Cellular and Molecular Life Sciences

Review

Protein factors in pre-mRNA 3′-end processing

C. R. Mandel, Y. Bai and L. Tong*

Department of Biological Sciences, Columbia University, New York, NY 10027 (USA), Fax: +1-212-865-8246,
e-mail: ltong@columbia.edu

Abstract. Most eukaryotic mRNA precursors (pre-mRNAs) must undergo extensive processing, including cleavage and polyadenylation at the 3′-end. Processing at the 3′-end is controlled by sequence elements in the pre-mRNA (*cis* elements) as well as protein factors. Despite the seeming biochemical simplicity of the processing reactions, more than 14 proteins have been identified for the mammalian complex, and more than 20 proteins have been identified for the yeast complex. The 3′-end processing machinery also has important roles in transcription and splicing. The mammalian machinery contains several sub-complexes, including cleavage and polyadenylation specificity factor, cleavage stimulation factor, cleavage factor I, and cleavage factor II . Additional protein factors include poly(A) polymerase, poly(A)-binding protein, symplekin, and the C-terminal domain of RNA polymerase II largest subunit. The yeast machinery includes cleavage factor IA, cleavage factor IB, and cleavage and polyadenylation factor.

Introduction

In eukaryotes, messenger RNA precursors (pre-mRNAs) are transcribed in the nucleus from the genomic DNA by RNA polymerase II (polII). These pre-mRNAs must undergo extensive co-transcriptional processing before they can be transported to the cytoplasm for translation into proteins. The processing events include: capping, where the guanosine at the 5′-end of the pre-mRNA is methylated at the 7 carbon; splicing to remove the intronic sequences of the pre-mRNA; and cleavage and polyadenylation at the 3′-end [1–4]. All these modifications are essential for the maturation of the pre-mRNAs. This review focuses on the processing events at the 3′-end.

A poly(A) polymerase (PAP) activity was first isolated from calf thymus in 1960 [5], and its role in pre-mRNA 3′-end processing was recognized 10 years later [6–8]. Subsequent studies showed that the primary RNA transcripts extended beyond the site of polyadenylation, indicating that eukaryotic pre-mRNA 3′-end processing requires both cleavage and polyadenylation [9–11]. These observations are in sharp contrast to those in bacteria, where the 3′-ends of mRNAs are formed by transcriptional termination. Further studies showed that a large collection of protein factors (with more than 14 molecules in mammals and more than 20 molecules in yeast, totaling over 1 mega-dalton) are required for the 3′-end processing, despite the apparent simplicity of the cleavage and polyadenylation reactions. This machinery is directed to the correct cleavage site by sequence elements within the pre-mRNA 3′-end.

* Corresponding author.

This review attempts to summarize the current knowledge on pre-mRNA 3′-end processing in mammals and yeast, two model systems in which these modifications have been studied in the greatest detail. Special emphasis will be placed on recent developments in this area, including structural information on the protein factors that have become available in the past few years. There are also many excellent earlier reviews on pre-mRNA 3′-end processing [1–4, 12–23].

The importance of pre-mRNA 3′-end processing

The processing of pre-mRNA 3′-ends has crucial functional importance in eukaryotes, and disruption of this processing has catastrophic effects on cell growth and viability [2]. First, 3′-end processing promotes the transport of mRNAs from the nucleus to the cytoplasm [18]. Substitution of the pre-mRNA 3′-end polyadenylation site with a ribosomal RNA cleavage site produces an mRNA that is cleaved but not polyadenylated. This substitution decreased the ratio of cytoplasmic to nuclear mRNA concentration by tenfold, indicating a reduction in mRNA transport and consequently a reduction in protein expression [24].

Secondly, 3′-end processing promotes the stability of mRNAs [20, 22]. In the cytoplasm, mRNAs are degraded from the 3′-end first, indicating the importance of protecting the 3′-end [20]. The addition of the poly(A) tail and subsequently the binding of poly(A)-binding protein (PABP) has been shown to prevent degradation in mammalian cells [25] and *Xenopus* oocytes [26]. In fact, just the presence of PABP without a poly(A) tail may be sufficient to prevent mRNA degradation [27].

Thirdly, 3′-end processing enhances the translation of mRNAs into proteins. The poly(A) tail and PABP interact with the methyl cap at the 5′-end to promote translation [20, 22, 28, 29]. Studies in yeast have shown that the presence of the poly(A) tail alone is sufficient to initiate efficient translation, but the presence of both the poly(A) tail and the 5′-cap is optimal for translation [30].

Finally, 3′-end processing is intricately coupled to the transcription and splicing machineries [13, 16, 17, 23, 31]. The 3′-end processing complex interacts with transcription factors and the C-terminal domain (CTD) of polII to help control transcriptional initiation, and a proper poly(A) signal is essential for transcriptional termination. Alterations in these interactions lead to improper polyadenylation and mRNA degradation [4].

