857

858 ***Figure 4 – figure supplement 1: CstF77 reduces RNA 3' polyadenylation rate.*** (**A**) Polyadenylation activity
859 assay of mPSF in the presence of varying molar ratios of wild-type CsftF77 or a CstF77 mutant ($CstF_{mut}$) incapable
860 of binding hFip1. Polyadenylated RNA is indicated as RNA-$(A)_n$.

A

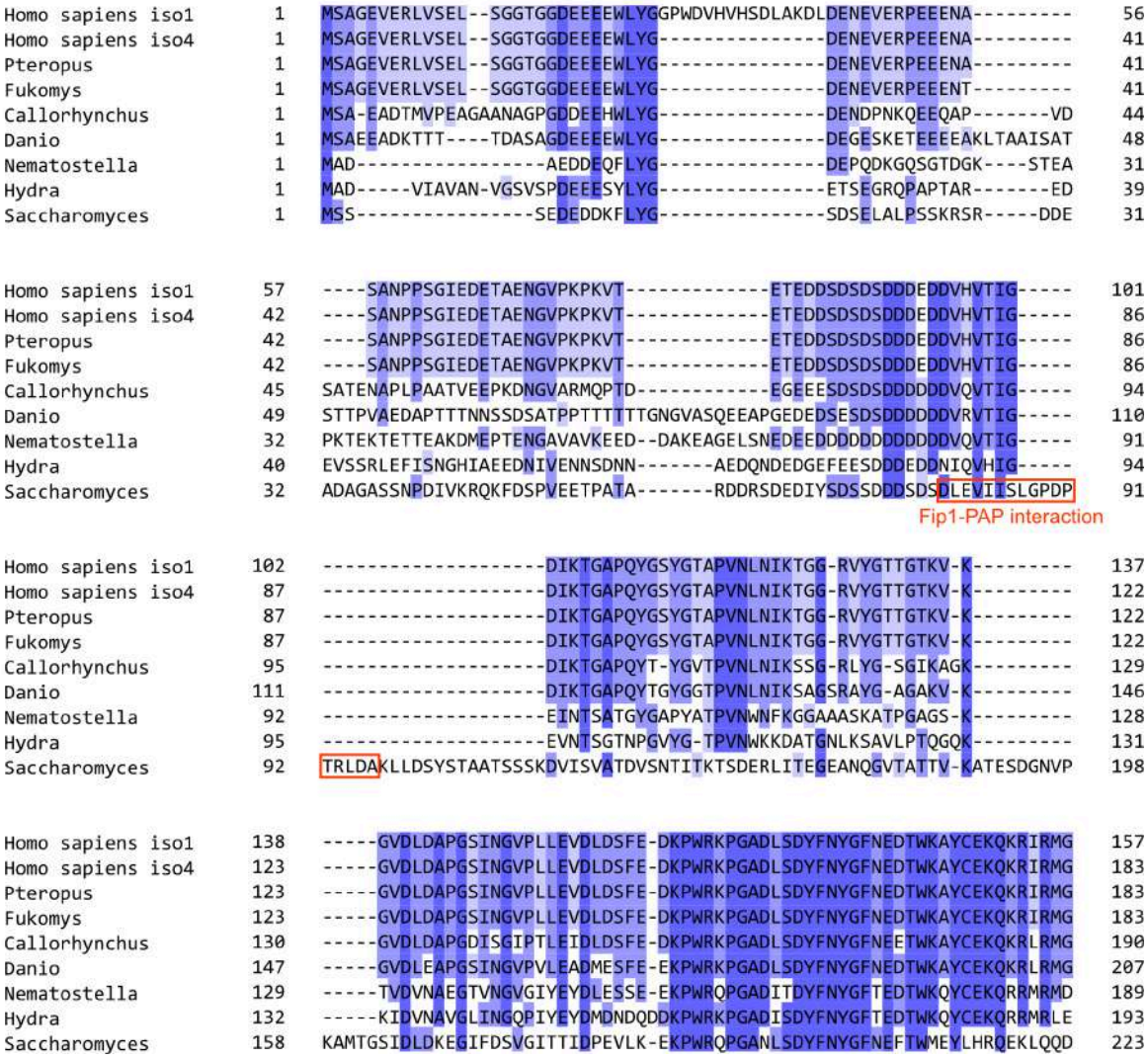


Figure 4 – figure supplement 2: Pap1 interaction motif in Fip1 orthologs is poorly conserved. (A) Multiple sequence alignment of hFip1 isoform 1, hFip1 isoform 4 and selected orthologs, colored by sequence similarity. Residues of yeast Fip1 that directly interact with Pap1 (Meinke et al., 2008, PDB ID: 3C66) are indicated in red.

TABLES

Table 1. Crystallographic data collection and refinement statistics.

	hFip1-CPSF30	hFip1-CstF77
Data collection		
Space group	$P2_1$	$P6_122$
Cell dimensions		
a, b, c (Å)	60.127, 115.125 , 66.444	157.612, 157.612, 161.005
α, β, γ (°)	90, 116.781, 90	90, 90, 120
Wavelength (Å)	1.2809	1.0000
Resolution (Å)	48.65-2.201 (2.28-2.201)	56.31-2.55 (2.641-2.55)
Total reflections	226720 (15294)	1577004 (162259)
Unique reflections	37698 (3244)	38981 (3836)
R_{merge} (%)	7.5 (95.9)	9.2 (186.1)
R_{pim} (%)	3.2 (46.9)	1.5 (28.8)
$I/\sigma I$	13.5 (1.1)	36.0 (2.6)
Cc(1/2)	0.998 (0.557)	1 (0.836)
Completeness (%)	92.3 (80.22)	99.96 (100.00)
Redundancy	6.0 (4.7)	40.5 (42.3)
Refinement		
Resolution (Å)	48.65-2.201	56.31-2.55
No. reflections	37698	38975
R_{work} / R_{free}	0.2406/0.2622	0.2410/0.2647
No. non-hydrogen atoms		
Protein	4607	5188
Ligand/ion	8	98
Water	67	25
B-factors (Å^2)		
Protein	56.53	65.34
Ligand/ion	63.69	69.46
Water	49.83	55.9
R.m.s. deviations		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	1.03	1.1
Ramachandran plot		
% favored	95.83	97.9
% allowed	4.17	2.1
% outliers	0	0

Values in parentheses are for highest-resolution shell.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY METHODS

Protein expression and purification

MacroLab Series 2 Biobricks adaption

In order to allow for biobricks-type assembly to generate polypromoter plasmids for *E. coli* expression using ligation-independent cloning analogous to MacroLab Series 438, MacroLab Series 2 vectors were modified by inserting whole expression cassettes as gBlocks (IDT) into 2B-T with Gibson assembly carrying the necessary modifications. The resulting vectors denoted 16-B (for His_6-TEV-tagged protein expression), 16-M (for His_6-MBP-TEV-tagged protein expression), and 16-M_ΔHis (for MBP-TEV-tagged protein expression), have two *PmeI* restriction sites flanking the T7 expression cassette, an internal *SspI* site for target gene insertion, and a *SwaI* site downstream of the T7 terminator for biobricks-type assembly using ligation-independent cloning. Gene assembly proceeds following the Series 438 vectors assembly protocol (Gradia et al., 2017).

MBP-CPSF30 proteins for pull-down analysis

His_6-MBP-TEV-$CPSF30^{1-243}$ mutants were expressed and purified as described for SEC-MALS. For CPSF30 point mutants Y151E, Y127E/Y151E, Y127E/Y155E, a high salt wash (25 mM Tris-HCl pH 7.5, 1 M NaCl) during Ni-IMAC purification was included prior to elution with buffer A buffer supplemented with 200 mM imidazole.

GFP-hFip1 proteins for pull-down analysis

His_6-GFP-TEV-$hFip1^{1-195}$ mutants (W150E, F161E, W170E) were expressed and purified following the same purification strategy as for the His_6-GFP-TEV-$hFip1^{1-195}$ wt protein.

GST-hFip1 proteins for pull-down analysis

His_6-GST-TEV-$hFip1^{36-80}$ was expressed in *E. coli* BL21 star (DE3) cells overnight at 18 °C by induction with 0.5 mM IPTG at OD_{600} of about 0.6-0.8. Cells were lysed by sonication in buffer A supplemented with 1 mM DTT, 1 µM Peptsatin A, and 400 µM AEBSF protease inhibitor. Lysate was cleared by centrifugation (20 min, 20,000 × *g*, 4°C). Clarified lysate was subjected to a GSTrap Fast Flow (Cytiva) column, washed with buffer A supplemented with 1 mM DTT prior to elution in buffer A supplemented with 1 mM DTT and 10 mM GSH. Protein was further purified on a Superdex 200 (Cytiva) column

893 equilibrated in buffer A supplemented with 1 mM DTT. Protein was concentrated in centrifugal filter
894 (Amicon Ultra-15, MWCO 30 kDa, Merck Millipore), aliquoted, flash frozen, and stored at -80 °C. His_6-
895 GST-TEV-hFip1$^{80-195}$ was expressed and purified using the same protocol, but changing to buffer B.

896 ***Pull-down assays***

897 For all pull-down assays, bound proteins were eluted with 1X SDS-PAGE loading buffer on ice and
898 analyzed by SDS-PAGE on 4-20% Mini-PROTEAN TGX Precast Protein Gels (Bio-Rad) without prior
899 heating to preserve the GFP fluorescence. GFP fluorescence was visualized on a Typhoon FLA 9500
900 laser scanner (Cytiva) at 473 nm and subsequently stained with Coomassie brilliant blue R250.

901 *Pull-down analysis of hFip1-CPSF30 interaction*

902 For pull-down analysis of hFip1 mutants binding to CPSF30, 120 µg of purified His_6-MBP-TEV-
903 CPSF30$^{1-243}$ wt protein was incubated with 120 µl Amylose resin (NEB) equilibrated in pull-down wash
904 buffer and gently agitated at 4 °C for 1 h. The beads were washed three times with 0.5 ml of pull-down
905 wash buffer and equally distributed in four tubes. His_6-GFP-TEV-hFip1$^{1-195}$ wt and point mutants
906 (W150E, F161E, W170E) were added in 5-fold molar excess to the beads. After incubation at 4 °C for
907 1 h, gently agitated, unbound protein was washed off by adding three times 0.5 ml pull-down wash
908 buffer. For pull-down analysis of the hFip1-interaction with the ZF of CPSF30, 15 µg of His_6-MBP-TEV-
909 CPSF30$^{1-243}$ wt and ZF mutants were incubated each with 30 µl Amylose resin (NEB) equilibrated in
910 pull-down wash buffer and gently agitated at 4 °C for 1 h. Unbound protein was washed off three times
911 with 0.5 ml of pull-down wash buffer and His_6-GFP-TEV-hFip1$^{1-195}$ wt was added in 4-fold molar excess
912 to the resin and incubated at 4 °C for 1 h, gently agitated. Beads were washed three times with 0.5 ml
913 pull-down wash buffer.

914 *Pull-down analysis of hFip1-PAP interaction*

915 For pull-down analysis of the hFip1:PAP interaction, 20 µg of His_6-GST-TEV-hFip1 truncation constructs
916 (hFip1$^{1-195}$, hFip1$^{36-195}$, hFip1$^{80-195}$, hFip1$^{36-80}$, and hFip1$^{1-35}$) were incubated each with 15 µl Glutathione
917 Sepharose 4 Fast Flow beads (Cytiva) equilibrated in pull-down wash buffer and gently agitated at 4 °C
918 for 1 h. Unbound protein was washed off three times with 0.5 ml of pull-down wash buffer and His_6-
919 GFP-TEV-PAP$^{1-504}$ was added in 4-fold molar excess to the resin and incubated at 4 °C for 1 h, gently
920 agitated. Beads were washed three times with 0.5 ml pull-down wash buffer.

921

Pull-down analysis of mPSF-CstF77 interaction

For pull-down analysis of the mPSF:CstF77 interaction, Ni-IMAC purified mPSF complexes from *Sf9* cells containing hFip1$^{1-195}$ and N-terminal truncations hereof (hFip1$^{36-195}$, hFip1$^{80-195}$) were incubated with 20 µl Strep-Tactin (IBA Lifesciences) beads in buffer containing 20 mM HEPES-KOH pH 8.0, 150 mM KCl, 0.05% Tween-20, 0.5 mM TCEP and gently agitated at 4 °C for 1 h. Unbound protein was washed off three times with 0.5 ml of buffer containing 20 mM HEPES-KOH pH 8.0, 150 mM KCl, 0.05% Tween-20, 0.5 mM TCEP and 10 µg His$_6$-MBP-TEV-CstF77$^{21-549}$ was added to the resin and incubated at 4 °C for 1 h, gently agitated. Beads were again washed three times with 0.5 ml.

Polyadenylation assay with titration of CstF77

Polyadenylation reactions were performed according to standard procedure described above. Reaction mix for polyadenylation assay with varying CstF77 concentrations (Figure 4 – figure supplement 1) contained 80 nM of either mPSF complex (CPSF160-WDR33$^{1-410}$-CPSF30-hFip1$^{1-198}$ or CPSF160-WDR33$^{1-410}$-CPSF30-hFip1$^{36-198}$), 160 nM PAP$^{1-504}$, and a final concentration of 500 µM ATP. His$_6$-TEV-CstF77$^{21-549}$ or His$_6$-TEV-CstF77$_{mut}^{21-549}$ were added at 160 nM (denoted as + or 1) or 1280 nM (denoted 8).

Bioinformatic analysis of CstF77

Color-coded electrostatic surface representation of CstF77 was generated for the biological assembly of murine CstF77$^{20-549}$ (PDB ID: 2OOE) in PyMOL 2.5.0 (Schrödinger LLC, 2021) using the protein contact potential option. The CstF77$^{241-549}$-hFip1$^{1-35}$ structure from this study and cryo-EM structure of human CPSF160-WDR33-CPSF30-PAS RNA-CstF77 complex (PDB ID: 6URO) were superimposed onto murine CstF77 using PyMOL's align command to identify and visualize the hFip1- and mPSF-binding regions, respectively. Analysis of evolutionary conservation of CstF77 was carried out using the ConSurf web server (Ashkenazy et al., 2016) with murine CstF77$^{20-549}$ (PDB ID: 2OOE) as input and applying standard settings (sequence alignment with MAFFT, homologues taken from UniRef90). The degree of conservation was visualized in PyMOL by color-coding (green: variable, violet: conserved) the protein surface according to the conservation scores which are written into the tempFactor column of the ConSurf web server output PDB file.

SUPPLEMENTARY TABLES

Supplementary Table 1. Expression constructs.

#	Protein	Vector	Figure	Expression
pLM B092	His_6-MBP-TEV-HsCstF77$^{21-549}$	1M	Fig. 3A, Fig. 3E, Fig. 4A, Fig. 3-Supplement 2A	*E. coli*
pLM B123	His_6-TEV-CstF77$^{21-559}$	16-B	Fig. 4B Fig. 4-Supplement 1A	*E. coli*
pLM B142	His_6-GST-TEV-MmPAP$^{1-504}$	2GT	Fig. 2D	*E. coli*
pLM B156	His_6-MBP-TEV-HsPAP$^{1-504}$	1M	Fig. 4B Fig. 4-Supplement 1A	*E. coli*
pLM B157	His_6-GFP-TEV-HsPAP$^{1-504}$	2GFP-T	Fig. 1G Fig. 2B Fig. 4A Fig. 2-Supplement 1B	*E. coli*
pLM B164	His_6-MBP-TEV-CstF77$^{241-549}$	1M	Fig. 3B	*E. coli*
pLM B167	His_6-MBP-TEV-CstF77$^{21-549}$ R395A/R402A/K431A	1M	Fig. 3E	*E. coli*
pLM B168	His_6-MBP-TEV-CstF77$^{21-549}$ R395A/R402A/K431A	1M	Fig. 3E	*E. coli*
pLM B170	His_6-TEV-CstF77$^{21-549}$ R395A/R402A/K431A	16B	Fig. 4-Supplement 1A	*E. coli*
pMC B051	His_6-MBP-TEV-CPSF30$^{118-178}$	1B	Fig. 1B-E	*E. coli*
pMC B054	His_6-MBP-TEV-CPSF30 wt	16-M	Fig. 1F, Fig. 1-Supplement 2A Fig. 1-Supplement 2B	*E. coli*
pMC B055	His_6-MBP-TEV-CPSF30 F131E	16-M	Fig. 1-Supplement 2A	*E. coli*
pMC B056	His_6-MBP-TEV-CPSF30 F155E	16-M	Fig. 1F, Fig. 1-Supplement 2A	*E. coli*
pMC B057	His_6-MBP-TEV-CPSF30 F131E/F155E	16-M	Fig. 1-Supplement 2A	*E. coli*

pMC B058	His_6-MBP-TEV-CPSF30 Y127E	16-M	Fig. 1F, Fig. 1-Supplement 2A	*E. coli*
pMC B059	His_6-MBP-TEV-CPSF30 Y127E/F131E	16-M	Fig. 1-Supplement 2A	*E. coli*
pMC B060	His_6-MBP-TEV-CPSF30 Y151E	16-M	Fig. 1-Supplement 2A	*E. coli*
pMC B061	His_6-MBP-TEV-CPSF30 Y151E/F155E	16-M	Fig. 1-Supplement 2A	*E. coli*
pMC B062	His_6-MBP-TEV-CPSF30 Y127E/Y151E	16-M	Fig. 1-Supplement 2A	*E. coli*
pMC B063	His_6-MBP-TEV-CPSF30 Y127E/F155E	16-M	Fig. 1F, Fig. 1-Supplement 2A	*E. coli*
pMC C011	His_6-GST-TEV-hFip1$^{130-195}$	13S-A	Fig. 1B-E	*E. coli*
pMC C015	His_6-GST-TEV-hFip1$^{80-195}$	2GT	Fig. 2-Supplement 1B	*E. coli*
pMC C030	His_6-TEV-hFip1$^{130-195}$	16-B		*E. coli*
pMC C049	His_6-TEV-GFP- hFip1$^{130-195}$	16-B	Fig. 1F, Fig. 1-Supplement 2A, Fig. 1-Supplement 2B	*E. coli*
pMC C050	His_6-GST-TEV-hFip1$^{36-80}$	2GT	Fig. 2-Supplement 1B	*E. coli*
pMC C059	His_6-GST-TEV-hFip1$^{1-35}$	2GT	Fig. 3A, Fig. 3B, Fig. 3E	*E. coli*
pMC C060	His_6-GST-TEV-hFip1$^{1-195}$	2GT	Fig. 2A, Fig. 3A	*E. coli*
pMC C066	His_6-TEV-GFP-TEV-hFip1 W150E	16-B‡	Fig. 1-Supplement 2B	*E. coli*
pMC C067	His_6-TEV-GFP-TEV-hFip1 F161E	16-B‡	Fig. 1-Supplement 2B	*E. coli*
pMC C068	His_6-TEV-GFP-TEV-hFip1 W170E	16-B‡	Fig. 1-Supplement 2B	*E. coli*
pMC C073	His_6-GST-TEV-hFip1$^{36-195}$	2GT	Fig. 3A	*E. coli*
pMC C093	His_6-GST-TEV-hFip1$^{1-35}$ E22A + E23A	2GT	Fig. 3E	*E. coli*
pMC C094	His_6-GST-TEV-hFip1$^{1-35}$	2GT	Fig. 3E	*E. coli*

	W25A + L26A + Y27A			
pMC C096	His_6-GST-TEV-hFip1$^{1-35}$ W25A	2GT	Fig. 3E	*E. coli*
pMC N015	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2C	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-117}$	438-A		
pMC N018	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2C	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-178}$	438-A		
	StrepII-GFP-TEV-hFip1$^{130-195}$	438-RGFP		
pMC N018A	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2B, Fig. 2C, Fig. 2-Supplement 1A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
pMC N018C-2	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2C, Fig. 2-Supplement 1A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{1-378}$	438-RGFP		
pMC N018G	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2A, Fig. 2B, Fig. 2C, Fig. 2D, Fig. 4B Fig. 2-Supplement 1A, Fig. 3-Supplement 2A, Fig. 4-Supplement 1A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-0	His_6-TEV-CPSF160$^{1-1443}$	438-B		*Sf9*

	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-8	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ F155E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-10	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y127E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-12	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y151E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-14	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y127E/Y151E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-15	His_6-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2A	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		

	CPSF30$^{1-243}$ Y127E + F155E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y127E/F155E	438-A		
	GFP-TEV-hFip1$^{1-195}$	DNA fragment*		
pMC N018G-21	His_6-TEV-FLAG-CPSF160$^{1-1443}$	438-B**	Fig. 1G	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-22	His_6-TEV-FLAG-CPSF160$^{1-1443}$	438-B**	Fig. 1G	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y127E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-23	His_6-TEV-FLAG-CPSF160$^{1-1443}$	438-B**	Fig. 1G	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y151E	438-A		
	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018G-24	His_6-TEV-FLAG-CPSF160$^{1-1443}$	438-B**	Fig. 1G	*Sf9*
	His_6-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$ Y127E/Y151E	438-A		

	StrepII-GFP-TEV-hFip1$^{1-195}$	438-RGFP		
pMC N018H	His$_6$-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2C, Fig. 2-Supplement 1A	*Sf9*
	His$_6$-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{130-378}$	438-RGFP		
pMC N018I	His$_6$-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2C, Fig. 2-Supplement 1A	*Sf9*
	His$_6$-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{130-195}$	438-RGFP		
pMC N018J	His$_6$-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2B, Fig. 2C, Fig. 4B, Fig. 2-Supplement 1A, Fig. 3-Supplement 2A	*Sf9*
	His$_6$-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{36-195}$	438-RGFP		
pMC N018K	His$_6$-TEV-CPSF160$^{1-1443}$	438-B	Fig. 2B, Fig. 2C, Fig. 2-Supplement 1A Fig. 3-Supplement 2A	*Sf9*
	His$_6$-TEV-2xStrepII-TEV-WDR33$^{1-410}$	438-D		
	CPSF30$^{1-243}$	438-A		
	StrepII-GFP-TEV-hFip1$^{80-195}$	438-RGFP		

‡GeneArt Strings: Insert with LIC overhangs ordered as DNA fragment that already contains GFP-hFip1

*GeneArt Strings: PmeI-digested insert ordered as DNA fragment and directly subcloned

**FLAG-tag inserted by FLAG-overhang in Oligo for PCR.

954 ***Supplementary Table 2.*** RNA oligos.

#	Accession No.	Database	Source
rLM 011	CUGC**AAUAAA**CAACUUAACAACAAAAA	unlabeled	IDT
rLM 015	CUGC**AAUAAA**CAACUUAACGUCAAAAA	5'-Cy5	IDT
rLM 016	CUGC**AGUACA**CAACUUAACGUCAAAAA	5'-Cy5	IDT

955

956 ***Supplementary Table 3.*** hFip1 ortholog input sequences for multiple sequence alignment.

Organism	Accession No.	Database
Homo sapiens (isoform 1)	Q6UN15	UniProtKB
Homo sapiens (isoform 4)	Q6UN15-4	UniProtKB
Pteropus vampyrus	XP_011360945.1	NCBI Reference Sequence
Fukomys damarensis	XP_010638552.1	NCBI Reference Sequence
Xenopus tropicalis	NP_001037890.1	NCBI Reference Sequence
Callorhinchus milii	XP_007890936.1	NCBI Reference Sequence
Danio rerio	NP_001006042.1	NCBI Reference Sequence
Drosophila melanogaster	NP_649476.1	NCBI Reference Sequence
Nematostella vectensis	XP_032230323.1	NCBI Reference Sequence
Hydra vulgaris	XP_004208725.1	NCBI Reference Sequence
Saccharomyces cerevisiae	NP_012626.1	NCBI Reference Sequence