

Structural biology of poly(A) site definition

Qin Yang and Sylvie Doublié*

3' processing is an essential step in the maturation of all messenger RNAs (mRNAs) and is a tightly coupled two-step reaction: endonucleolytic cleavage at the poly(A) site is followed by the addition of a poly(A) tail, except for metazoan histone mRNAs, which are cleaved but not polyadenylated. The recognition of a poly(A) site is coordinated by the sequence elements in the mRNA 3' UTR and associated protein factors. In mammalian cells, three well-studied sequence elements, UGUA, AAUAAA, and GU-rich, are recognized by three multisubunit factors: cleavage factor I_m (CFI_m), cleavage and polyadenylation specificity factor (CPSF), and cleavage stimulation factor (CstF), respectively. In the yeast *Saccharomyces cerevisiae*, UA repeats and A-rich sequence elements are recognized by Hrp1p and cleavage factor IA. Structural studies of protein–RNA complexes have helped decipher the mechanisms underlying sequence recognition and shed light on the role of protein factors in poly(A) site selection and 3' processing machinery assembly. In this review we focus on the interactions between the mRNA *cis*-elements and the protein factors (CFI_m, CPSF, CstF, and homologous factors from yeast and other eukaryotes) that define the poly(A) site. © 2011 John Wiley & Sons, Ltd. *WIREs RNA* 2011 2 732–747 DOI: 10.1002/wrna.88

INTRODUCTION

3' processing of transcripts generated by RNA polymerase II is a fundamental event for the maturation of messenger RNA (mRNA) in eukaryotes. 3' processing has crucial functional importance for regulating terminal intron removal, mRNA stability, efficient nuclear export, and subsequent translation.^{1–3} Dysfunction of the process can perturb cell growth⁴ and is associated with a variety of human diseases.⁵ The majority of eukaryotic pre-mRNAs 3' ends are processed in two tightly coupled steps: an initial site-specific endonucleolytic cleavage followed by the addition of a poly(A) tail at the 3' end of the upstream cleavage product.^{4,6} Metazoan histone mRNAs represent an exception in that they do not receive a poly(A) tail after cleavage.⁷ Despite the fact that only two enzymatic activities, cleavage and poly(A) addition, are required for the reaction, 20 polyadenylation factors have been identified in yeast,^{4,8} and more than 80 protein factors

have been co-purified with the human 3' processing machinery.⁹ The complexity of the 3' processing machinery ensures a seamless cross-talk with other steps of gene expression,¹⁰ precise regulation of the 3' processing,¹¹ and proper definition of the position where the cleavage reaction takes place, which we will refer to as the poly(A) site.¹²

Recognition of the poly(A) site is coordinated by a series of intimately coupled protein:RNA interactions between 3' processing factors and RNA sequence elements (so called *cis*-elements) located on either side of the poly(A) site.^{4,6} These interactions influence not only the assembly of the 3' processing machinery and subsequent efficiency of cleavage and polyadenylation,^{1,4,6,13} but also the selection of alternative poly(A) sites,^{11,12} which is an essential mechanism for generating the transcriptome diversity of complex eukaryotes.^{14,15} An extensive body of reviews is available on the topic of the 3' processing machinery assembly^{1,3–8,11,13,16–19} and structural information of individual factors.¹ In this review, we focus primarily on the molecular mechanisms and structural basis of the interactions between *cis*-elements and protein factors they are associated with.

*Correspondence to: Sylvie.Doublié@uvm.edu

Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT, USA

DOI: 10.1002/wrna.88

3' PROCESSING MACHINERY

Despite the tightly coupled nature of the cleavage and polyadenylation reactions, researchers have managed to isolate these activities from cell extracts, identify participating protein factors and clone the relevant genes.^{4,20–27} In mammals, effective cleavage requires four multisubunit complexes—cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors I_m and II_m (CFI_m and CFII_m)—and a single subunit poly(A) polymerase (PAP). Of these, only CPSF, PAP and poly(A)-binding protein (PABP) are required for the polyadenylation step.^{4,6} In yeast, cleavage factor IA, IB, and II (CFIA, CFIB, and CFII) participate in the cleavage reaction, and CFIA, CFIB, polyadenylation factor I (PFI), poly(A) polymerase (Pap1p) and poly(A)-binding protein (Pab1p) are needed for *in vitro* reconstitution of the polyadenylation activity.^{4,6} In spite of the apparent difference in complex assembly, many individual subunits share high homology with their mammalian counterparts. For example, Rna14p and Rna15p of the yeast CFIA complex are homologous to the 77 and 64 kDa subunit of CstF, whereas Clp1p and Pcf11p of CFIA are homologous to components of mammalian CFII_m. The relationships between mammalian and yeast 3' processing factors are summarized in Table 1 and Figure 1.

A TRIPARTITE MECHANISM FOR POLY(A) SITE RECOGNITION

In mammals, the poly(A) site is recognized through a tripartite mechanism cooperatively mediated by three sets of consensus *cis*-elements: the AAUAAA hexamer, U/GU-rich elements, and UGUA elements⁵⁷ (Figure 1(a)). The AAUAAA hexamer was the first *cis*-element identified to modulate 3' processing,⁵⁸ partially owing to its high conservation.^{15,59} A bioinformatics study showed that AAUAAA and a close variant AUUAAA represent 70 and 15%, respectively, of constitutive poly(A) sites from the ~14,000 human genes investigated. Twelve percent of genes contain 10 types of other variants and only 3% do not harbor any known hexamer.¹⁵ AAUAAA is located 10–35 nucleotides 5' to the poly(A) site.^{60,61} The distance is critical for poly(A) site selection,⁶² as CPSF contacts both AAUAAA hexamer and poly(A) site. The 160 kDa subunit (CPSF160) specifically recognizes the hexamer,²⁵ whereas the poly(A) site is bound by CPSF73, which has recently been shown to be the endonuclease that carries out the cleavage reaction.³⁰ The downstream element is preferentially

GU- and U-rich as determined by biochemical analyses, SELEX assays, and bioinformatics studies (Ref 63 and references within). U/GU-rich elements are generally located 15–30 nucleotides 3' to the poly(A) site²⁷ and are bound by the three-subunit complex CstF. Multiple UGUA elements are generally present within 40–100 nucleotides upstream of the poly(A) site.⁶⁰ More recently CFI_m was shown to bind two UGUA elements simultaneously.⁵² Mutations in these elements influence CFI_m association,⁵² poly(A) site selection, and 3' processing efficiency.⁵⁷ In addition to the three-key *cis*-elements mentioned above, additional auxiliary elements have been identified, including upstream U-rich and UAUA repeats and downstream G-rich sequences.⁶⁰ These auxiliary elements serve as platforms for binding additional 3' processing and regulatory factors to modulate 3' end formation (reviewed in Ref 11).

Despite the apparent differences in sequence and position of the *cis*-elements between yeast and mammals,¹ their 3' processing machinery share a lot of similarities, and possibly a similar tripartite recognition mechanism¹¹: UA repeats, an A-rich element and U-rich regions recognized by several multimeric processing factors (Figure 1(b)). The UA repeats are located at a variable nucleotide distance upstream of the poly(A) site⁶⁴ and bound specifically by Hrp1p, the sole subunit of yeast CFIB.⁵³ The UA repeats are followed by an A-rich element, with a preference for AAUAAN.⁶⁴ Similar to the AAUAAA hexamer in mammals, the A-rich element is also about 10–30 nucleotides 5' to the poly(A) site.⁶⁵ But interestingly, the A-rich element is not bound by CFII, the yeast counterpart of CPSF. Rather, it interacts with CFIA, a homolog of the mammalian CstF complex. While it is known that the presence of U-rich regions surrounding the poly(A) site is critical for efficient 3' processing,⁶⁶ the subunit(s) of cleavage and polyadenylation factor (CPF) they associate with remain to be identified.^{1,66}

MAMMALIAN CFI_m

CFI_m functions at an early step of the 3' processing complex assembly. A smaller 25 kDa subunit (CFI_m25) and three larger 59, 68, or 72 kDa subunits were shown to comigrate with CFI_m activity in HeLa cell nuclear extracts.²³ The human 68 and 59 kDa subunits are encoded by separate genes (CPSF6 and CPSF7, respectively). The 72 kDa form is likely to be the result of alternative splicing of the CPSF6 transcript.⁶⁷ CFI_m is required for the cleavage reaction *in vitro*⁶⁸ and the activity could be reconstituted with the CFI_m25 and 68 kDa subunits.²¹

TABLE 1 | List of Published Structures of 3' Processing Factors

Protein/Complex Name	Organism	PDB ID	Reference
Individual Protein Factors			
CFI _m 25	<i>H. sapiens</i>	2C3L, 2J8Q, 3BAP, 3BHO	28, 29
CPSF73 (Ysh1p/Brr5p) ¹ MBL and β -CASP domain	<i>H. sapiens</i>	2I7T, 2I7V	30
CPSF73 MBL, β -CASP and KH domain	<i>P. horikoshii</i>	3AF5	31
	<i>M. mazei</i>	2XR1	32
CPSF100 (Ydh1p/Cft2p) MBL and β -CASP domain	<i>S. cerevisiae</i>	2I7X	30
CstF64 (Rna15p) N-terminal RRM domain	<i>H. sapiens</i>	1P1T	33
	<i>S. cerevisiae</i>	2X1B	34
CstF64 (Rna15p) C-terminal domain	<i>H. sapiens</i>	2J8P	35
CstF77 (Rna14p) HAT domain	<i>E. cuniculi</i>	2UY1	36
	<i>M. musculus</i>	2O0E, 2OND	37
CstF50 N-terminal dimerization domain	<i>D. melanogaster</i>	2XZ2	38
Symplekin (Pta1p) N-terminal HEAT domain	<i>D. melanogaster</i>	3GS3	39
hPcf11 (Pcf11p) N-terminal CTD interaction domain	<i>S. cerevisiae</i>	1SZ9, 2BF0	40, 41
PAP (Pap1p)	<i>B. taurus</i>	1F5A, 1Q78, 1Q79	42, 43
	<i>S. cerevisiae</i>	1FA0, 201P, 2HHP	44, 45
PABPN1 (Pab1p) RRM domain	<i>H. sapiens</i>	3B4D, 3B4M	46
Protein–Protein Complex			
CFI _m 25, CFI _m 68 RRM complex	<i>H. sapiens</i>	3Q2S	47
CPSF30 ZF2/3, Influenza NS1A complex	<i>H. sapiens</i>	2RHK	48
Clp1p, Pcf11p clp1 binding region complex	<i>S. cerevisiae</i>	2NPI	49
Pcf11p, RNA pol II CTD complex	<i>S. cerevisiae</i>	1SZA	41
Pap1p, Fip1 N-terminal acidic domain complex	<i>S. cerevisiae</i>	3C66	50
Symplekin N-terminal HEAT domain, Ssu72, CTD phosphopeptide complex	<i>H. sapiens</i>	3O2Q	51
Protein–RNA complex			
CFI _m 25–UGUA complex	<i>H. sapiens</i>	3MDG, 3MDI	52
CFI _m 25–CFI _m 68 RRM–UGUA complex	<i>H. sapiens</i>	3Q2T	47
CPSF73 MBL, β -CASP and KH domain–RNA complex	<i>P. horikoshii</i>	3AF6	31
Rna15p RRM–G/GU complex	<i>S. cerevisiae</i>	2X1A, 2X1F	34
Hrp1p RRM1/2–AUUAU complex	<i>S. cerevisiae</i>	2CJK	53
Rna15p RRM–Hrp1p RRM1/2–RNA complex	<i>S. cerevisiae</i>	2KM8	54
Pap1p–AAAAA complex	<i>S. cerevisiae</i>	2Q66	55
PABPC RRM1/2–Poly(A) complex	<i>H. sapiens</i>	1CVJ	56

¹Yeast homologs of mammalian factors are shown in parentheses.

CFI_m increases both the rate and the efficiency of the cleavage reaction²¹ by binding to an upstream sequence element.^{57,69} SELEX experiments identified the RNA-binding specificity of CFI_m toward UGUA sequences.⁶⁹ Interactions between CFI_m and UGUA elements are essential for CFI_m-mediated regulation of the cleavage reaction *in vitro*⁶⁹ and *in vivo*.⁵⁷

CFI_m25 contains a Nudix domain (residues 77–202),⁷⁰ which generally catalyzes the hydrolysis of

nucleoside diphosphates linked to other moieties, X⁷¹ (Figure 2(a)). CFI_m25 possesses the canonical $\alpha/\beta/\alpha$ Nudix fold^{28,29} (Figure 2(b)); however, two unique features of CFI_m25 converted the Nudix hydrolase domain into a sequence-specific single-stranded RNA-binding domain (RBD).^{28,29,52} First, CFI_m25 lacks two of the four conserved glutamate residues that coordinate divalent ions and are critical for catalysis in the Nudix enzyme.^{28,72} As a result, CFI_m25 is able to

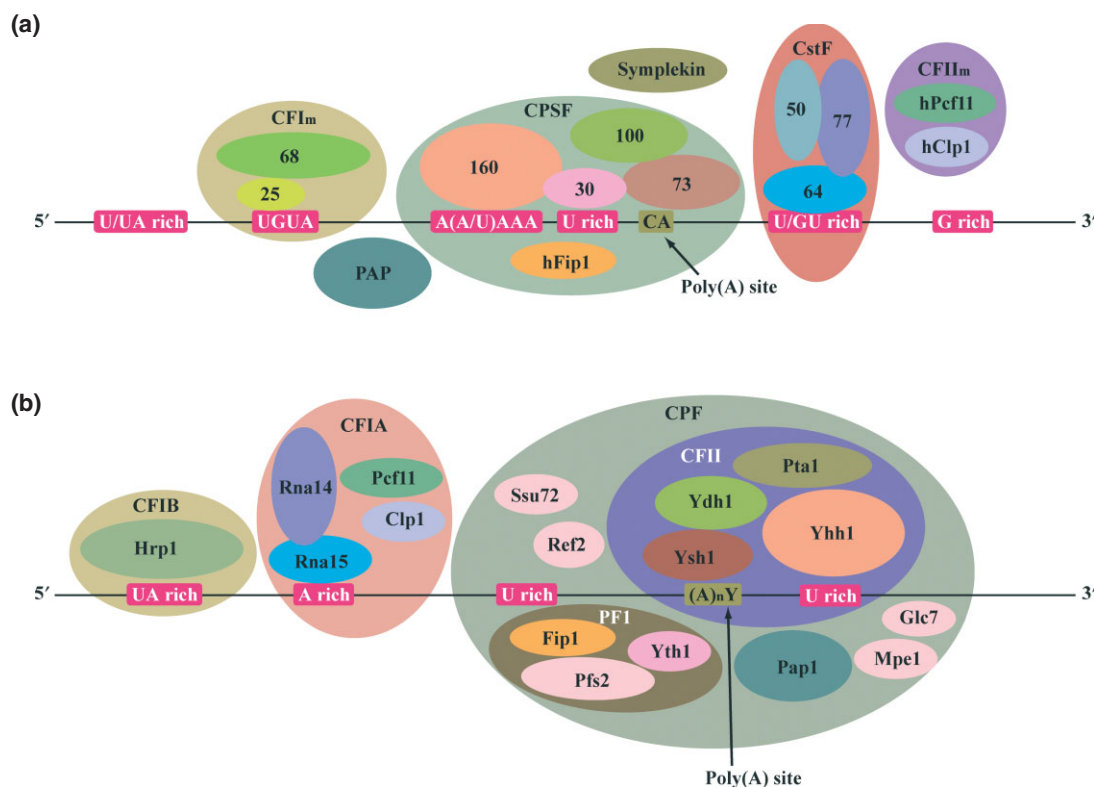


FIGURE 1 | Schematic representation of the (a) mammalian and (b) yeast 3' processing machinery. The RNA *cis*-elements are shown in magenta. Homologous protein factors in yeast and mammals are depicted in the same color.

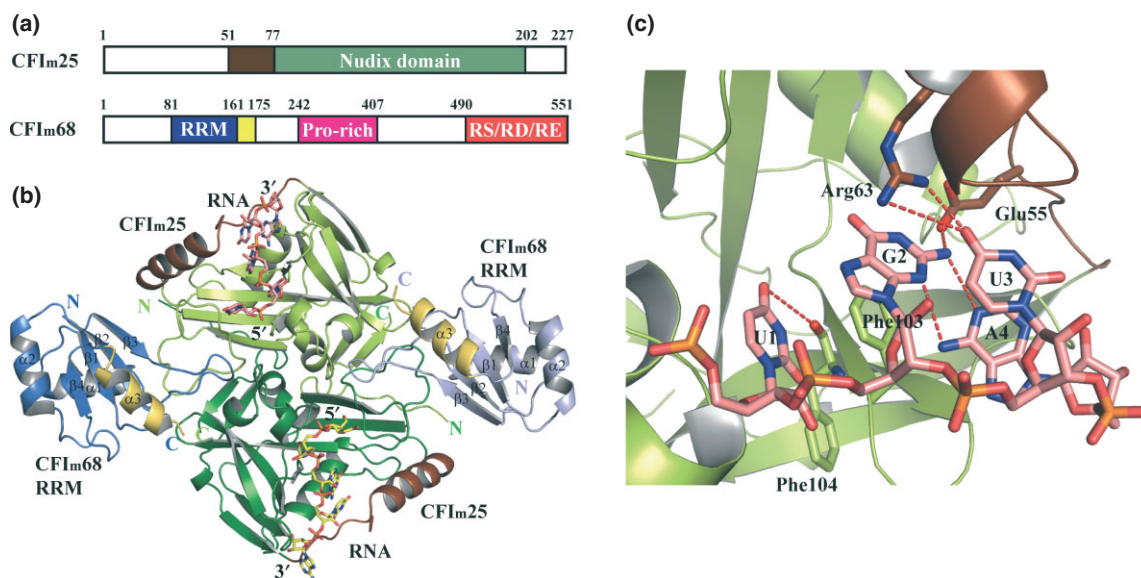


FIGURE 2 | Cleavage factor I_m (CFIm). (a) Domain organization of CFIm subunits CFIm25 and CFIm68. (b) Crystal structure of the CFIm25/CFIm68 RRM/RNA complex (PDB ID: 3Q2T). The CFIm25 homodimer is colored in green and lime. The highly conserved loop/helix motif (residues 51–76) is highlighted in brown. Each CFIm25 monomer binds to one UGUA element, shown as a stick model and colored in yellow and salmon. CFIm68 monomers are colored in blue and cyan. The additional C-terminal α -helix of the RRM is highlighted in gold. (c) Close-up view of the CFIm25-UGUA interactions. The protein and RNA are colored as in (b). Hydrogen bonds are represented by red dashed lines.

bind, but not hydrolyze dinucleotide substrates, such as diadenosine tetraphosphate (Ap₄A).²⁸ Second, the highly conserved loop–helix motif (residues 51–76) immediately preceding the Nudix domain not only occludes the conventional substrate binding pocket of Nudix proteins, but also provides key residues to form an alternative binding pocket⁵² (Figure 2(b)). The crystal structure of a CFI_m25–RNA complex illustrated that CFI_m25 utilizes this new binding pocket to specifically bind UGUA elements.⁵² U1, G2, and U3 are recognized by the main chain amide and carbonyl of Phe104, side chain of Glu55, and side chain of Arg63, respectively. A4 is recognized through an intramolecular sugar-edge/Watson–Crick base pair interaction with G2.⁷³ Phe103 provides additional hydrophobic forces to stabilize UGUA (Figure 2(c)).⁵² CFI_m25 forms a stable homodimer in the absence or presence of RNA,^{28,29,52} which allows it to bind two UGUA elements simultaneously.⁵² Ap₄A is bound in the RNA-binding pocket via the same protein residues (Arg63 and Phe103) that interact with the UGUA element. This mutually exclusive binding property hints at a potential regulatory role of small molecules in 3' processing. A dramatic conformational change of conserved residue Arg63 suggests it might serve as a sensor to distinguish between RNA and small molecules.⁵²

In addition to RNA binding, CFI_m25 also participates in the assembly of the 3' processing machinery. The N-terminal 60 residues and residues 81–160 of CFI_m25 have been shown to interact with the C-terminal 60 residues of PAP⁷⁴ and PABPN1,⁷⁰ respectively. Acetylation of Lys23 of CFI_m25 and lysines 635/644/730/734 of PAP decreases the strength of the interaction between CFI_m25 and PAP.⁷⁵

The larger subunit of CFI_m (CFI_m68 or CFI_m59) possesses three distinct domains: an N-terminal RNA recognition motif (RRM), also known as RBD, a central proline-rich region, and a C-terminal RS-, RD-, RE-dipeptide repeat region⁷⁰ (Figure 2(a)). This domain organization is similar to that of the splicing regulator (SR) family proteins, which are essential in regulating pre-mRNA splicing.⁷⁶ Indeed, the C-terminal RS/RD alternating charge domain has been demonstrated to interact with splicing factors^{70,77} and CFI_m has also been shown to co-purify with the spliceosome.^{78–80}

The N-terminal RRM domain of CFI_m68 is necessary and sufficient for interaction with CFI_m25.⁷⁰ A recently reported crystal structure⁴⁷ showed that the RRM of CFI_m68 displays the canonical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ topology, with two α -helices packed on one side of the four-stranded antiparallel β -sheet (Figure 2(b)). In addition to the core $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold, an extended

C-terminal helix α_3 is present in CFI_m68 (Figure 2(b)), which has been observed in several other RRM containing proteins (reviewed in Ref 81). In CFI_m68, α_3 sits on top of the β -sheet and blocks the canonical RNA-binding surface.⁴⁷ Furthermore, a comparison with other RRM sequences indicated that CFI_m68 is missing one of the three critical aromatic residues that stack onto the RNA bases.⁴⁷ Taken together, these observations provide an explanation as to why CFI_m68 does not bind RNA on its own.⁷⁰

The crystal structure of the CFI_m25–CFI_m68 RRM–RNA complex⁴⁷ reveals that CFI_m is a heterotetramer organized in an unexpected manner in that two monomers of CFI_m68 flank the CFI_m25 homodimer rather than interact with individual CFI_m25 monomers (Figure 1(b)). The interactions are mainly mediated by the loops connecting β_1/α_1 and β_2/β_3 of the CFI_m68 RRM. Mutational analysis showed that both loops are involved in RNA binding, whereas the β_2/β_3 loop is critical for CFI_m complex formation. CFI_m25 retains the same homodimer conformation and ability to bind two UGUA elements in the context of the CFI_m complex.⁴⁷ CFI_m68 enhances the RNA-binding affinity by stabilizing the RNA loop connecting two UGUA elements.⁴⁷ The weak RNA-binding affinity provided by the β -sheet allows CFI_m68 to bind loops of various lengths.⁴⁷ Moreover, the potential long range looping of RNA sequences may allow CFI_m to loop out an entire poly(A) site and thereby induce the selection of an alternative poly(A) site.⁴⁷ This hypothesis is consistent with recent observations that CFI_m modulates alternative polyadenylation.^{82,83}

Besides its fundamental role in the 3' processing, and postulated involvement in coupling splicing and 3' processing,⁷⁷ CFI_m has also been shown to connect 3' processing and mRNA export by interacting with mRNA export receptor NXF1.⁸⁴

CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR

CPSF is composed of a 160, 100, 73, 30 kDa subunits and a recently identified fifth subunit human Fip1 (hFip1)^{1,20} (Figure 1(a)). CPSF is required for *in vitro* reconstitution of both cleavage and polyadenylation reactions.^{4,6} It provides a platform to anchor other 3' processing factors, such as symplekin,⁸⁵ PAP,²⁰ and many others.^{1,9} CPSF also recognizes the canonical 3' processing signal AAUAAA hexamer.²⁵ Most importantly, CPSF is responsible for the endonucleolytic cleavage reaction.^{30,86}

CPSF73 was brought into the spotlight after more than two decades of hunting for the 3' end

endonuclease.⁸⁶ A crystal structure of the N-terminal of CPSF73 in complex with zinc ions provided the molecular basis for the nuclease activity.³⁰ CPSF73 is a member of β -CASP (metallo- β -lactamases-associated CPSF Artemis SNM1 PSO2) subfamily,⁸⁷ which belongs to the large metallo- β -lactamase (MBL) family,⁸⁸ members of which have been characterized as zinc-dependent nucleases.⁸⁹

The N-terminus of CPSF73 comprises two separate domains, a central β -CASP domain (residues 209–394), and an integral MBL domain folded from two separate regions (residues 1–208 and 395–460)³⁰ (Figure 3(c)). The β -CASP domain is made of a central parallel β -sheet flanked by two α -helices on each side⁸⁹ (Figure 3(a) and (b)). This $\alpha/\beta/\alpha$ domain architecture is also seen in other β -CASP proteins, such as CPSF100,³⁰ RNA degradation protein TTHA0252,⁹⁰ and RNase J.⁹¹ The MBL domain is the catalytic domain.³⁰ It adopts the canonical MBL four-layer sandwich fold, with two

β -sheets flanked by α -helices.³⁰ Similar to other MBL proteins, zinc is required for the nuclease activity of CPSF73⁹²: dialysis of the nuclear extract, or addition of zinc-specific chelators inhibit the cleavage reaction, and activity can be rescued by adding zinc back.⁸⁶ As expected, two zinc ions were observed in the active site.³⁰ Both ions are coordinated by the conserved residues from the characteristic motifs of the MBL family^{30,92} and align perfectly well to interact with the oxygen atoms from SO_4^{2-} , which originates from the crystallization solution and possibly mimics a phosphate group of the RNA substrate³⁰ (Figure 3(b)). An acid catalysis mechanism was proposed with His396 acting as the general acid, which would be activated by the adjacent Glu204³⁰; mutating either residue decreases the nuclease activity⁹³ and causes cell death in yeast, in which Brr5/Ysh1 is homologous to CPSF73.⁸⁶ The catalytic center lies at the interface between the MBL and β -CASP domains.⁸⁷ Residues from the β -CASP domain do not contribute to zinc

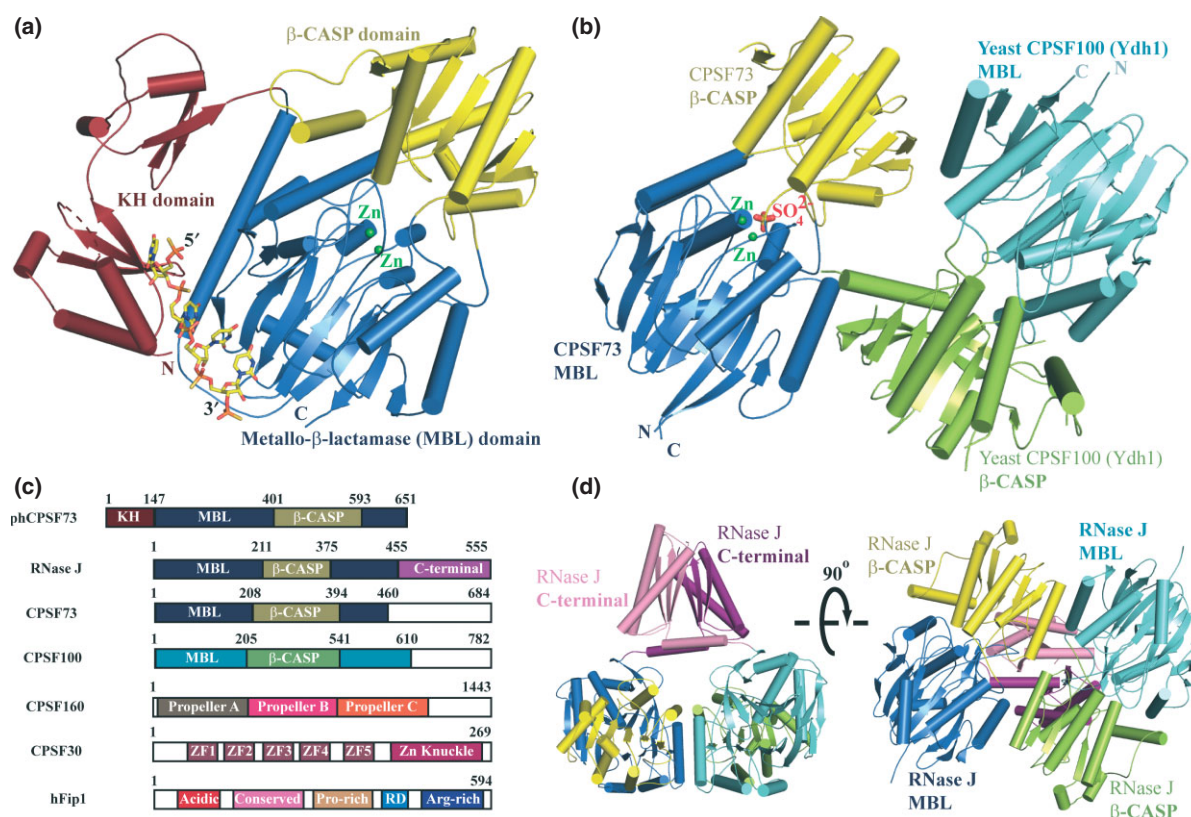


FIGURE 3 | Cleavage and polyadenylation specificity factor (CPSF). (a) Crystal structure of the archaeal *Pyrococcus horikoshii* CPSF73 in complex with RNA (PDB ID: 3AF6³¹). The KH domain, metallo- β -lactamase (MBL) domain, and β -CASP domain are shown in cartoon mode, with α -helices in cylindrical representation, and colored in red, blue, and yellow, respectively. The RNA is shown as a stick model (yellow). The two zinc ions are depicted as green spheres. (b) Crystal structures of human CPSF73 (2I7V³⁰) and yeast CPSF100 (Ydh1) (2I7X³⁰). MBL and β -CASP domains are shown in blue and yellow, respectively, for CPSF73, and in cyan and green, respectively, for CPSF100 (Ydh1). The zinc ions and sulfate bound by CPSF73 are shown. The position and orientation of CPSF73 and CPSF100 are based on the superposition with the RNase J homodimer. (c) Domain organization of CPSF subunits and RNase J. (d) Homodimer conformation of RNase J (3BK1⁹¹).

binding, and the β -CASP domain docks so close to the MBL domain that a conformational change was proposed to allow RNA to access the binding site.³⁰

A recently reported structure of the archaeal *Pyrococcus horikoshii* CPSF73 (PkCPSF73) unveiled an interesting feature³¹ (Figure 3(a)). The β -CASP and MBL domains of PkCPSF73 are structurally very similar to those of CPSF73, but PkCPSF73 possesses two additional KH domains, a common single-stranded RNA-binding motif,⁹⁴ situated at the N-terminus³¹ (Figure 3(a)). In collaboration with the MBL domain, the KH domains of PkCPSF73 are capable of binding RNA,³¹ suggesting that PkCPSF73 might be able to bind RNA sequence specifically.

CPSF100 is also a member of the β -CASP family.⁸⁷ The crystal structure of yeast CPSF100 (Cft2/Ydh1) revealed an expected MBL plus β -CASP domain organization, with each domains folded into a three-dimensional architecture similar to that of CPSF73³⁰ (Figure 3(b)). Due to the absence of all six zinc-coordinating residues, no zinc was detected in the Ydh1 structure and the protein exhibits no nuclease activity.³⁰ In contrast to yeast, the higher eukaryotic CPSF100 might possess a weak nuclease activity.⁹³ CPSF100 lacks only one out of the six residues that are responsible for binding zinc in higher eukaryotes,⁹³ which might result in a single metal binding zinc binding site or a combination of a weaker and stronger site. In support of this hypothesis, MBL family proteins have been shown to remain active, albeit with lower activity, in the presence of one ion (reviewed in Ref 92). Moreover, CPSF100 has an arginine (Arg543) at the position corresponding to His396 in CPSF73,^{30,93} which could also serve as general acid.⁹⁵

The simultaneous presence of CPSF73 and CPSF100 in every pre-mRNA processing machinery, including yeast,⁹⁶ plants,⁹⁷ metazoan poly(A), and histone-pre-mRNA,⁸⁵ suggests a tight association of these two subunits. The C-terminal 245 amino acids of CPSF73 are critical for binding to CPSF100.⁹⁸ A recently solved structure of the RNase J homodimer, a member of the β -CASP subfamily, demonstrated the importance of the C-terminal region in dimerization⁹¹ (Figure 3(d)). CPSF73 might associate with CPSF100 in a similar fashion (Figure 3(b)). The interaction between CPSF73 and CPSF100 is specific,⁹⁸ because the C-terminal of CPSF73 does not bind to RC-74, an isoform of CPSF100 in humans.⁹⁸ In addition to protein–protein interactions, the C-terminal region of CPSF73 might be responsible for the exonuclease activity detected in histone-pre-mRNA processing,⁹⁹ as no exonuclease activity was detected with N-terminal MBL- β -CASP domain

alone.³⁰ Intriguingly, the C-terminal domain of RNase J is required for the exonuclease activity.⁹¹ However, another possibility is that the exonuclease activity requires the presence of other component(s) of the 3' processing complex.⁹⁹

The other three subunits of CPSF—CPSF160, CPSF30, and hFip1—are capable of binding RNA. UV cross-linking using either nuclear extracts¹⁰⁰ or recombinant CPSF160²⁵ has demonstrated a preference toward the AAUAAA hexamer or its close variants, suggesting that CPSF160 interacts with AAUAAA directly. However, the residues or domains involved in binding have yet to be identified.¹ The intimate association of CPSF160 and CPSF73 may restrict the allowed distance between the AAUAAA hexamer and poly(A) site in mammals.⁶⁰ Yhh1/Cft1, the yeast homolog of CPSF160, is able to bind RNA close to the A-rich *cis*-element, but the sequence preference and exact site of binding remain to be determined.¹⁰¹

CPSF30 contains five CCCH zinc fingers (ZF1–5) followed by a CCHC zinc knuckle²² (Figure 3(c)), two common RNA-binding motifs, for which there are several examples of structures in complex with single-stranded RNA.⁹⁴ The zinc finger motifs and zinc knuckle of CPSF30 have been shown to bind RNA. Similar to other Zinc finger motifs,⁹⁴ CPSF30 displays a high affinity toward U-rich sequences.²² Similar sequences have been found between the AAUAAA hexamer and poly(A) site.⁶⁰ Moreover, Yth1, the yeast homolog of CPSF30, binds to the U-rich region surrounding the poly(A) site, as illustrated by footprinting assays.¹⁰² Besides binding RNA, the zinc finger motifs are also mediating protein–protein interactions, e.g., CPSF160 and PAPB.¹⁰³ A complex structure of Influenza virus Non-Structural protein 1A (NS1A) and ZF2/ZF3 of CPSF30 illustrated how this virus suppresses host antiviral systems by blocking the 3' processing machinery.⁴⁸

hFip1 is an integral component of the CPSF complex.²⁰ Disorder prediction programs suggest that this 66 kDa protein is highly flexible²⁰ (Figure 3(c)). hFip1 is composed of an acidic domain at the N-terminus, followed by a highly conserved segment and a proline-rich region. The C-terminus contains an RD-rich and an Arg-rich domain, which are missing in the yeast homolog Fip1p.²⁰ A crystal structure of a complex of yeast Pap1p with a peptide of Fip1p reveals that Fip1p binds to Pap1p through its N-terminal acidic domain (residues 80–105).⁵⁰ GST pull down assays indicated that hFip1 binds to CPSF30 through its conserved segment and to CPSF160 through the C-terminal RD-rich and Arg-rich domains.²⁰

Interestingly, UV cross-linking demonstrated that hFip1 binds to U-rich RNA sequences preferentially through its C-terminal Arg-rich domain. A subsequent footprinting assay showed that the region between the AAUAAA and poly(A) site is protected by hFip1, which is consistent with its U-rich sequence preference.²⁰ The close spatial relationship between CPSF160, hFip1, and the *cis*-elements they recognize suggests a cooperative poly(A) site recognition enhanced by protein–protein interactions. Similar interactions exist in the Hrp1p–Rna15p–RNA ternary complex (see below).

CLEAVAGE STIMULATION FACTOR

Both the CstF complex and its yeast homolog are essential for the cleavage reaction,¹⁰⁴ by recognizing the downstream U/GU-rich *cis*-elements that define the poly(A) site.¹⁰⁵ On the other hand, CstF is not required for the polyadenylation step,¹⁰⁶ although its presence could enhance the reaction.¹⁰⁷ CstF is composed of three subunits, 50, 64 and 77 kDa, all of which are required for activity¹⁰⁵ (Figure 1(a)). Far-Western blotting assays demonstrated that CstF77 bridges CstF64 and CstF50, which do not interact with each other.¹⁰⁸

CstF64 was initially implicated in sequence-specific polyadenylation for its ability to UV cross-link to AAUAAA-containing sequences in nuclear extracts,¹⁰⁹ which was later shown to be a result of interactions between CstF and CPSF complex.^{104,110} Instead of the AAUAAA hexamer, CstF64 preferentially binds to GU-rich downstream *cis*-elements^{111–113} (Figure 1(a)). The conserved N-terminal RRM of CstF64, or its yeast homolog Rna15p, is necessary and sufficient for sequence-specific recognition.^{33,34,113} The crystal structure of the Rna15p RRM–RNA complex³⁴ revealed the molecular basis of the preference toward the GU dinucleotide (Figure 3(c)): The RRM is folded in the typical $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ conformation. Two RNA recognition sites are present in the RRM, which are located on loop β_1/α_1 (site-I) and the central β -sheet (site-II)—the canonical RNA-binding surface.³⁴ The specificity of both sites is attained by hydrogen bonds between base-edges and main chain carbonyl and amide atoms, which distinguish G and U from A and C. Stacking forces provided by Tyr27 and Arg87 further strengthen the RNA-binding affinity of site-I³⁴ (Figure 4(b)). The conserved aromatic residue on β_1 (Tyr21) is responsible for stabilizing the RNA substrate at site-II³⁴ (Figure 4(b)). Another notable feature is the movement of the C-terminal region of the RRM upon RNA binding. In the absence of RNA,

the C-terminal region occludes the RNA-binding site by packing on top of the β -sheet through either a short α -helix (CstF64³³) or a loop sampling various conformations (Rna15p³⁴) (Figure 4(b)). Upon RNA addition, the C-terminal region unfolds³³ and exposes the RNA-binding site.^{33,34}

The highly conserved C-terminal region (residues 529–577) of CstF64 forms a three-helix bundle, with two 4-turn helices connected by a 2-turn helix³⁵ (Figure 4(e)). Domains with a similar conformation have been reported in Dia1, a cytoskeletal remodeling protein,¹¹⁵ and the RNA-silencing suppressor p21,¹¹⁶ both of which are involved in protein–protein interactions.^{115,116} Similarly, Rna15p binds to yeast Pcf11, another subunit of yeast CFIA, through its C-terminal domain.³⁵ Growth deficiency or cell death caused by deleting the C-terminal region of Rna15p reinforces the importance of this region.³⁵

CstF64 has also been shown to bind other 3' processing factors, such as CstF77 and symplekin, through the hinge region, which is located immediately after the RRM domain^{117,118} (Figure 4(e)), whereas in Rna15p, residues between the hinge and C-terminal domain are also required for binding Rna14p,³⁵ the yeast homolog of CstF77.

CstF77 is composed of an N-terminal HAT (half a tetratricopeptide repeat [TPR]) domain (residues 1–550) and a C-terminal proline-rich region (residues 560–630)¹⁰⁵ (Figure 4(f)). The HAT repeat is a derivative of the 34 amino acid TPR,¹¹⁹ in that it retains the hallmark aromatic side chains, but lacks the highly conserved alanine and glycine residues.¹²⁰ TPR is commonly involved in protein–protein interactions, including both TPR–TPR interactions and TPR–non-TPR.¹¹⁹ The crystal structures of murine³⁷ and *Encephalitozoon cuniculi*³⁶ CstF77 reveals that it forms a tail-to-tail homodimer through HAT–HAT interactions. CstF77 contains 12 HAT repeats. Six of the repeats at the C-terminus (HAT-C) are involved in the dimerization^{36,37} (Figure 4(a) and (e)). The cavity formed at the V shape dimerization interface is a potential protein–protein interaction region,³⁷ as seen in human PEX5 in complex with a targeting signal peptide.¹²¹ Indeed, the HAT-C domain interacts with the second propeller of CPSF160 (residues 400–973) as demonstrated by yeast two-hybrid assays,³⁷ even though the nature of the interaction is unknown. The first five HAT repeats are not involved in dimerization. They orient roughly perpendicularly to the HAT-C, and form a flat platform with the conserved residues exposed at the surface.³⁶ This platform is another hot spot for CstF77 to interact with other 3' processing factors.³⁶

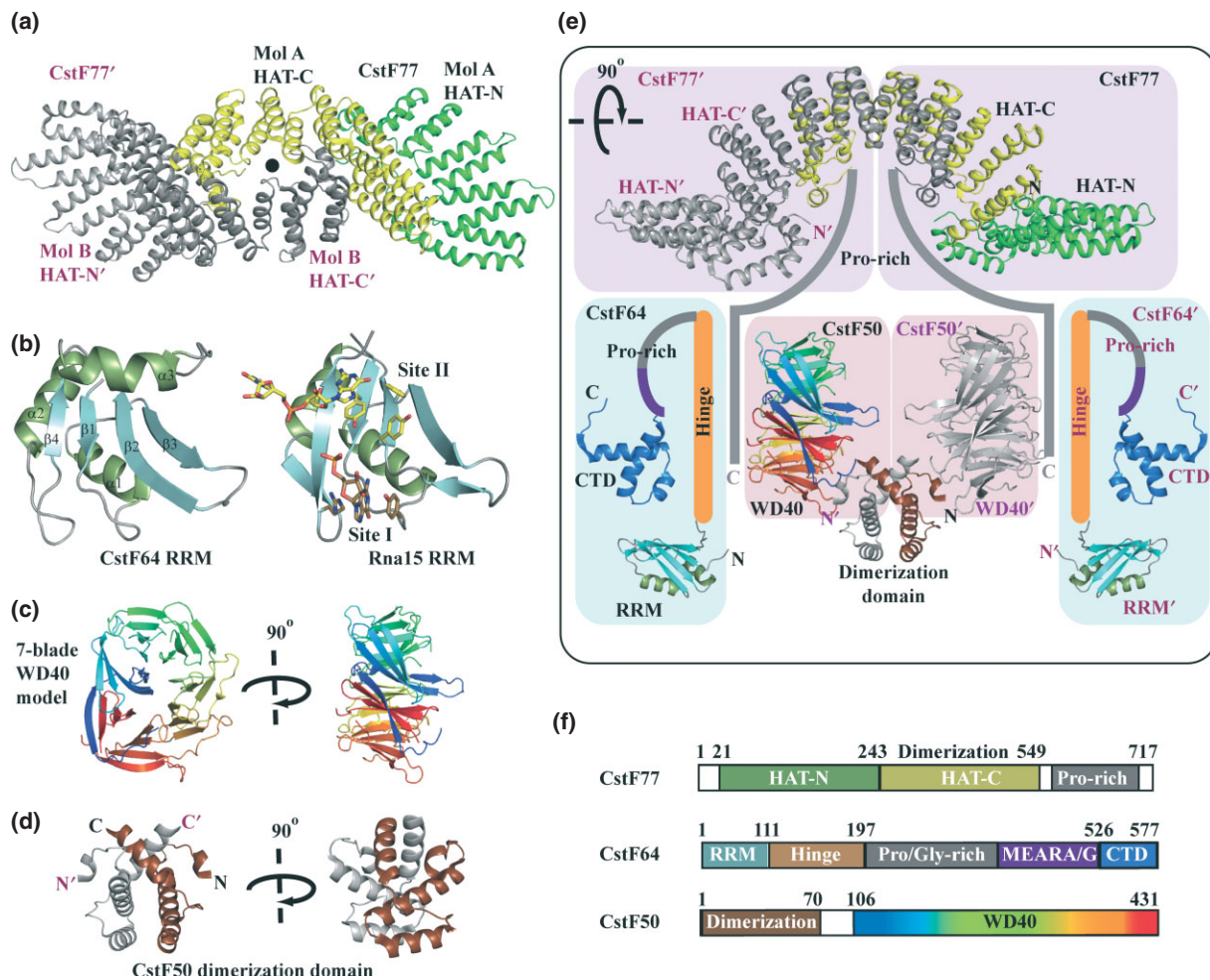


FIGURE 4 | Cleavage stimulation factor (CstF). (a) Crystal structure of the CstF77 homodimer (200E³⁷). The HAT-N domain is shown in green. The HAT-C domain, which is involved in dimerization, is shown in yellow. The second monomer in the dimer is shown in gray. (b) Structure of human CstF64 (1P1T³³) and the yeast CstF64 homolog Rna15p/RNA complex (2X1A, 2X1F³⁴). Two RNA molecules bound at sites I and II are shown in tan and yellow, respectively. (c) A crystal structure of a seven-blade WD40 domain from Prp19 (3LRV¹¹⁴) is presented as a reference for the size and shape of CstF50. (d) Crystal structure of the CstF50 N-terminal dimerization domain (2XZ2³⁸). (e) A model for the hexameric architecture (2:2:2 stoichiometry) of the CstF complex. The second monomer in the dimers is labeled with a prime (') symbol. The atomic structures are represented on the same scale. (f) Domain organization of the CstF subunits.

The C-terminal proline-rich region of CstF77 is a key element for the assembly of the CstF complex. It interacts with the hinge region of CstF64^{37,108,117} and the WD-40 repeat of CstF50¹⁰⁸ (Figure 4(e)). The dimeric organization of CstF77 suggests a 2:2:2 stoichiometry of the CstF complex³⁷ (Figure 4(e)). One benefit from the dimerization is the presence of two RRM from CstF64, which could enhance RNA-binding ability.⁸¹ In fact, in yeast, analytical ultracentrifugation demonstrated that Rna14p and Rna15p form an A₂B₂ tetramer,¹²² and the addition of Rna14p increases RNA binding significantly.¹²² The C-terminal domain of the *Arabidopsis thaliana* homolog of CstF77 has been shown to bind RNA, but this might be a plant-specific property.¹²³

The CstF50 structure is dominated by seven WD40 repeats (residues 93–425).¹²⁴ The WD40 repeat is a conserved structural motif that generally comprises 40–60 amino acids and ends with a C-terminal tryptophan-aspartic acid dipeptide (WD).¹²⁵ Proteins containing WD40 repeats function in a variety of cellular processes by providing a platform for protein complexes assembly.^{125,126} It was shown by far-Western blotting assay that CstF50 binds to CstF77 through its WD40 repeats.¹⁰⁸ Although the structure of the N-terminal homodimerization domain of CstF50 was recently published³⁸ no three-dimensional structure of the C-terminal region has been reported yet; it is plausible that the WD40 repeats of CstF50 fold

into the conserved globular seven-blade β -propeller¹⁰⁸ (Figure 4(c)) and the depression formed at the center of β -propeller anchors protein partners, as seen in the histone methyltransferase subunit WDR5/Histone H3 peptide complex.¹²⁷ In support of this prediction, deletion of any repeats greatly reduces CstF50–CstF77 binding.¹⁰⁸ Besides CstF77, the WD40 domain of CstF50 also binds to BARD1,^{128–130} which modulates the cell cycle in response to DNA damage together with its binding partner BRCA1.¹³¹ Pfs2, the possible yeast homolog of CstF50, interacts with Rna14p and bridges yeast CFIA and CPF complexes through its WD40 domain.¹³²

The N-terminus of CstF50 does not contain any predicted domain. This region is responsible for CstF50 self-association,^{38,108} which might promote CstF dimerization³⁷ (Figure 4(d) and (e)). This region is also critical for CstF50 binding to the CTD of RNA polymerase II,¹³³ and hence to establish a link between 3' processing and transcription.⁸

Hrp1p and Rna15p, THE YEAST HOMOLOGS OF MAMMALIAN CFI_m AND CstF64

Hrp1p, also known as Nab4,¹³⁴ is a component of *S. cerevisiae* CFI.²⁴ Hrp1p may be the closest *S. cerevisiae* ortholog of CFI_m.⁵⁷ Hrp1p is required for polyadenylation. However, whether it is essential for reconstitution cleavage *in vitro* is debatable given the opposing biochemical data available.^{65,134,135} The function of Hrp1p is associated with its recognition of the AU-rich enhancement element (EE) located upstream to the cleavage site^{65,136,137} (Figure 1(b)). The molecular basis of the specificity toward AU dinucleotide repeats was uncovered by the NMR structure of a Hrp1p–RNA complex.⁵³ Hrp1p contains two RRM domains located centrally and arranged in tandem (Figure 5(b)). Both RRM domains are involved in AU recognition.⁵³ Similar to other sequence-specific RRM domains,⁹⁴ the base recognition is mediated by hydrogen bonding between the base-edges and the side chains/main chain atoms on the β -sheet.⁵³ A notable feature of the structure is that besides the β -sheet face, the β 1/ α 1 loop also participates in RNA recognition,⁵³ which is reminiscent of the RNA-binding site-I of CstF64.³⁴ The C-terminal of Hrp1p contains repeats of a RGGF/Y tetra-peptide, which have no known function.⁶⁵

A recently solved solution structure of Hrp1p–Rna15p–RNA ternary complex illustrated how poly(A) site recognition can be modulated by protein–protein interactions.⁵⁴ Unlike its mammalian

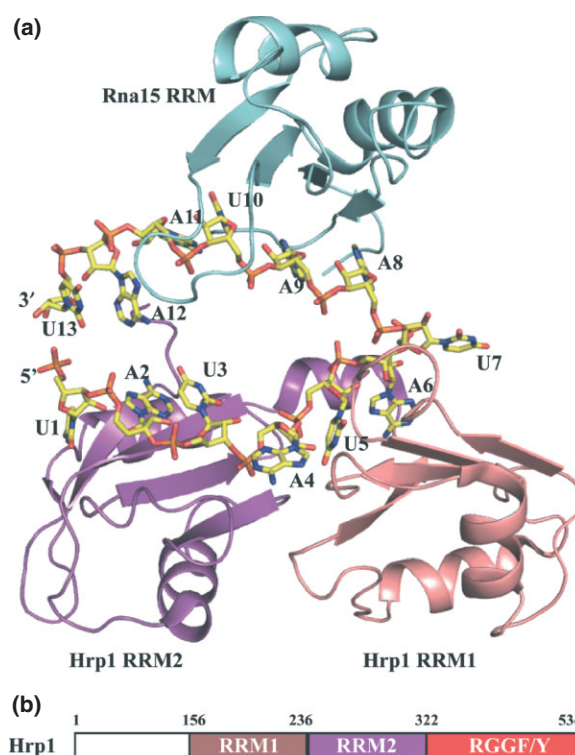


FIGURE 5 | Hrp1p. (a) NMR structure of the Hrp1p–Rna15p–RNA complex (2KM8⁵⁴). Hrp1p RRM1 and RRM2 are shown in salmon and magenta, respectively. Rna15p RRM is colored in cyan. The RNA molecule is shown in yellow. (b) Domain organization of Hrp1p.

homolog CstF64, which has a clear preference for the U/GU-rich element,³³ Rna15p displays a different sequence specificity depending on whether it is unliganded or bound to another factor^{34,54}: in solution, Rna15p alone binds to GU-rich sequences preferentially with a binding affinity of $\sim 20 \mu\text{M}$ (Figure 4(c)), and in a selection experiments, the selectivity of GU-rich sequence is only fivefold better than that with A-rich elements.³⁴ But in the presence of Hrp1p, Rna15p interacts with A-rich elements to facilitate poly(A) site selection¹³⁸ (Figure 5(a)). In the Hrp1p–Rna15p–RNA complex structure, Hrp1p interacts with the 5' AU repeats, whereas Rna15p binds to AAUAAU at the 3' of the RNA molecule, which illustrates the molecular basis for A-rich element recognition by Rna15p.⁵⁴ This structure also reveals a direct interaction between the RRM of Rna15p and RRM1 of Hrp1p.⁵⁴ Together with the RRM2 of Hrp1p, these three RRM domains are assembled in a 'horseshoe' configuration, in which the RNA-binding surface (β -sheet) of each RRM domain is rotated $\sim 90^\circ$ relative to the adjacent domain⁵⁴ (Figure 5(a)). This domain organization would permit the binding of a continuous RNA molecule by the complex, which

is in agreement with the observation that A-rich elements are generally located immediately downstream of AU repeats.⁶⁴ Taken together, the recognition of the upstream *cis*-elements is possibly a two-step reaction: Hrp1p specifically recognizes AU repeats and loads the low specificity Rna15p onto the A-rich elements. Rna14p may further enhance binding among Hrp1p, Rna15p, and *cis*-elements.^{122,138} In keeping with its proposed positioning role, a Hrp1p variant altered poly(A) site usage *in vivo*.¹³⁹

CONCLUSION

Two decades of biochemical and genetic studies have built a general mechanistic framework for mRNA 3' processing.^{1,4,6,8,10,11} Emerging structural information of individual subunits has provided invaluable building blocks for the gigantic 3' processing network (Table 1). Time and effort has also been devoted toward understanding the 3' processing machinery assembly through protein–protein complexes.^{1,39,54,85,118} Besides the protein factors, *cis*-elements on the 3' UTR region are also integral part of 3' processing and poly(A) site determination. With this review we have attempted to summarize the

current progress in *cis*-elements recognition with an emphasis on how structural data help further our knowledge of poly(A) site selection. As illustrated by the CFI_m-RNA structure, the CFI_m heterotetramer may be able to loop out an entire poly(A) site and utilize an alternative site.⁴⁷ It is plausible that the CstF heterohexamer may employ a similar looping mechanism to bind multiple U/GU-rich elements. Furthermore, the yeast Hrp1p–Rna15p–RNA complex provides a perfect example of how protein–protein interactions influence poly(A) site definition: A specific and tight interaction between Hrp1p and AU-rich element recruits Rna15p to a less preferred A-rich element.⁵⁴ In addition to obtaining the structures of individual subunits/factors, determining the high-order molecular architecture of 3' processing complex remains a major challenge. Great progress has been made toward purifying a functional 3' processing complex, which permits the structural analysis of the intact complex.⁹ While a structural picture of 3' processing complex is beginning to emerge, more structural information will be required for a complete understanding of complex assembly, poly(A) site selection, and cross-talk with other gene expression steps.

ACKNOWLEDGMENTS

The authors thank Dr. Gregory Gilmartin for helpful comments on this manuscript. Work on RNA processing in our laboratory was supported in part by a grant from the National Institutes of Health (GM62239 to S.D.)

REFERENCES

1. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* 2008, 65:1099–1122.
2. Proudfoot N. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* 2004, 16:272–278.
3. Edmonds M. A history of poly A sequences: from formation to factors to function. *Prog Nucleic Acid Res Mol Biol* 2002, 71:285–389.
4. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* 1999, 63:405–445.
5. Danckwardt S, Hentze MW, Kulozik AE. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J* 2008, 27:482–498.
6. Wahle E, Ruegsegger U. 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* 1999, 23:277–295.
7. Dominski Z, Marzluff WF. Formation of the 3' end of histone mRNA. *Gene* 1999, 239:1–14.
8. Proudfoot N, O'Sullivan J. Polyadenylation: a tail of two complexes. *Curr Biol* 2002, 12:R855–R857.
9. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL. Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 2009, 33:365–376.
10. Moore MJ, Proudfoot NJ. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 2009, 136:688–700.
11. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* 2010, 38:2757–2774.

12. Lutz CS. Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol* 2008, 3:609–617.
13. Hunt AG. Messenger RNA 3' end formation in plants. *Curr Top Microbiol Immunol* 2008, 326:151–177.
14. Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* 2009, 106:7028–7033.
15. Tian B, Hu J, Zhang H, Lutz CS. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* 2005, 33:201–212.
16. Shi Y, Chan S, Martinez-Santibanez G. An up-close look at the pre-mRNA 3'-end processing complex. *RNA Biol* 2009, 6:522–525.
17. Gilmartin GM. Eukaryotic mRNA 3' processing: a common means to different ends. *Genes Dev* 2005, 19:2517–2521.
18. Lutz CS, Moreira A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *WIREs RNA* 2011, 2:22–31.
19. Chan S, Choi E-A, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. *WIREs RNA* 2011, 3:321–335.
20. Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* 2004, 23:616–626.
21. Rügsegger U, Blank D, Keller W. Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell* 1998, 1:243–253.
22. Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev* 1997, 11:1703–1716.
23. Rügsegger U, Beyer K, Keller W. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* 1996, 271:6107–6113.
24. Kessler MM, Zhao J, Moore CL. Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends. *J Biol Chem* 1996, 271:27167–27175.
25. Murthy KG, Manley JL. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* 1995, 9:2672–2683.
26. Minvielle-Sebastia L, Preker PJ, Keller W. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science* 1994, 266:1702–1705.
27. MacDonald CC, Wilusz J, Shenk T. 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* 1994, 14:6647–6654.
28. Coseno M, Martin G, Berger C, Gilmartin G, Keller W, Doublé S. Crystal structure of the 25 kDa subunit of human cleavage factor Im. *Nucleic Acids Res* 2008, 36:3474–3483.
29. Tresaugues L, Stenmark P, Schuler H, Flodin S, Welin M, Nyman T, Hammarstrom M, Moche M, Graslund S, Nordlund P. The crystal structure of human cleavage and polyadenylation specific factor-5 reveals a dimeric Nudix protein with a conserved catalytic site. *Proteins* 2008, 73:1047–1052.
30. Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L. Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 2006, 444:953–956.
31. Nishida Y, Ishikawa H, Baba S, Nakagawa N, Kuramitsu S, Masui R. Crystal structure of an archaeal cleavage and polyadenylation specificity factor subunit from *Pyrococcus horikoshii*. *Proteins* 2010, 78:2395–2398.
32. Mir-Montazeri B, Ammelburg M, Forouzan D, Lupas AN, Hartmann MD. Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *J Struct Biol* 2011, 173:191–195.
33. Perez Canadillas JM, Varani G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. *EMBO J* 2003, 22:2821–2830.
34. Pancevac C, Goldstone DC, Ramos A, Taylor IA. Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3'-end processing factors. *Nucleic Acids Res* 2010, 38:3119–3132.
35. Qu X, Perez-Canadillas JM, Agrawal S, De Baecke J, Cheng H, Varani G, Moore C. The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing. *J Biol Chem* 2007, 282:2101–2115.
36. Legrand P, Pinaud N, Minvielle-Sebastia L, Fribourg S. The structure of the CstF-77 homodimer provides insights into CstF assembly. *Nucleic Acids Res* 2007, 35:4515–4522.
37. Bai Y, Auperin TC, Chou CY, Chang GG, Manley JL, Tong L. Crystal structure of murine CstF-77: dimeric association and implications for polyadenylation of mRNA precursors. *Mol Cell* 2007, 25:863–875.
38. Moreno-Morcillo M, Minvielle-Sebastia L, Mackereith C, Fribourg S. Hexameric architecture of CstF supported by CstF-50 homodimerization domain structure. *RNA* 2011, 17:412–418.
39. Kennedy SA, Frazier ML, Steiniger M, Mast AM, Marzluff WF, Redinbo MR. Crystal structure of the

- HEAT domain from the Pre-mRNA processing factor Symplekin. *J Mol Biol* 2009, 392:115–128.
40. Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* 2005, 12:144–151.
 41. Meinhart A, Cramer P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 2004, 430:223–226.
 42. Martin G, Keller W, Doublié S. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J* 2000, 19:4193–4203.
 43. Martin G, Moglich A, Keller W, Doublié S. Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase. *J Mol Biol* 2004, 341:911–925.
 44. Bard J, Zhelkovsky AM, Helmling S, Earnest TN, Moore CL, Bohm A. Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science* 2000, 289:1346–1349.
 45. Balbo PB, Toth J, Bohm A. X-ray crystallographic and steady state fluorescence characterization of the protein dynamics of yeast polyadenylate polymerase. *J Mol Biol* 2007, 366:1401–1415.
 46. Ge H, Zhou D, Tong S, Gao Y, Teng M, Niu L. Crystal structure and possible dimerization of the single RRM of human PABPN1. *Proteins* 2008, 71:1539–1545.
 47. Yang Q, Coseno M, Gilmartin GM, Doublié S. Crystal structure of a human cleavage factor CFIm25/CFIm68/RNA complex provides an insight into poly(A) site recognition and rna looping. *Structure* 2011, 19:368–377.
 48. Das K, Ma LC, Xiao R, Radvansky B, Aramini J, Zhao L, Marklund J, Kuo RL, Twu KY, Arnold E, et al. Structural basis for suppression of a host antiviral response by influenza A virus. *Proc Natl Acad Sci U S A* 2008, 105:13093–13098.
 49. Noble CG, Beuth B, Taylor IA. Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. *Nucleic Acids Res* 2007, 35:87–99.
 50. Meinke G, Ezeokonkwo C, Balbo P, Stafford W, Moore C, Bohm A. Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered protein. *Biochemistry* 2008, 47:6859–6869.
 51. Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* 2010, 467: 729–733.
 52. Yang Q, Gilmartin GM, Doublié S. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci U S A* 2010, 107:10062–10067.
 53. Perez-Canadillas JM. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J* 2006, 25:3167–3178.
 54. Leeper TC, Qu X, Lu C, Moore C, Varani G. Novel protein-protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1. *J Mol Biol* 2010, 401:334–349.
 55. Balbo PB, Bohm A. Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure* 2007, 15: 1117–1131.
 56. Deo RC, Bonanno JB, Sonenberg N, Burley SK. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* 1999, 98:835–845.
 57. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* 2005, 19:1315–1327.
 58. Proudfoot NJ, Brownlee GG. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* 1976, 263:211–214.
 59. Beaulieu E, Freier S, Wyatt JR, Claverie JM, Gautheret D. Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 2000, 10:1001–1010.
 60. Hu J, Lutz CS, Wilusz J, Tian B. Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* 2005, 11:1485–1493.
 61. Shen Y, Ji G, Haas BJ, Wu X, Zheng J, Reese GJ, Li QQ. Genome level analysis of rice mRNA 3'-end processing signals and alternative polyadenylation. *Nucleic Acids Res* 2008, 36:3150–3161.
 62. Fitzgerald M, Shenk T. The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. *Cell* 1981, 24:251–260.
 63. Salisbury J, Hutchison KW, Graber JH. A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. *BMC Genomics* 2006, 7:55.
 64. Graber JH, Cantor CR, Mohr SC, Smith TF. Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res* 1999, 27:888–894.
 65. Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL. Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev* 1997, 11:2545–2556.
 66. Dichtl B, Keller W. Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J* 2001, 20:3197–3209.
 67. Ruepp MD, Schümperli D, Barabino SML. mRNA 3' end processing and more—multiple functions of mammalian cleavage factor I-68. *WIREs RNA* 2011, 2:79–91.

68. Takagaki Y, Ryner LC, Manley JL. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* 1988, 52:731–742.
69. Brown KM, Gilmartin GM. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* 2003, 12:1467–1476.
70. Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SM. Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J Biol Chem* 2004, 279:35788–35797.
71. McLennan AG. The Nudix hydrolase superfamily. *Cell Mol Life Sci* 2006, 63:123–143.
72. Mildvan AS, Xia Z, Azurmendi HF, Saraswat V, Legler PM, Massiah MA, Gabelli SB, Bianchet MA, Kang LW, Amzel LM. Structures and mechanisms of Nudix hydrolases. *Arch Biochem Biophys* 2005, 433:129–143.
73. Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, Allain FH. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* 2006, 25:163–173.
74. Kim H, Lee Y. Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I. *Biochem Biophys Res Commun* 2001, 289:513–518.
75. Shimazu T, Horinouchi S, Yoshida M. Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3'-end processing. *J Biol Chem* 2007, 282:4470–4478.
76. Graveley BR. Sorting out the complexity of SR protein functions. *RNA* 2000, 6:1197–1211.
77. Millevoi S, Loulengue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J* 2006, 25:4854–4864.
78. Awasthi S, Alwine JC. Association of polyadenylation cleavage factor I with U1 snRNP. *RNA* 2003, 9:1400–1409.
79. Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* 2002, 12:1231–1245.
80. Zhou Z, Sim J, Griffith J, Reed R. Purification and electron microscopic visualization of functional human spliceosomes. *Proc Natl Acad Sci U S A* 2002, 99:12203–12207.
81. Clery A, Blatter M, Allain FH. RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* 2008, 18:290–298.
82. Kubo T, Wada T, Yamaguchi Y, Shimizu A, Handa H. Knock-down of 25 kDa subunit of cleavage factor Im in HeLa cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Res* 2006, 34:6264–6271.
83. Sartini BL, Wang H, Wang W, Millette CF, Kilpatrick DL. Pre-messenger RNA cleavage factor I (CFIm): potential role in alternative polyadenylation during spermatogenesis. *Biol Reprod* 2008, 78:472–482.
84. Ruepp MD, Aringhieri C, Vivarelli S, Cardinale S, Paro S, Schumperli D, Barabino SM. Mammalian pre-mRNA 3' end processing factor CF Im 68 functions in mRNA export. *Mol Biol Cell* 2009, 20:5211–5223.
85. Sullivan KD, Steiniger M, Marzluff WF. A core complex of CPSF73, CPSF100, and Symplekin may form two different cleavage factors for processing of poly(A) and histone mRNAs. *Mol Cell* 2009, 34:322–332.
86. Ryan K, Calvo O, Manley JL. Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* 2004, 10:565–573.
87. Callebaut I, Moshous D, Mornon JP, de Villartay JP. Metallo- β -lactamase fold within nucleic acids processing enzymes: the β -CASP family. *Nucleic Acids Res* 2002, 30:3592–3601.
88. Neuwald AF, Liu JS, Lipman DJ, Lawrence CE. Extracting protein alignment models from the sequence database. *Nucleic Acids Res* 1997, 25:1665–1677.
89. Dominski Z. Nucleases of the metallo- β -lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol* 2007, 42:67–93.
90. Ishikawa H, Nakagawa N, Kuramitsu S, Masui R. Crystal structure of TTHA0252 from *Thermus thermophilus* HB8, a RNA degradation protein of the metallo- β -lactamase superfamily. *J Biochem* 2006, 140:535–542.
91. de la Sierra-Gallay IL, Zig L, Jamalli A, Putzer H. Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol* 2008, 15:206–212.
92. Bebrone C. Metallo- β -lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. *Biochem Pharmacol* 2007, 74:1686–1701.
93. Kolev NG, Yario TA, Benson E, Steitz JA. Conserved motifs in both CPSF73 and CPSF100 are required to assemble the active endonuclease for histone mRNA 3'-end maturation. *EMBO Rep* 2008, 9:1013–1018.
94. Auweter SD, Oberstrass FC, Allain FH. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* 2006, 34:4943–4959.
95. Castro C, Smidansky ED, Arnold JJ, Maksimchuk KR, Moustafa I, Uchida A, Gotte M, Konigsberg W, Cameron CE. Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol* 2009, 16:212–218.
96. Kyburz A, Sadowski M, Dichtl B, Keller W. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic Acids Res* 2003, 31:3936–3945.
97. Xu R, Zhao H, Dinkins RD, Cheng X, Carberry G, Li QQ. The 73 kD subunit of the cleavage and

- polyadenylation specificity factor (CPSF) complex affects reproductive development in Arabidopsis. *Plant Mol Biol* 2006, 61:799–815.
98. Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF. A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol* 2005, 25:1489–1500.
99. Yang XC, Sullivan KD, Marzluff WF, Dominski Z. Studies of the 5' exonuclease and endonuclease activities of CPSF-73 in histone pre-mRNA processing. *Mol Cell Biol* 2009, 29:31–42.
100. Moore CL, Chen J, Whoriskey J. Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. *EMBO J* 1988, 7:3159–3169.
101. Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* 2002, 21:4125–4135.
102. Barabino SM, Ohnacker M, Keller W. Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J* 2000, 19:3778–3787.
103. Chen Z, Li Y, Krug RM. Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *EMBO J* 1999, 18:2273–2283.
104. Wilusz J, Shenk T, Takagaki Y, Manley JL. A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol Cell Biol* 1990, 10:1244–1248.
105. Takagaki Y, Manley JL. A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. *Nature* 1994, 372:471–474.
106. Wahle E, Lustig A, Jenö P, Maurer P. Mammalian poly(A)-binding protein II. Physical properties and binding to polynucleotides. *J Biol Chem* 1993, 268:2937–2945.
107. Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev* 1998, 12:2522–2534.
108. Takagaki Y, Manley JL. Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* 2000, 20:1515–1525.
109. Wilusz J, Shenk T. A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. *Cell* 1988, 52:221–228.
110. Gilmartin GM, Nevins JR. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. *Mol Cell Biol* 1991, 11:2432–2438.
111. Deka P, Rajan PK, Perez-Canadillas JM, Varani G. Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein. *J Mol Biol* 2005, 347:719–733.
112. Beyer K, Dandekar T, Keller W. RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J Biol Chem* 1997, 272:26769–26779.
113. Takagaki Y, Manley JL. RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* 1997, 17:3907–3914.
114. Vander Kooi CW, Ren L, Xu P, Ohi MD, Gould KL, Chazin WJ. The Prp19 WD40 domain contains a conserved protein interaction region essential for its function. *Structure* 2010, 18:584–593.
115. Rose R, Weyand M, Lammers M, Ishizaki T, Ahmadian MR, Wittinghofer A. Structural and mechanistic insights into the interaction between Rho and mammalian Dia. *Nature* 2005, 435:513–518.
116. Ye K, Patel DJ. RNA silencing suppressor p21 of Beet yellows virus forms an RNA binding octameric ring structure. *Structure* 2005, 13:1375–1384.
117. Hatton LS, Eloranta JJ, Figueiredo LM, Takagaki Y, Manley JL, O'Hare K. The Drosophila homologue of the 64 kDa subunit of cleavage stimulation factor interacts with the 77 kDa subunit encoded by the suppressor of forked gene. *Nucleic Acids Res* 2000, 28:520–526.
118. Hockert JA, Yeh HJ, MacDonald CC. The hinge domain of the cleavage stimulation factor protein CstF-64 is essential for CstF-77 interaction, nuclear localization, and polyadenylation. *J Biol Chem* 2010, 285:695–704.
119. Lamb JR, Tugendreich S, Hieter P. Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* 1995, 20:257–259.
120. Preker PJ, Keller W. The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem Sci* 1998, 23:15–16.
121. Gatto GJ Jr, Geisbrecht BV, Gould SJ, Berg JM. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nat Struct Biol* 2000, 7:1091–1095.
122. Noble CG, Walker PA, Calder LJ, Taylor IA. Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic Acids Res* 2004, 32:3364–3375.
123. Bell SA, Hunt AG. The Arabidopsis ortholog of the 77 kDa subunit of the cleavage stimulatory factor (AtCstF-77) involved in mRNA polyadenylation is an RNA-binding protein. *FEBS Lett* 2010, 584:1449–1454.

124. Takagaki Y, Manley JL. A human polyadenylation factor is a G protein β -subunit homologue. *J Biol Chem* 1992, 267:23471–23474.
125. Li D, Roberts R. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. *Cell Mol Life Sci* 2001, 58:2085–2097.
126. Smith TF, Gaitatzes C, Saxena K, Neer EJ. The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* 1999, 24:181–185.
127. Couture JF, Collazo E, Trievel RC. Molecular recognition of histone H3 by the WD40 protein WDR5. *Nat Struct Mol Biol* 2006, 13:698–703.
128. Kleiman FE, Manley JL. The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell* 2001, 104:743–753.
129. Kleiman FE, Manley JL. Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science* 1999, 285:1576–1579.
130. Edwards RA, Lee MS, Tsutakawa SE, Williams RS, Nazeer I, Kleiman FE, Tainer JA, Glover JN. The BARD1 C-terminal domain structure and interactions with polyadenylation factor CstF-50. *Biochemistry* 2008, 47:11446–11456.
131. Scully R, Xie A, Nagaraju G. Molecular functions of BRCA1 in the DNA damage response. *Cancer Biol Ther* 2004, 3:521–527.
132. Ohnacker M, Barabino SM, Preker PJ, Keller W. The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. *EMBO J* 2000, 19:37–47.
133. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 1997, 385:357–361.
134. Minvielle-Sebastia L, Beyer K, Krecic AM, Hector RE, Swanson MS, Keller W. Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO J* 1998, 17:7454–7468.
135. Gross S, Moore C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I. *Proc Natl Acad Sci U S A* 2001, 98:6080–6085.
136. Valentini SR, Weiss VH, Silver PA. Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. *RNA* 1999, 5:272–280.
137. Chen S, Hyman LE. A specific RNA-protein interaction at yeast polyadenylation efficiency elements. *Nucleic Acids Res* 1998, 26:4965–4974.
138. Gross S, Moore CL. Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. *Mol Cell Biol* 2001, 21:8045–8055.
139. Kim Guisbert KS, Li H, Guthrie C. Alternative 3' pre-mRNA processing in *Saccharomyces cerevisiae* is modulated by Nab4/Hrp1 in vivo. *PLoS Biol* 2007, 5:e6.

FURTHER READING

Dominski Z. The hunt for the 3' endonuclease. *WIREs RNA* 2010, 1:325–340.