Recent studies show that polyadenylation is not restricted to the nucleus. PAP enzymes have been identified in the cytoplasm and in the mitochondria [32–34]. These enzymes are found from yeast to humans, and have important functions in many cellular processes.

Sequence elements in the pre-mRNA that direct 3′-end processing

The 3′-end cleavage and polyadenylation reaction is directed by sequence elements within the untranslated region of the pre-mRNA (the so-called *cis* elements). These sequence elements are found in almost every eukaryotic pre-mRNA that is polyadenylated, but they are not found in histone pre-mRNAs that are cleaved but not polyadenylated [15]. Disruption of the position and sequence of these *cis* elements reduces the efficiency of 3′-end processing, consistent with their conservation in pre-mRNAs. Recent studies analyzing mRNAs and cDNAs indicate that polyadenylation can occur at multiple sites for over half the genes in humans ($\sim 54\%$) and for about one-third of the genes ($\sim 32\%$) in mice [35, 36]. These alternative polyadenylation sites may occur in the same exon or in an alternatively spliced 3′ exon. Although polyadenylation can occur at multiple sites, those containing the optimal sequence elements are cleaved more efficiently.

The sequence elements in yeast and mammals share some recognizable similarity, but also have significant differences (Fig. 1A, B). The yeast sequence elements differ in their sequence and location from mammalian sequence elements (Fig. 1B). Additionally, yeast sequence elements are also less conserved than their mammalian counterparts. Some of the sequence elements in yeast appear in duplicate, while mammalian sequence elements generally occur only once (Fig. 1A). Another difference is the presence of U-rich elements that flank the cleavage site in yeast. As a result, we discuss the mammalian and yeast sequence elements separately below.

Sequence elements in mammals

Mammalian pre-mRNAs contain three primary sequence elements that define the polyadenylation site and two auxiliary sequence elements that enhance/regulate the 3′-end processing reaction (Fig. 1A). The three primary sequence elements consist of the hexamer AAUAAA polyadenylation signal (PAS), the cleavage site, and the G/U-rich downstream element (DSE) [2]. The two auxiliary sequence elements consist of an upstream element and a DSE [2].
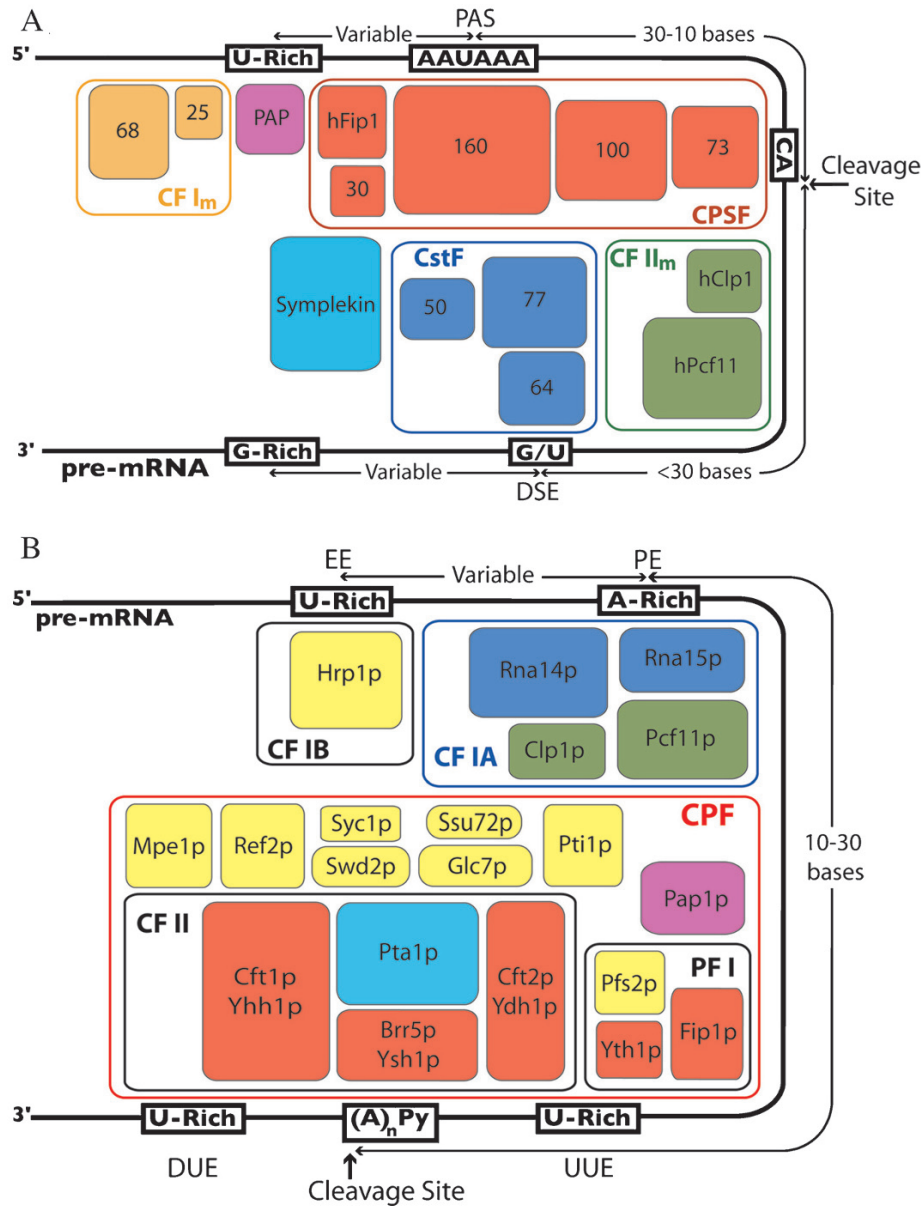
Figure 1. Pre-mRNA 3′-end processing machinery. (*A*) Schematic drawing of the pre-mRNA 3′-end processing complex in mammals. The *cis* elements in the pre-mRNA are also indicated. Cleavage and polyadenylation specificity factor (CPSF)-160 recognizes the polyadenylation signal (PAS), whereas cleavage stimulation factor (CstF)-64 recognizes the downstream element (DSE). (*B*) Schematic drawing of the pre-mRNA 3′-end processing complex in yeast.

The PAS. The PAS was the first identified sequence element proposed to play a role in 3′-end processing [37]. This AAUAAA hexamer was found in 80–90 % of sequenced mRNAs. Variation of the second position to a U, creating the AUUAAA hexamer, is present in about 10 % of the mRNAs [38]. A study of over 4000 human expressed sequence tags (ESTs) found that 77 % contained a PAS, and of these 75 % contained the AAUAAA sequence and 20 % contained the AUUAAA sequence [36]. A more recent study of over 13 000 human and mice ESTs showed that the number of ESTs that did not contain a PAS is only about 4 % [35]. AAUAAA (70 %) and AUUAAA (15 %) are the most common hexamers in the PAS.

Mutational studies have confirmed the necessity of the PAS for proper 3′-end processing. Patients suffering from α- and β-thalassemia carry a point mutation in the final base of the motif (A to G) and have low levels of mRNA [39]. Additionally, four separate point mutations (AACAAA, AAUUAA, AAUACA and AAUGAA) in the PAS produced significantly reduced levels of polyadenylated RNA and increased amounts of unprocessed pre-mRNA in *Xenopus laevis* oocytes [40]. Interestingly, two of the tested mutations are found in naturally occurring pre-mRNAs, yet they failed to produce polyadenylated mRNA.

Not only is the PAS sequence conserved, its distance from the poly(A) site is also conserved, which ranges from 10 to 30 bases downstream of the PAS (Fig. 1A).

Deleting parts of the region between the PAS and the cleavage site produced a new cleavage site further downstream that corresponded to the length of the deleted segment [41].

The DSE. The DSE is less conserved and more diffuse than the PAS [2, 19]. The presence of this element was suggested by the observation that a deletion of more than 10 nucleotides downstream of the cleavage site reduced polyadenylation threefold [42]. The DSE has been observed in two forms, a GU-rich element that has a sequence of YGUGUUYY (Y = pyrimidine) [43, 44] and a U-rich sequence (UUUUU) [43, 45], although pre-mRNAs can have neither, one, or both of these sequences [46]. Point mutations within the DSE have a smaller effect on cleavage activity, whereas deletions of short segments of the DSE have more significant impacts [47, 48]. While the DSE is fairly tolerant of sequence abnormalities, it is less tolerant of positioning abnormalities. The DSE is generally located within 30 nucleotides of the cleavage site (Fig. 1A), although in a few instances DSEs have been observed further downstream [49].

The cleavage site. The cleavage site of pre-mRNAs is positioned between the PAS and the DSE, generally within a region of 13 nucleotides [50]. The nucleotide sequence surrounding the cleavage site is not strictly conserved [51, 52]. In vertebrate pre-mRNAs, 70 % of the cleavage sites are located at the 3′ side of an adenosine residue [53], with a nucleotide preference of A > U > C >> G. The nucleotide preceding the cleavage site is cytosine in 59 % of 269 pre-mRNA sequences examined [53], making the optimal cleavage site CA (Fig. 1A).

The auxiliary upstream element. This element, located upstream of the PAS (Fig. 1A), does not have a consensus sequence, but often consists of a U-rich element (UUUU) [2] or similar sequences (UGUA, UAUA) [54]. The efficiency of cleavage and polyadenylation is enhanced by the presence of this auxiliary element, as it promotes the binding of other polyadenylation factors to the cleavage site [55–58]. Most auxiliary upstream elements have been identified with the enhanced expression of intronless genes, which are normally expressed at lower levels than transcripts that contain introns [59].

The auxiliary DSE. While many examples of auxiliary upstream elements are known, fewer auxiliary elements have been identified downstream of the cleavage site. These auxiliary elements are generally G-rich, but they lack a conserved sequence and distance from the cleavage site. In addition, more than one auxiliary sequences can be present in a gene [2, 60–64].

Sequence elements in yeast

The yeast poly(A) site is defined by four sequence elements: the AU-rich efficiency element (EE), the A-rich positioning element (PE), the cleavage site, and the U-rich elements flanking the cleavage site (Fig. 1B).

The EE. The AU-rich EE is located furthest upstream but at a variable distance from the cleavage site [2]. The sequence UAUAUA provided the greatest effect on 3′-end processing with the U at the first and fifth positions being the most critical for function [65]. A large-scale analysis of the 3′ untranslated region of 1017 yeast nuclear transcripts showed that more than half of them (52 %) contained the optimal EE sequence (UAUAUA) [66]. While the EE improves the efficiency, it is not required for cleavage.

The PE. The A-rich PE is located downstream of the EE and approximately 10–30 nucleotides upstream of the cleavage site [2]. The position and sequence of the PE are critical for efficient 3′-end processing. Although many A-rich sequences for the PE have been identified, the most effective sequences are AAUAAA or AAAAAA [67]. The PE received its name because most mutations in the PE disrupt the position but not the efficiency of 3′-end processing [68], although a single-point mutations in the PE of the TRP4 gene decreased the efficiency of 3′-end processing [69]. Deletion of the entire PE can also decrease the efficiency of 3′-end processing [2]. The PE is similar to its mammalian counterpart, the PAS, but serves a less critical role in 3′-end processing.

The cleavage site. The cleavage site is generally defined by a sequence element containing a pyrimidine followed by multiple adenosines $Y(A)_n$ [2], with the cleavage occurring at the 3′ side of an adenosine, similar to the reaction in mammals [70, 71]. Mutation of the PE forces cleavage to occur at unspecific locations in the 3′ untranslated region [19].

The upstream and downstream U-rich elements (UUE and DUE). Surprisingly, deletion of the PE and EE in cytochrome c pre-mRNA (CYC-1) does not abolish cleavage activity, signifying the importance of other sequences near the cleavage site [72]. Sequence analysis revealed conserved U-rich sequences just upstream and downstream of the cleavage site [73]. Mutations in either UUE or DUE reduce cleavage activity but do not abolish it. If these mutations are made in combination with mutations to the cleavage

site, cleavage is drastically reduced, indicating a synergistic role of the U-rich sequences with the cleavage site [72]. These U-rich sequences are also found in plants, but are absent in mammals [74].

The 3′-end processing machinery and its sub-complexes

Only two enzymatic activities are required for pre-mRNA 3′-end processing, cleavage and polyadenylation. However, studies with mammalian nuclear extracts [75] and yeast extracts [76], as well as studies by other biochemical and genetic methods, suggest that a large number of protein factors are required for 3′-end processing. Recent studies from (tandem) affinity purification of protein complexes in yeast, followed by identification of the components by mass spectrometry, confirmed the known protein factors in 3′-end processing as well as identifying many new ones [77–82]. Currently, more than 14 proteins have been identified for the mammalian 3′-end processing machinery (Fig. 1A), and more than 20 proteins appear associated with the yeast 3′-end processing machinery (Fig. 1B). Therefore, a mega-dalton complex is required for eukaryotic pre-mRNA 3′-end processing, despite the seeming biochemical simplicity of the reaction. While many of the protein factors in this machinery regulate pre-mRNA 3′-end processing, other factors in this machinery mediate transcriptional initiation, transcriptional termination, splicing and other events.

The mammalian 3′-end processing complex contains several sub-complexes (Fig. 1A), including cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I (CF I$_m$), and cleavage factor II (CF II$_m$). In addition, PAP, PABP, symplekin, and PolII CTD also belong to this machinery. All the protein factors except PABP are required for *in vitro* cleavage reaction, while only CPSF, PAP and PABP are required for *in vitro* polyadenylation [2].

The yeast 3′-end processing complex is different from the mammalian complex, but also shares significant similarities [12]. The sub-complexes of the yeast machinery include the cleavage and polyadenylation factor (CPF), cleavage factor IA (CF IA), and cleavage factor IB (CF IB) (Fig. 1B). CPF can be further separated into cleavage factor II (CF II) and polyadenylation factor I (PF I) [83]. CF II contains subunits that are homologous to those in mammalian CPSF, except that the homologs of mammalian CPSF-30 and hFip1, Yth1p and Fip1p, actually belongs to PF I rather than CF II (Fig. 1B). CF IA contains subunits that are homologous to those in mammalian CstF and

CF II$_m$, except that mammalian CstF-50 does not appear to have a homolog in yeast. The functions of some of these sequence homologs can be different in yeast and in mammals. For example, mammalian CPSF-160 recognizes the AAUAAA PAS, whereas its homolog in yeast (Cft1p/Yhh1p) is mostly associated with the cleavage site. Similarly, while mammalian CstF-64 recognizes the G/U-rich DSE, yeast Rna15p recognizes the A-rich PE.

The yeast machinery also contains many additional protein factors in the CPF (Fig. 1B). Even though some of them have mammalian homologs, the functional roles of these homologs in pre-mRNA 3′-end processing have yet to be established. Even in yeast, reconstitution of the cleavage reaction *in vitro* requires only CF IA, CF IB, and CF II, while *in vitro* polyadenylation requires CPF, CF IA, CF IB, and Pab1p. The additional factors play generally a modulatory role.

Despite the identification of this large collection of protein factors, the identity of the nuclease that actually catalyzes the cleavage reaction was not known. Recent studies have provided strong structural and biochemical evidence that CPSF-73 is the endoribonuclease for pre-mRNA 3′-end processing.

The protein factors involved in pre-mRNA 3′-end processing are described in more detail in the following sections, organized based on the mammalian machinery. For those mammalian proteins that have yeast homologs, the names of the yeast proteins are given in the parentheses. The function and sequence characteristics of protein factors that are characterized in both yeast and mammals are summarized in Table 1.

Cleavage and polyadenylation specificity factor

Mammalian CPSF contains five subunits, CPSF-30, CPSF-73, CPSF-100, CPSF-160 and hFip1 (Table 1). They are all required for efficient cleavage and polyadenylation of pre-mRNAs, and they all have clear sequence homologs in the yeast 3′-end processing machinery.

CPSF-160 (Cft1p/Yhh1p). CPSF-160 is the largest subunit of CPSF and is one of the more studied proteins of the pre-mRNA 3′-end processing complex. CPSF-160 is conserved throughout eukaryotes (24 % identity and 51 % similarity between CPSF-160 and Cft1p/Yhh1p). It is an essential factor in yeast and mammals; it interacts directly with pre-mRNA to direct the cleavage reaction, and is also required for polyadenylation. Besides its role in 3′-end processing, CPSF-160 associates with factors involved in transcriptional initiation (TFIID) [84] and elongation (PolII CTD) [85], and plays a role in transcriptional

Table 1. Protein factors in pre-mRNA 3′-end processing in yeast and mammals.

Mammalian factor (yeast factor)	Mammalian protein (yeast homolog)	Mammalian protein partners (yeast protein partners)	Protein function and characteristics
CPSF (CPF)	CPSF-160 (Cft1p/Yhh1p)	CstF-77, Pol II CTD, PAP, hFip1, TFIID (Fip1p, Pol II CTD, Cft2p/Ydh1p and Swd2p)	Binds RNA (mammalian AAUAAA site, yeast cleavage site), three possible β-propellers
CPSF (CPF)	CPSF-100 (Cft2p/Ydh1p)	CPSF-73 (Pol II CTD, Brr5p/Ysh1p, Cft1p/Yhh1p, Pta1p, Pfs2p, Ssu72p, Pcf11p, YDL094C)	Binds RNA at the U-rich sequence in yeast, may act as a scaffolding protein, contains a non-metal binding metallo-β-lactamase domain
CPSF (CPF)	CPSF-73 (Brr5p/Ysh1p)	CPSF-100 (Pta1p, Cft2p/Ydh1p, Yth1p, Clp1p, Syc1p, Swd2p)	Endonuclease for the cleavage reaction, contains a metal binding metallo-β-lactamase domain
CPSF (CPF)	CPSF-30 (Yth1p)	hFip1 (Fip1p, Brr5p/Ysh1p)	Binds U-rich RNA sequences, contains five zinc fingers and a zinc knuckle (mammalian)
CPSF (CPF)	hFip1 (Fip1p)	PAP, CPSF-160, CPSF-30, CstF-77 (Pap1p, Yth1p, Rna14p)	Directs PAP to correct polyadenylation site and binds U-rich sequences, contains a conserved region that binds CPSF-30
CstF (CF IA)	CstF-77 (Rna14p)	CPSF-160, CstF64, CstF50, Pol II CTD (Rna15p, Pcf11p)	Scaffolding protein, links CstF and CPSF, contains a HAT domain and a Pro-rich segment
CstF (CF IA)	CstF-64 (Rna15p)	CstF-77, Symplekin (Pcf11p, Rna14p)	Binds to RNA through an N-terminal RRM (G/U rich sequence in mammals, A-rich sequence in yeast)
CstF	CstF-50 (None)	CstF-77, Pol II CTD, BARD1	Binds Pol II CTD, contains seven WD-40 repeats
CF I$_m$	CF I$_m$-25kD (None)	PAP, CF I$_m$-68kD	CF I$_m$ aids AAUAAA site recognition, contains a conserved PAP interaction domain
CF I$_m$	CF I$_m$-68kD (None)	CF I$_m$-25kD	CF I$_m$ aids AAUAAA site recognition, contains an N-terminal RBD
CF II$_m$ (CF IA)	hPcf11 (Pcf11p)	(Pol II CTD, Rna14p, Rna15p, Clp1p, Swd2p, Cft1p/Yhh1p, Cft2p/Ydh1p, Brr5p/Ysh1p, Pta1p)	Interacts with Pol II CTD, acts as a scaffolding protein, contains two zinc finger domains
CF II$_m$ (CF IA)	hClp1 (Clp1p)	CF I$_m$ and CPSF (Brr5p/Ysh1p, Cft2p/Ydh1p, Pcf11p)	Tethers CPSF with CF I$_m$, contains Walker A and B motifs for ATP binding
– (CPF)	Symplekin (Pta1p)	CstF-64, CPSF (Brr5p/Ysh1p, Syc1p, Cft2p/Ydh1p, Swd2p, Glc7p, Pti1p, Ssu72p)	Bridges interactions between CPSF and CstF, regulated by phosphorylation
–	PAP (Pap1p)	hFip1, CF I$_m$, CPSF-160 (Fip1p)	Adds poly(A) tail to cleaved pre-mRNA
–	Pol II CTD (Pol II CTD)	hPcf11, CPSF-160, CstF-50, CstF-77 (Cft2p/Ydh1p, Cft1p/Yhh1p, Pcf11p)	Transcriptional termination, contains YSPTSPS repeats
–	PABII (Pab1p)	(Rna15p)	Stimulates polyadenylation and stabilizes poly(A) tail, contains two RRM domains

termination [4]. CPSF-160 is also involved in cytoplasmic polyadenylation in *Xenopus* oocytes [32]. Mammalian CPSF-160 binds directly to the PAS (Fig. 1A), with high affinity for the perfect PAS sequence (AAUAAA) and lower affinity for other PAS sequences [2]. UV cross-linking, nuclease protection and other experiments demonstrated this direct binding, using both nuclear extracts and purified recombinant protein [86–88]. Deletion of nucleotides beyond the cleavage site reduces the ability of CPSF-160 to bind to the PAS. Mutations of the PAS abolished the cross-linking of nuclear extract [89], while they only have a small effect on the cross-linking of the recombinant CPSF-160 [88], suggesting that

other proteins are required for the highly specific binding to the PAS by CPSF-160. CPSF-160 can also bind at a site other than the PAS in HIV pre-mRNA [90].
Efficient binding of CPSF-160 to the PAS also depends on the CstF. CPSF-160 was cross-linked to the PAS along with CstF-64 and the C protein of heterogeneous ribonucleoprotein particles (hnRNPs). The interaction with CstF is likely mediated by CstF-77 (see section on CstF-77 below). CPSF-160 interacts with other subunits of CPSF, including CPSF-100 and hFip1, as well as PAP [88, 91]. Cleavage and polyadenylation are abolished if yeast Cft1p/Yhh1p is depleted from cell extracts by its
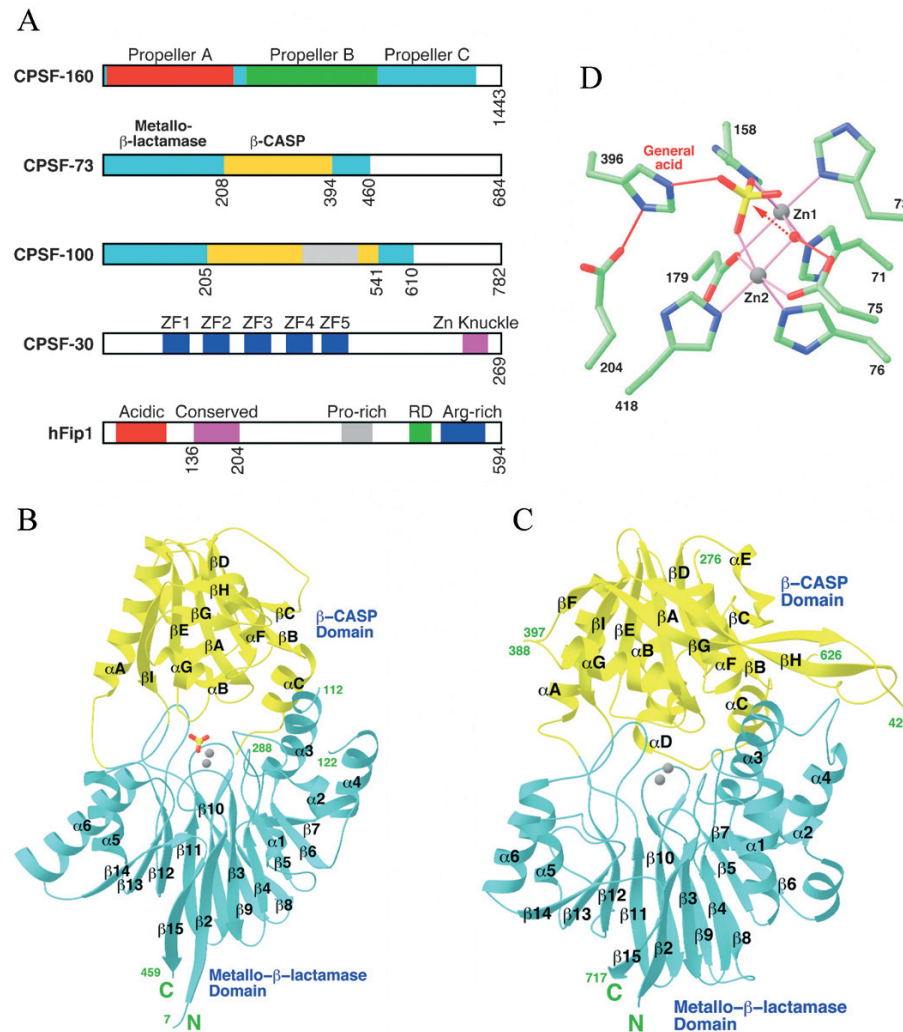
Figure 2. Cleavage and polyadenylation specificity factor (CPSF). (*A*) Schematic drawing of the domain organization of the five CPSF subunits in humans. (*B*) Crystal structure of the metallo-β-lactamase and β-CASP domains (residues 1–460) of human CPSF-73 [112]. (*C*) Schematic drawing of the coordination of two metal ions in the active site of CPSF-73. The bridging hydroxide is the general base, and His396 is the general acid for the ribonuclease reaction. (*D*) Crystal structure of the metallo-β-lactamase and β-CASP domains (residues 1–720) of yeast Cft2p/Ydh1p [112]. The zinc ions observed in the structure of CPSF-73 are shown for reference.

antibody [92]. Cft1p/Yhh1p binds directly to RNA [93], but not at the PE, the equivalent of the mammalian PAS. Instead, binding was observed primarily near the A-rich cleavage site (Fig. 1B), although the exact site is still not known [93].

Human CPSF-160 contains 1443 amino acid residues (Fig. 2A). There are two RNP-type binding motifs near the N terminus (RNP1, residues 379–386 and RNP2, residues 344–349), but it is not known whether these residues directly interact with RNA [88]. In yeast, deletion experiments showed that residues 500–750 of Cft1p/Yhh1p are required for RNA binding *in vitro* [93]. Sequence analysis of this region predicts that it contains five β-propeller repeats [93]. This region of the sequence is highly similar to the spliceosomal U2 snRNP (human SAP130, yeast Rse1p) [94] and the UV damage recognition protein *Xeroderma pigmentosum* group E (XPE) [95]. Sixteen β-propeller repeats were identified in human CPSF-160 based on sequence analysis [96].

CPSF-160 also bears weak sequence homology with the ubiquitin ligase DDB1, for which structural information has been obtained [97, 98]. DDB1 contains three β-propellers (each with seven repeats) and a C-terminal domain. CPSF-160 probably forms a similar structure, although it contains several large insertions in the β-propellers that do not have counterparts in DDB1. The RNA-binding region of Cft1p/Yhh1p would be located in the second β-propeller.

CPSF-73 (Brr5p/Ysh1p). Gene disruption of Brr5p/Ysh1p reveals that CPSF-73 is required for cell viability [99]. It was recognized recently that CPSF-73 contains a metallo-β-lactamase domain at its N terminus, and it is a founding member of the β-CASP superfamily of proteins (Fig. 2A) [100–102]. The name β-CASP was derived from the metallo-β-lactamase domain they all contain and after the members <u>C</u>PSF-73 and CPSF-100, <u>A</u>rtemis, <u>S</u>NM1 and <u>P</u>SO2. The highest amino acid sequence conser-

vation between CPSF-73 and Brr5p/Ysh1p is for their metallo-β-lactamase domains [103].

Most β-CASP proteins are involved in nucleic acid binding and/or processing. Several members of the β-CASP family are known to be nucleases [102]. Artemis, when complexed with DNA-dependent protein kinase, functions as an endonuclease on ssDNA overhangs [104, 105]. Bacterial RNase J1 has 5'-to-3' exoribonuclease activity on ribosomal RNA precursors [106]. RNase Z, which contains only the metallo-β-lactamase domain (and therefore is not a β-CASP family member), is the ribonuclease essential for the 3'-end processing of tRNA precursors [107, 108]. These observations suggest that CPSF-73 could have nuclease activity, and could actually be the endoribonuclease for the cleavage step of pre-mRNA 3'-end processing.

UV cross-linking of nuclear extracts revealed that CPSF-73 binds directly to the cleavage site in an AAUAAA-dependent manner [109], suggesting that it is at the correct position to carry out the cleavage reaction. In addition, CPSF-73 is cross-linked in a U7-dependent manner to the histone pre-mRNA cleavage site, and it may be the endoribonuclease involved in histone pre-mRNA 3'-end processing [110].

Most β-CASP proteins contain conserved residues that can bind two metal ions (zinc, iron or others). Disruption of these putative metal binding residues results in 3'-end processing defects and cell death [109, 111]. Dialysis of the nuclear extract leads to partial loss of cleavage activity, which can be rescued by the addition of $ZnCl_2$. Moreover, cleavage reaction using nuclear extract is inhibited by zinc-specific chelators N,N,N',N'-tetrakis(2-pyridylmethyl)ethylenediamine (TPEN) and ortho-phenanthroline (OP), as well as high concentrations the metal chelator EDTA [109]. These results suggest that the cleavage reaction is zinc dependent. However, it is currently not known exactly which protein(s) in this machinery is sensitive to the loss of zinc.

The crystal structure of the N-terminal region (residues 1–460) of human CPSF-73 has been reported (Fig. 2B) [112]. The structure contains a metallo-β-lactamase domain and a β-CASP domain. Remarkably, the structure reveals that a stretch of about 60 residues just after the β-CASP domain actually belongs to the metallo-β-lactamase domain (Fig. 2A, B). Two metal ions, modeled as zinc based on the current data, are bound by the metallo-β-lactamase domain (Fig. 2C), and the binding site is located at the interface with the β-CASP domain. The binding modes of the two zinc ions in CPSF-73 are similar to those of the two zinc ions in RNase Z [107, 108] as well as a bacterial ribonuclease [113]. The structural studies identify a conserved His residue as

the general acid, which is activated by a conserved Asp/Glu residue (Fig. 2C). Mutation of this His residue is lethal in yeast [114]. In addition, the *brr5–1* phenotype in yeast, a cold sensitive mutant that exhibits defects in cleavage and polyadenylation [99], is caused by a single-site mutation, A407T [115], that is located next to this His residue in Brr5p/Ysh1p. The functional importance of this His-Glu pair is also confirmed in RNase Z [116].

Biochemical studies with the bacterially expressed and purified human CPSF-73 showed that it possessed weak ribonuclease activity towards pre-mRNA substrates [112], in the absence of the other proteins of the 3'-end processing machinery. This activity was enhanced greatly when the purified protein was pre-incubated with Ca^{2+} ions. The activity is endonucleolytic and has little sequence specificity, consistent with the fact that CPSF-160 and CstF-64 help define the exact cleavage site in the pre-mRNA substrate. Overall, the structural and biochemical studies offer convincing direct experimental evidence that CPSF-73 is the endoribonuclease for the cleavage reaction in pre-mRNA 3'-end processing (Fig. 1A).

While the C-terminal region of CPSF-73 is not as conserved as the N-terminal region, removal of the C-terminus of Brr5p/Ysh1p, by as little as the last 30 residues, results in cell death [111]. On the other hand, removal of the last 10 or 19 residues do not affect cell viability. Sequence analysis suggests that a leucine zipper, which is usually involved in protein-protein interactions, may be present in the final 30 residues of Brr5p/Ysh1p. The C terminus of CPSF-73 does interact with other proteins in the 3'-end processing complex, including CPSF-100 [117] and Pta1p [111]. Brr5p/Ysh1p has also been shown to interact strongly with Clp1p, but the binding region is unknown [118]. This association may bridge CPSF with CF II_m in mammals [119].

The C-terminal residues of Brr5p/Ysh1p have strong sequence homology to another member of the CPF in the yeast 3'-end processing complex, Syc1p (38% identity) [111]. This is discussed in more detail in the section on Syc1p later.

CPSF-73 has a second isoform in humans, known as RC-68 or Int9 [117, 120]. It may be involved in 3'-end processing of small nuclear RNAs and it interacts with the second isoform of CPSF-100, RC-74 or Int11 [117, 120].

CPSF-100 (Cft2p/Ydh1p). Cft2p/Ydh1p is critical for yeast cell viability [121], and conditional mutants of this protein disrupt both cleavage and polyadenylation *in vitro* and *in vivo* [118]. CPSF-100 has recognizable sequence homology to CPSF-73 (23% identity and 49% similarity for their metallo-β-lactamase

domains) [103]. This sequence conservation is comparable to that between CPSF-100 and Cft2p/Ydh1p (24% identity and 43% similarity for the entire proteins). Therefore, CPSF-100 is also a member of the β-CASP superfamily of proteins. However, the zinc-binding residues are not conserved in CPSF-100 (especially in Cft2p/Ydh1p), and therefore CPSF-100 cannot bind zinc and is not expected to be catalytically active [100, 101]. The exact function of this protein in 3′-end processing is currently not known. Cft2p/Ydh1p contains an insert of about 200 residues (residues 401–601) that is highly charged and hydrophilic [112], which explains its larger size compared to CPSF-73 but the functional role of this segment is not known. CPSF-100 contains a similar (though shorter) segment (Fig. 2A).

The crystal structure of the metallo-β-lactamase and β-CASP domains of Cft2p/Ydh1p was reported recently [112]. The overall structure is similar to that of CPSF-73 (Fig. 2D), and the structure confirms that the zinc-binding site is severely disrupted in this protein. Interactions between CPSF-100 and other proteins in the 3′-end processing complex have been identified. Cft2p/Ydh1p self-associates and also interacts with many proteins of the CPF (Cft1p/Yhh1p, Brr5p/Ysh1p, Pta1p, Pfs2p, Ssu72p, YDL094c), Pcf11p of CF IA, and the CTD of PolII [118]. