Molecular mechanism for the interaction between human CPSF30 and hFip1

Keith Hamilton and Liang Tong

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

Most eukaryotic pre-mRNAs must undergo 3′-end cleavage and polyadenylation prior to their export from the nucleus. A large number of proteins in several complexes participate in this 3′-end processing, including cleavage and polyadenylation specificity factor (CPSF) in mammals. The CPSF30 subunit contains five CCCH zinc fingers (ZFs), with ZF2–ZF3 being required for the recognition of the AAUAAA poly(A) signal. ZF4–ZF5 recruits the hFip1 subunit of CPSF, although the details of this interaction have not been characterized. Here we report the crystal structure of human CPSF30 ZF4–ZF5 in complex with residues 161–200 of hFip1 at 1.9 Å resolution, illuminating the molecular basis for their interaction. Unexpectedly, the structure reveals one hFip1 molecule binding to each ZF4 and ZF5, with a conserved mode of interaction. Our mutagenesis studies confirm that the CPSF30–hFip1 complex has 1:2 stoichiometry in vitro. Mutation of each binding site in CPSF30 still allows one copy of hFip1 to bind, while mutation of both sites abrogates binding. Our fluorescence polarization binding assays show that ZF4 has higher affinity for hFip1, with a K_d of 1.8 nM. We also demonstrate that two copies of the catalytic module of poly(A) polymerase (PAP) are recruited by the CPSF30–hFip1 complex in vitro, and both hFip1 binding sites in CPSF30 can support polyadenylation.

In eukaryotes, most mRNA precursors (pre-mRNAs) are cleaved and polyadenylated at their 3′ end prior to their export from the nucleus (Zhao et al. 1999; Shi and Manley 2015; Sun et al. 2020a). This sequence of events is carefully orchestrated, with multiple protein complexes binding to different *cis* elements in the pre-mRNA, providing both a tight regulation of cleavage/polyadenylation and the opportunity to select from multiple cleavage sites with various affinities (alternative polyadenylation [APA]) (Tian and Manley 2017; Gruber and Zavolan 2019). Switching between these sites can lead to changes in the length and sequence of the 3′ UTR of mRNAs, which has many effects in protein expression, mRNA stability, and localization (Mayr 2017).

In mammals, a large number of proteins in several complexes contribute to the selection of cleavage and polyadenylation sites. These include cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), mammalian cleavage factor I (CFIm), CFIIm, and poly(A) polymerase (PAP), which adds the poly(A) tail. CPSF consists of two subcomplexes: mammalian polyadenylation specificity factor (mPSF) that recognizes the conserved polyadenylation signal (AAUAAA) upstream of the cleavage site and recruits PAP, and mammalian cleavage factor (mCF) that carries out the cleavage reaction (Chan

et al. 2014; Schönemann et al. 2014). mPSF contains four proteins: CPSF160, WDR33, CPSF30, and hFip1, and structures of mPSF bound to AAUAAA RNA demonstrate that CPSF160 serves as a scaffold onto which the N-terminal regions of WDR33 and CPSF30 are organized to bind the RNA (Clerici et al. 2018; Sun et al. 2018). The C-terminal region of CPSF30 and hFip1 are not visible in these structures, presumably because they are not in a fixed conformation relative to the mPSF core. mCF contains CPSF73, CPSF100, and symplekin, and its structure is highly dynamic (Zhang et al. 2020) although it becomes ordered in the active state (Sun et al. 2020b).

CPSF30 has five CCCH zinc fingers (ZF1–ZF5) near the N terminus and a CCHC zinc knuckle near the C terminus, separated by a nonconserved and putatively disordered region (Fig. 1A). ZF1 is crucial for the interaction with CPSF160, and ZF2–ZF3 recognize the AAUAAA poly(A) signal (Clerici et al. 2018; Sun et al. 2018). ZF5 of the yeast CPSF30 homolog Yth1 is required for binding to Fip1 (Barabino et al. 1997, 2000), which in turn tethers PAP (Meinke et al. 2008). hFip1 may also participate in cleavage site selection by binding to U-rich regions in the pre-mRNA through its C-terminal arginine-rich region (Kaufmann et al. 2004), and by bridging the CFIm
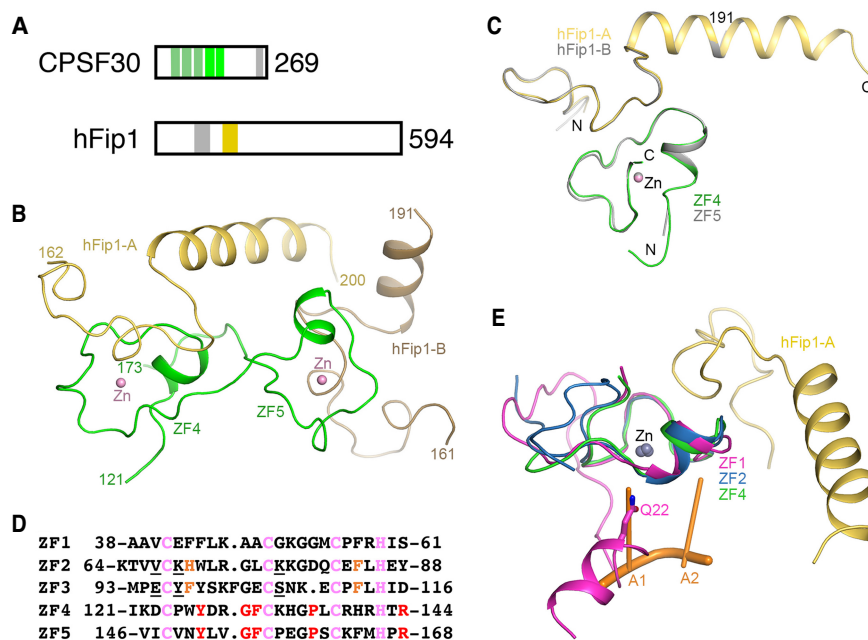
Figure 1. Structure of the human CPSF30–hFip1 complex. (*A*) Domain organizations of human CPSF30 and hFip1. The five zinc fingers of CPSF30 are shown in green, with ZF4 and ZF5 in a brighter color. The zinc knuckle of CPSF30 is shown in gray. The segments of hFip1 that interact with CPSF30 and PAP are shown in yellow and gray, respectively. (*B*) Schematic drawing of the structure of the human CPSF30–hFip1 complex. CPSF30 is in green. The hFip1 molecule bound to ZF4 is in yellow, and that bound to ZF5 is in brown. (*C*) Overlay of the ZF4–hFip1 complex (in color) and the ZF5–hFip1 complex (gray). Residue 191 at the C-terminal end of the hFip1 bound to ZF5 is labeled. (*D*) Sequence alignment of the CPSF30 zinc fingers. The ligands of the zinc ions are in pink, and residues in ZF4 and ZF5 that contribute to hFip1 binding are in red. Residues in ZF2 and ZF3 that contact the A-A dinucleotide are in orange, and residues whose main chain hydrogen bonds to the dinucleotide are underlined. (*E*) Overlay of the zinc fingers of CPSF30 and their binding partners. ZF1 is in magenta, ZF2 in blue and ZF4 in green. The A-A dinucleotide bound to ZF2 is shown in orange. ZF3 and ZF5 are not shown for clarity, as their structures are similar to that of ZF2 and ZF4, respectively. Structure figures are produced with PyMOL (www.pymol.org).

and CPSF complexes by interacting with the RS domain of CFIm68 or CFIm59 through its own RS domain (Zhu et al. 2018). These two interactions contribute to the selection of polyadenylation sites, making hFip1 an important regulator of APA.

It has previously been observed that CPSF30 interacts with residues 137–243 of hFip1 (Kaufmann et al. 2004), although the details of this interaction have not been characterized. Here we report the crystal structure of human CPSF30 ZF4–ZF5 in complex with residues 161–200 of hFip1 at 1.9 Å resolution, illuminating the molecular basis for their interaction. Unexpectedly, the structure reveals one hFip1 molecule binding to each ZF4 and ZF5, with a conserved mode of interaction. Our fluorescence polarization binding assays show that ZF4 has higher affinity for hFip1, with a K_d of 1.8 nM. We also demonstrate that two copies of the catalytic module of PAP (Bard et al. 2000; Martin et al. 2000) are recruited by the CPSF30–hFip1 complex in vitro.

Results

Overall structure of CPSF30–hFip1 complex

To obtain a more precise definition of the region of hFip1 that interacts with human CPSF30, we coexpressed ZF2–ZF5 of CPSF30 (residues 62–173) together with progressively shorter versions of hFip1 starting from residues 137–243 (Kaufmann et al. 2004), and assessed complex formation by gel filtration chromatography. The shortest version of hFip1 that could interact with CPSF30 contained residues 159–200, which we then used for crystallization.

Small crystals were observed from the initial crystallization screening using a sample containing ZF2–ZF5 of human CPSF30 (residues 62–173) and hFip1 (residues 159–200). However, these crystals took 2 mo to appear and could not be reproduced. We noticed that there were molds growing in the crystallization solution, and our earlier observations suggest that a protease secreted by the molds may have cleaved the protein(s), which was required for crystallization (Mandel et al. 2006; Bai et al. 2007b). We then introduced trypsin into the protein solution, which greatly improved the crystallization and produced crystals within a few days. We screened through various constructs for CPSF30 and hFip1, and the best crystals were obtained using a sample containing ZF4–ZF5 of CPSF30 (residues 114–173) coexpressed with hFip1 (residues 159–200), with trypsin at 1:280 weight ratio. The structure was determined at 1.9 Å resolution (Table 1) using the anomalous signal from the zinc atoms.

The electron density for one of the two hFip1 molecules stopped abruptly at residue Lys191. This residue is

Table 1. *Summary of crystallographic information*

Structure	Human CPSF30–hFip1 complex
Data collection	
Space group	$P6_4$
Cell dimensions	
a, b, c (Å)	79.0, 79.0, 48.7
α, β, γ (°)	90, 90, 120
Resolution (Å)[a]	48.7–1.9 (1.94–1.90)
R_{merge} (%)	4.6 (87.4)
$CC_{1/2}$	1.000 (0.788)
I/σI	30.9 (3.1)
Completeness (%)	100.0 (99.9)
No. of reflections	13793 (869)
Redundancy	12.6 (10.9)
Refinement	
Resolution (Å)	39.6–1.9 (1.95–1.90)
R_{work} (%)	17.9 (26.2)
R_{free} (%)	20.9 (24.9)
Number of atoms	1127
Protein	1043
Ligand/Ion	2
Water	82
B-factors (Å²)	
Protein	55.4
Ligand	41.4
Water	54.0
RMSD	
Bond lengths (Å)	0.017
Bond angles (°)	2.0
Ramachandran plot	
Favored (%)	96.6
Allowed (%)	100.0
Outliers (%)	0.00

[a]The numbers in parentheses are for the highest-resolution shell.

involved in crystal packing, and additional C-terminal residues would not be compatible with the crystal packing. In addition, the first residue observed for CPSF30 is Ile121, while residue 120 is Lys. Ile121 is also involved in crystal packing, and additional N-terminal residues here would not be compatible with the crystal packing either. In fact, Ile121 is located near Lys191 of the truncated hFip1 molecule in another asymmetric unit of the crystal. Therefore, in situ proteolysis by trypsin (or a fungal protease) of both CPSF30 and hFip1 was essential for this crystallization.

Both ZF4 and ZF5 of CPSF30 are well ordered in the structure, and the C-terminal extension beyond ZF5 (residues 169–173) is positioned between the two zinc fingers, helping to stabilize the structure (Fig. 1B). To our surprise, we observed two molecules of hFip1 in complex with each CPSF30 molecule in the structure, one primarily bound to ZF4 (hFip1-A) and the other to ZF5 (hFip1-B). Residues 162–200 are observed for hFip1-A, while residues 161–191 are observed for hFip1-B. This segment of hFip1 contains a loop (residues 161–181) followed by a long helix (residues 182–198). Residues at the N terminus (161–170) of this segment have weaker electron density as they do not have many contacts with CPSF30.

The conformations of the two zinc fingers are highly similar to each other (Fig. 1C), with an RMS distance of 0.54 Å for 23 equivalent Cα atoms between them, consistent with the high sequence conservation between them (Fig. 1D). Moreover, the binding modes of the two hFip1 molecules are similar to each other as well (Fig. 1C). An overlay of the ZF4–hFip1-A and ZF5–hFip1-B structures produces an RMS distance of 0.52 Å for 53 equivalent Cα atoms. The C-terminal helix of the two hFip1 molecules are located in the same position, even though that of hFip1-B is shorter because of the proteolysis (Fig. 1C). The similarity in the two complexes also indicate that crystal packing, especially for residue Lys191 in hFip1-B, has essentially no impact on the interactions between CPSF30 and hFip1.

The structures of the five zinc fingers of CPSF30 are similar to each other in general (Fig. 1E). However, they use different surfaces for interacting with their binding partners. ZF2 and ZF3 use one face of the zinc finger to bind A-A dinucleotides with the same binding mode (Sun et al. 2018) and to bind the NS1 protein of influenza virus (Das et al. 2008). ZF1 uses the same face to bind the N-terminal extension of CPSF30 (Clerici et al. 2018; Sun et al. 2018) with the side chain of Gln22 being located at the same position as the first base of the A-A dinucleotide. In contrast, ZF4 and ZF5 use the opposite face of the zinc finger to bind hFip1. Residues Pro125 in ZF4 and Pro157 in ZF5 would abolish the hydrogen-bonding interactions observed between the main chain of the equivalent residues in ZF2 and ZF3 and A-A dinucleotide (Fig. 1D). In fact, the Pro residues would also cause steric clash with the base of the nucleotides. Therefore, ZF4 and ZF5 are unlikely to bind RNA with high affinity.

Binding mode of hFip1 in CPSF30

Residues 173–188 of hFip1, in the loop prior to the helix and the first two turns of the helix, have extensive interactions with the zinc finger of CPSF30. For the interface between ZF4 and hFip1-A, 750 Å² of the surface area of each protein is buried here. The His142 ligand to the zinc ion is hydrogen-bonded to the main chain carbonyl of Asp174 (in hFip1-A), while the side chain of Asp174 has ionic interactions with that of Arg144 (ZF4) (Fig. 2A). The side chain hydroxyl of Tyr127 (ZF4) is hydrogen-bonded to the main chain carbonyl of Ser173 (hFip1-A), and its phenyl ring is sandwiched between the side chains of His142 (ZF4) and Asn177 (hFip1-A). The side chain of Phe131 (ZF4) is part of a cluster of aromatic residues from hFip1-A. This residue is conserved in ZF5, but not the other three zinc fingers of CPSF30 (Fig. 1D). Generally, residues in the CPSF30–hFip1 interface are highly conserved among their homologs (Fig. 2B,C).

For the interface between ZF5 and hFip1-B, 480 Å² of the surface area of each protein is buried here. Residues His142, Arg144, Tyr127, and Phe131 of ZF4 are conserved as His166, Arg168, Tyr151, and Phe155 in ZF5, and these residues maintain essentially the same interactions with hFip1-B (Fig. 2D). The main chain of Arg168 is placed against Arg129 in ZF4 and assumes a different
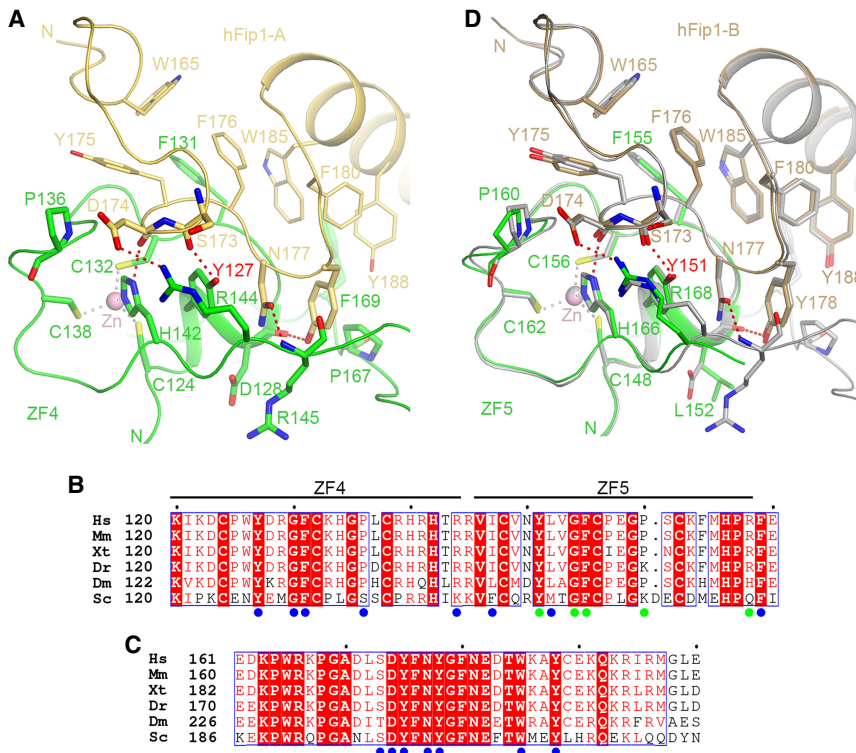
Figure 2. Detailed interactions between human CPSF30 and hFip1. (*A*) Interactions between ZF4 (green) and hFip1 (yellow). Side chains in the interface are shown as stick models. Hydrogen-bonding interactions are shown as dashed lines (red). Residue Tyr127 is labeled in red. (*B*) Sequence alignment of ZF4–ZF5 of CPSF30 homologs. Residues contributing >50 Å² buried surface area in the complex with hFip1 in ZF4 are highlighted in blue, and those in the complex with hFip1 in ZF5 are in green. (Hs) *Homo sapiens*, (Mm) *Mus musculus*, (Xt) *Xenopus tropicalis*, (Dr) *Danio rerio*, (Dm) *Drosophila melanogaster*, (Sc) *Saccharomyces cerevisiae*. (*C*) Sequence alignment of hFip1 homologs in the segment that interact with CPSF30. (*D*) Interactions between ZF5 (green) and hFip1 (brown), overlaid with the ZF4–hFip1 complex (gray). Residue Tyr151 is labeled in red. Panels *B* and *C* are modified from an output from ESPript (Gouet et al. 1999).

conformation compared with Arg144, but the guanidinium groups of the two Arg residues are located at the same position. The smaller buried surface area in this interface is because hFip1-A also has some contacts with ZF5 and the C-terminal extension beyond it, especially residue Phe169, which is part of the aromatic residue cluster in ZF4 (Fig. 2A).

There are also contacts between the C-terminal helix of hFip1-A and the loop of hFip1-B (Fig. 1B), giving rise to 300 Å² buried surface area in this interface. However, the bound conformations of the two hFip1 molecules are nearly the same (Fig. 2D), and therefore it is unlikely that these contacts have affected the binding modes of hFip1.

Mutation studies confirm two hFip1-binding sites in CPSF30

To assess the structural observations on the CPSF30–hFip1 interface, especially the 1:2 stoichiometry of the two proteins in the complex, we introduced mutations in the binding site and characterized their effects on the complex. We mutated Tyr127 in ZF4 and the equivalent Tyr151 in ZF5 to Ala, either separately or together.

We first characterized the complexes by gel filtration studies. We used the ZF2 C terminus construct of CPSF30 (residues 62–244, 21 kDa), as this protein alone has higher solubility and is more stable in solution (the ZF4–ZF5 protein alone has very low solubility; it can only be produced when coexpressed with hFip1). This recombinant protein is a monomer in solution by gel filtration, with an apparent molecular weight of 23 kDa (Fig.

3A). When mixed with threefold molar excess of hFip1 (residues 159–200, 5 kDa, corresponding to the segment that interacts with CPSF30 in the structure), wild-type CPSF30 produced a complex with an apparent molecular weight of 34 kDa, suggesting the presence of two copies of hFip1 and one CPSF30 molecule. By comparison, the Y127A and the Y151A mutants could only associate with one copy of hFip1 (29 kDa), while the Y127A/Y151A double mutant lost the ability to form a complex with hFip1 (23 kDa). These observations demonstrate the important role of Tyr127 and Tyr151 in the interactions with hFip1, and confirm that the CPSF30–hFip1 complex has 1:2 stoichiometry in vitro.

ZF4 has higher affinity for hFip1 than ZF5

We next used fluorescence anisotropy studies to determine the affinity between CPSF30 and hFip1. We used a hFip1 protein that contained residues 159–200. We created the S159C/C189S mutant to introduce a cysteine residue outside of the interface with CPSF30, and labeled this protein with FAM. The CPSF30 sample contained residues 62–244, from ZF2 to the C terminus. The K_d for the complex of wild-type CPSF30 and hFip1 was 0.69 ± 0.31 nM (Fig. 3B), indicating a strong interaction between the two proteins. The K_d for the Y151A mutant of CPSF30 was 1.8 ± 0.4 nM, while that for the Y127A mutant was 220 ± 32 nM. Finally, no binding was observed for the Y127A/Y151A double mutant at 1.5 µM concentration. These data indicate that ZF4 has high affinity for hFip1, while ZF5 has ~300-fold lower affinity for this peptide,
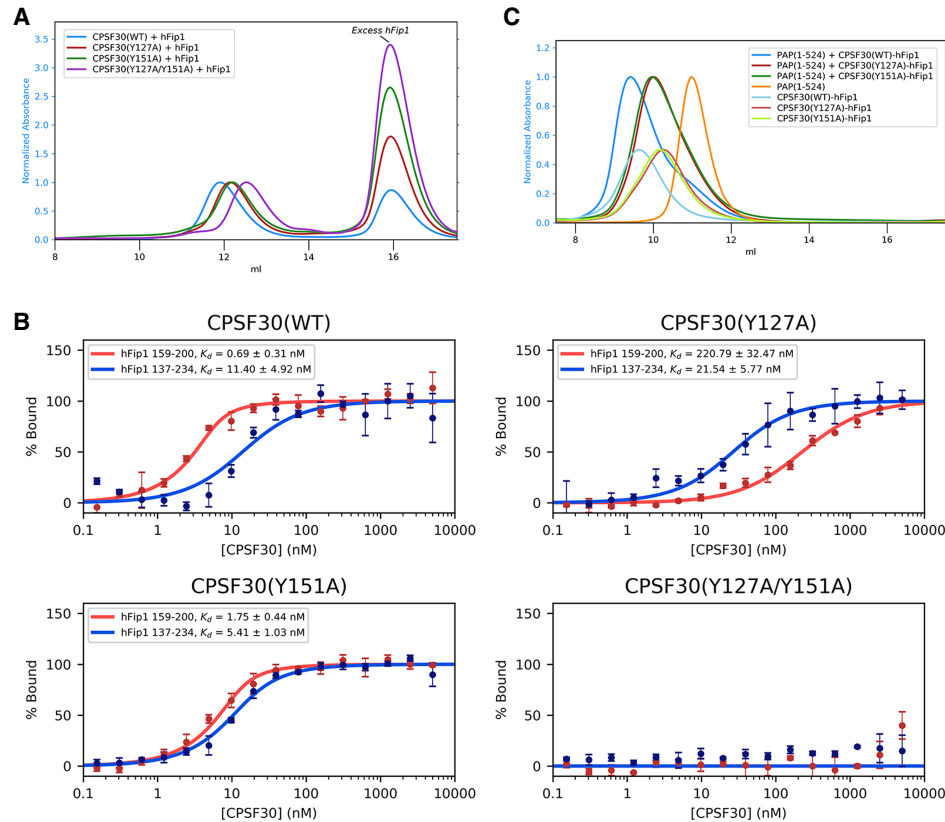
Figure 3. Biochemical and biophysical characterizations of the human CPSF30–hFip1 complex. (*A*) Gel filtration profiles for the mixtures of CPSF30 and hFip1 for wild-type CPSF30, Y127A mutant, Y151A mutant, and Y127A/Y151A double mutant. CPSF30 contained residues ZF2-C terminus (62–244), and hFip1 contained residues 159–200. The peak for excess hFip1 is indicated. (*B*) Fluorescence polarization binding assays for the CPSF30–hFip1 complex. hFip1 residues 159–200 (S159C/C189S mutant, red) and 137–234 (S159C/C189S/C216S mutant, blue) was labeled with FAM, and titrated with increasing concentrations of CPSF30 (residues 62–244). Error bars are ±1 standard deviation from triplicate experiments. (*C*) Gel filtration profiles for mixtures of CPSF30 (full-length), hFip1 (residues 79–200), and the catalytic module of PAP (residues 1–524). The maximum absorbance is arbitrarily scaled to 1. Gel filtration profiles for mixtures of CPSF30 (full-length) and hFip1 (residues 79–200) are also shown, with the maximum absorbance scaled to 0.5.

consistent with the smaller buried surface area in this complex.

We also used a longer hFip1 sample that contained residues 137–243, with the S159C/C189S/C216S mutation to allow labeling with FAM. The K_d values for wild-type and Y151A mutant CPSF30 were 11.4 ± 4.9 nM and 5.4 ± 1.0 nM, respectively (Fig. 3B). By comparison, the K_d values for the Y127A mutant was 22 ± 6 nM, while no binding was observed for the Y127A/Y151A double mutant at 1.5 μM concentration. These data suggest that longer hFip1 may enhance binding to ZF5.

The CPSF30–hFip1 complex can recruit two PAP molecules

The segment of yeast Fip1 that tethers PAP (Meinke et al. 2008) is weakly conserved in hFip1, located N-terminal to the segment that interacts with CPSF30 (Fig. 1A). The 1:2 stoichiometry of the CPSF30–hFip1 complex suggests that it could recruit two copies of PAP. We purified the catalytic module (residues 1–524, 60 kDa) of human

PAPα (Martin et al. 2000) and studied its mixture with CPSF30 and hFip1 by gel filtration. The hFip1 sample used in these studies contained residues 79–200 (13 kDa), to include the region that recruits PAP. The CPSF30 sample used in these studies was full-length (residues 1–244, 27 kDa). The hFip1 protein containing the region that recruits PAP is a dimer in solution, which produced a 2:4 complex of CPSF30–hFip1, while the Y127A and Y151A mutants reduced the complex to 2:2 (Fig. 3C). PAP alone is a monomer in solution (Fig. 3C).

In the mixture of PAP catalytic module with wild-type CPSF30 and hFip1, a complex with an apparent molecular weight of 161 kDa was observed (Fig. 3C), consistent with a 1:2:2 complex of CPSF30–hFip1–PAP (173 kDa). By comparison, the Y127A or Y151A mutant of CPSF30 formed a complex with an apparent molecular weight of 101 kDa, consistent with a 1:1:1 complex (100 kDa). Therefore, these studies confirm that the CPSF30–hFip1 complex can recruit two molecules of the catalytic module of PAP in vitro. They also suggest that the hFip1 dimer dissociated upon formation of the complex with the PAP catalytic

module, probably because the region that mediates this dimerization is also the region that interacts with PAP. The Y127A/Y151A double mutant of full-length CPSF30 has poor behavior in solution and could not be purified for this experiment, likely because it cannot interact with hFip1.

We next tested the ability of wild-type and mutant CPSF30 proteins to support polyadenylation. We observed clear polyadenylation activity with full-length wild-type CPSF30 as well as the Y127A and the Y151A mutants using an RNA containing the AAUAAA poly(A) signal (Fig. 4). In contrast, no polyadenylation activity was observed in a reaction lacking CPSF30 or using an RNA containing the AAGAAA poly(A) signal (Fig. 4). Nonspecific polyadenylation was observed in reactions lacking CPSF30 and hFip1 but containing Mn^{2+} as the divalent cation. These results suggest that both hFip1 binding sites in CPSF30 can support polyadenylation.

Discussion

Overall, our studies have revealed the molecular basis for the recruitment of hFip1 by CPSF30. The residues in the interface are highly conserved among CPSF30 (Fig. 2B) and hFip1 (Fig. 2C) homologs, from yeast to humans, indicating that this binding mode is likely conserved in most organisms. In addition, hFip1 is mostly an unstructured protein, with few conserved segments. The segment that interacts with CPSF30 is the most conserved region among its homologs, especially for fungal Fip1.

Combined with the earlier study on the recognition of the AAUAAA poly(A) signal (Clerici et al. 2018; Sun et al. 2018), we have now defined the molecular basis for the functions of each of the five zinc fingers in CPSF30. The overall structures of the zinc fingers are similar, in the shape of a short, oval cylinder. The zinc atom and its three Cys ligands are located near one end of this cylinder (the "bottom" face). This bottom face is used by ZF2 and ZF3 to recognize an A-A dinucleotide and by ZF1 to recognize a segment of CPSF30 itself (Fig. 1E). In contrast, ZF4 and ZF5 use the "top" face of this cylinder, where the His ligand to the zinc ion is located, to bind hFip1. Therefore, different side chains of the zinc fingers participate in binding the partners (Fig. 1D), depending on whether they are pointed toward the top or bottom face.

Residue Lys191 of hFip1 is in a helix, but it is a cleavage site for trypsin in hFip1-B. A helical conformation for this residue would not be able to access the active site of trypsin, indicating that residues in this region can undergo conformational changes between a helical conformation and a more extended conformation. Trypsin can cleave at this residue when it is in the extended conformation.

Two K_d values would be expected from the fluorescence anisotropy studies on wild-type CPSF30 binding to hFip1, because ZF4 and ZF5 appear to have different inherent affinity for hFip1. However, there is no evidence for this in the binding curve. It could be possible that the interactions between the two hFip1 molecules in the complex enhance the affinity of hFip1 for ZF5, so that the two zinc
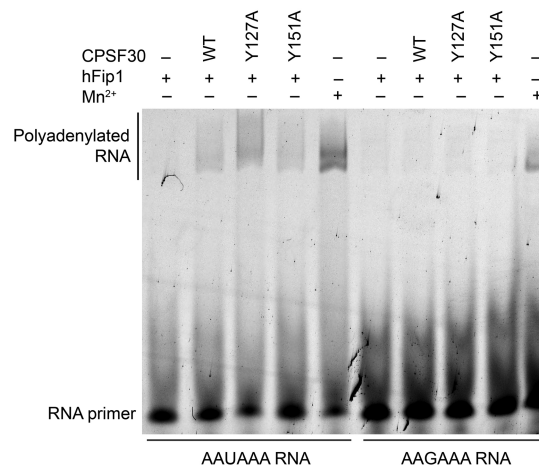

Figure 4. Both hFip1-binding sites in CPSF30 support polyadenylation. The RNA primer contained the AAUAAA or AAGAAA poly(A) signal, with a FAM label at the 5′ end. Full-length wild-type, Y127A, and Y151A mutants of CPSF30 (100 nM) together with hFip1 (100 nM) and CPSF160-WDR33 (50 nM) showed polyadenylation activity with the AAUAAA RNA, while a reaction lacking CPSF30 showed no activity. Nonspecific polyadenylation was observed in the presence of 1 mM Mn^{2+} as the divalent cation. The reactions were carried out in triplicate, and only one replicate is shown. The other two replicates produced essentially the same results.

fingers have comparable apparent affinity for hFip1. The total buried surface area for hFip1-B in this complex is 770 Å², which is more comparable with that for hFip1-A, 1070 Å².

The CPSF30–hFip1 complex provides a link between the recognition of the AAUAAA poly(A) signal and the recruitment of PAP to the machinery. Unexpectedly, our studies revealed a 1:2 stoichiometry between CPSF30 and hFip1 in this complex in vitro, although ZF4 has higher affinity for hFip1. It remains to be established whether this stoichiometry is also true in the active 3′-end processing machinery in vivo. Recent mass spectrometry studies have found up to two copies of Fip1 and Pap1 in the yeast machinery (Casanal et al. 2017). Several other factors in the processing machinery are dimeric, such as CstF77 (Bai et al. 2007a; Legrand et al. 2007; Paulson and Tong 2012), CstF50 (Moreno-Morcillo et al. 2011), and especially the CFIm25-CFIm68 complex (Yang et al. 2011). CFIm recognizes the UGUA sequence element (Hu et al. 2005; Venkataraman et al. 2005; Yang et al. 2010), which is enriched near the distal cleavage site (Wang et al. 2018; Zhu et al. 2018). Therefore, CFIm plays an important role in APA. At the same time, CFIm also interacts with PAP (Dettwiler et al. 2004) and hFip1 (Venkataraman et al. 2005), and therefore a dimeric CFIm in the complex may be compatible with having two copies of hFip1 and PAP.

Materials and methods

Protein expression and purification

Human CPSF30 was cloned into a pET28a vector modified to be ampicillin resistant, either with no affinity tag (residues 114–

173), or with an N-terminal His-tagged yeast SMT3 (residues 62–244 or 1–244). We used isoform 2 of CPSF30, where residues 191–215 are missing. With this isoform, full-length CPSF30 has 244 residues. hFip1 (residues 159–200 or 79–200) and human PAPα (residues 1–524) were each cloned into a pET28a vector, in-frame with N-terminal His-tagged yeast SMT3 as a solubility tag. Mutations to CPSF30 and hFip1 were carried out using the Quik-Change protocol (Agilent).

All expressions were carried out in LB media using *E. coli* BL21 (DE3) cells. Each was induced with 0.5 mM IPTG and incubated for 18 h at 17°C prior to harvesting by centrifugation. The resulting cell pellets were flash-frozen and stored at –80°C until use.

CPSF30, hFip1, PAPα, and CPSF30–hFip1 complexes were all purified using the same protocol, except that in samples containing CPSF30, 100 μM ZnSO₄ was added to all buffers, and those with full-length CPSF30 had at least 500 mM NaCl at all times. First, the cells were lysed using sonication in buffer containing 50 mM Tris (pH 7.5), 200 mM NaCl, 30 mM imidazole, 10 mM β-mercaptoethanol, and 2 mM PMSF. The lysate was centrifuged to clarify and the supernatant incubated with 5 mL Ni-NTA beads (Qiagen) at 4°C. The beads were then washed with two column volumes of buffer containing 2.5 M NaCl, then three column volumes with 100 mM NaCl, and finally eluted with buffer containing 100 mM NaCl and 250 mM imidazole. One-hundred micrograms of UlpI protease was added to the elution and allowed to cleave overnight at 4°C. After cleavage, the sample was run over a 5-mL Fast Flow HiPrep Q column (GE Healthcare) using a 100 mM to 2 M NaCl gradient. Fractions were analyzed using SDS-PAGE, and the appropriate ones were purified further using a Superdex 200 16/60 column (GE Healthcare) with a buffer containing 20 mM Tris (pH 7.5), 100 mM NaCl, and 10 mM DTT. CPSF30 alone was purified in the same manner as above, but with 50 mM arginine and at least 500 mM NaCl in the buffer during all steps, and a second Ni-NTA purification postcleavage and prior to MonoQ purification.

Human CPSF160/WDR33 (residues 1–425) complex was expressed and purified using a previously described protocol (Sun et al. 2018; Hamilton et al. 2019).

Protein crystallization

Crystals of the human CPSF30 (residues 114–173) and hFip1 (residues 159–200) complex were obtained with 14 mg/mL protein using the sitting-drop vapor-diffusion method at 20°C. Trypsin (50 μg/mL) was added to the protein sample prior to setup, and the reservoir solution contained 0.1 M sodium malonate (pH 5.7) and 16% (w/v) PEG 3350. Crystals appeared overnight, and were picked after 2 mo and flash-frozen with liquid nitrogen using 30% (v/v) glycerol as the cryo-protectant.

Data collection and structure determination

X-ray diffraction data were collected at the NE-CAT 24-ID-C beamline at the Advanced Photon Source. The diffraction images were processed using XDS (Kabsch 2010), and the phases were solved using the CRANK2 pipeline (Skubák and Pannu 2013) in the CCP4 suite (Collaborative Computational Project, Number 4 1994) using the anomalous scattering of the zinc atoms. The structure was rebuilt using Coot (Emsley and Cowtan 2004) and refined with Refmac5 (Murshudov et al. 1997). The atomic structure and X-ray diffraction data have been deposited in the Protein Data Bank (entry code 7K95).

Labeling of hFip1 with FAM

hFip1 (100 μM; residues 159–200 or 137-234) mutated to only have one cysteine that was outside of the CPSF30 binding region (S159C/C189S or S159C/C189S/C216S) was mixed with 2 mM fluorescein-5-maleimide (Cayman Chemical Company) in a buffer consisting of 20 mM Tris (pH 7.5), 100 mM NaCl, and 1 mM TCEP. The reaction was allowed to proceed overnight at 4°C in darkness. The next day, unreacted maleimide was rendered inactive by the addition of 10 mM DTT. The fluorescein-labeled hFip1 was then purified using a Superose 12 10/300 column (GE Healthcare).

Fluorescence polarization binding assays

Fluorescence polarization assays were performed on a Synergy Neo2 plate reader (Biotek) with polarizing filters (485 ± 20 nm excitation, 528 ± 20-nm emission). The buffer contained 20 mM Tris (pH 7.5), 200 mM NaCl, 1 mM DTT, 0.03% NP-40, 0.1 μM BSA, and 5 nM FAM-labeled hFip1 for wild-type CPSF30 and 10 nM labeled hFip1 for the CPSF30 mutants. Each titration was carried out in triplicate, and the results were analyzed using a locally developed Python program.

Analytical gel filtration experiments

Fifty microliters of each sample was mixed and incubated for 1 h on ice and then run on a Superose 12 10/300 column (GE Healthcare). The buffer contained 20 mM Tris (pH 7.5), 200 mM NaCl, and 10 mM DTT. The molecular weight of the eluted samples was calculated using a linear equation calibrated with a set of protein standards (Bio-Rad).

Polyadenylation assays

Assays were carried out following a modified version of a protocol detailed previously (Wahle 1995). All reactions were carried out in triplicate in a volume of 7.5 μL using a buffer that consisted of 20 mM Tris (pH 8.0), 10% (v/v) glycerol, 150 mM NaCl, 5 mM MgCl₂, 10 μM EDTA, 0.03% (v/v) NP-40, 10 mM DTT, 1 unit/reaction RNasin Plus RNase inhibitor (Promega), 1 μM BSA (Sigma), and 1 μM PABPN1 (residues 45–296). The AAUAAA RNA primer is from the human adenovirus L3 polyadenylation site, with the sequence UUCAAUAAAGGCAAAUGUUUUUAUUUGUACA. The reactions were heated at 30°C for 5 min prior to addition of 1 mM ATP (Sigma), then allowed to react for 10 min at 30°C prior to stopping with the addition of 7.5 μL 2× stop buffer (40 mM Tris at pH 8.0, 8 M urea, 100 mM EDTA) and heating for 10 min to 65°C. The reactions were separated by running them on an 8% (w/v) acrylamide urea gel in TBE buffer, and then visualized using a Typhon FLA 7000 (GE Healthcare).

References

Bai Y, Auperin TC, Chou C-Y, Chang G-G, Manley JL, Tong L. 2007a. 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

Bai Y, Auperin TC, Tong L. 2007b. The use of in situ proteolysis in the crystallization of murine CstF-77. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63:** 135–138. doi:10.1107/S1744309107002904

Barabino SML, Hubner 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

Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moore CL, Bohm A. 2000. Structure of yeast poly(A) polymerase alone and in complex with 3′-dATP. *Science* **289:** 1346–1349. doi:10.1126/science.289.5483.1346

Casanal A, Kumar A, Hill CH, Easter AD, Emsley P, Degliesposti G, Gordiyenko Y, Santhanam B, Wolf J, Wiederhold K, et al. 2017. Architecture of eukaryotic mRNA 3′-end processing machinery. *Science* **358:** 1056–1059. doi:10.1126/science.aao6535

Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR III, 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, 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

Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50:** 760–763. doi:10.1107/S0907444994003112

Das K, Ma LC, Xiao R, Radvansky B, Aramini J, Zhao L, Marklund J, Kuo RL, Twu KY, Arnold E et al. 2008. Structural basis for suppression of a host antiviral response by influenza A virus. *Proc Natl Acad Sci* **105:** 13093–13098. doi:10.1073/pnas.0805213105

Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SML. 2004. Distinct sequence motifs within the 68-kDa subunit of cleavage factor I$_m$ mediate RNA binding, protein–protein interactions, and subcellular localization. *J Biol Chem* **279:** 35788–35797. doi:10.1074/jbc.M403927200

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

Gouet P, Courcelle E, Stuart DI, Metoz F. 1999. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15:** 305–308. doi:10.1093/bioinformatics/15.4.305

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

Hamilton K, Sun Y, Tong L. 2019. Biophysical characterizations of the recognition of the AAUAAA polyadenylation signal. *RNA* **25:** 1673–1680. doi:10.1261/rna.070870.119

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

Kabsch W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66:** 133–144. doi:10.1107/S0907444909047374

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

Legrand P, Pinaud N, Minvielle-Sebastia L, Fribourg S. 2007. The structure of the CstF-77 homodimer provides insights into CstF assembly. *Nucl Acid Res* **35:** 4515–4522. doi:10.1093/nar/gkm458

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

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

Mayr C. 2017. Regulation by 3′-untranslated regions. *Annu Rev Genet* **51:** 171–194. doi:10.1146/annurev-genet-120116-024704

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

Moreno-Morcillo M, Minvielle-Sebastia L, Mackereth C, Fribourg S. 2011. Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure. *RNA* **17:** 412–418. doi:10.1261/rna.2481011

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255. doi:10.1107/S0907444996012255

Paulson AR, Tong L. 2012. Crystal structure of the Rna14–Rna15 complex. *RNA* **18:** 1154–1162. doi:10.1261/rna.032524.112

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

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

Skubák P, Pannu NS. 2013. Automatic protein structure solution from weak X-ray data. *Nat Commun* **4:** 2777. doi:10.1038/ncomms3777

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

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

Sun Y, Zhang Y, Aik WS, Yang XC, Marzluff WF, Walz T, Dominski Z, Tong L. 2020b. Structure of an active human histone pre-mRNA 3′-end processing machinery. *Science* **367:** 700–703. doi:10.1126/science.aaz7758

Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol* **18:** 18–30. doi:10.1038/nrm.2016.116

Venkataraman K, Brown KM, Gilmartin GM. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19:** 1315–1327. doi:10.1101/gad.1298605

Wahle E. 1995. Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* **270:** 2800–2808. doi:10.1074/jbc.270.6.2800

Wang R, Zheng D, Yehia G, Tian B. 2018. A compendium of conserved cleavage and polyadenylation events in mammalian genes. *Genome Res* **28:** 1427–1441. doi:10.1101/gr.237826.118

Yang Q, Gilmartin GM, Doublie S. 2010. Structural basis of UGUA recognition by the Nudix protein CFI$_m$25 and implications for a regulatory role in mRNA 3′ processing. *Proc Natl Acad Sci* **107:** 10062–10067. doi:10.1073/pnas.1000848107

Yang Q, Coseno M, Gilmartin GM, Doublié S. 2011. Crystal structure of a human cleavage factor CFI$_m$25/CFI$_m$68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure* **19:** 368–377. doi:10.1016/j.str.2010.12.021

Zhang Y, Sun Y, Shi Y, Walz T, Tong L. 2020. Structural insights into the human pre-mRNA 3′-end processing machinery. *Mol Cell* **77:** 800–809.e6. doi:10.1016/j.molcel.2019.11.005

Zhao J, Hyman L, Moore CL. 1999. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63:** 405–445. doi:10.1128/MMBR.63.2.405-445.1999

Zhu Y, Wang X, Forouzmand E, Jeong J, Qiao F, Sowd GA, Engelman AN, Xie X, Hertel KJ, Shi Y. 2018. Molecular mechanisms for CFIm-mediated regulation of mRNA alternative polyadenylation. *Mol Cell* **69:** 62–74.e4. doi:10.1016/j.molcel.2017.11.031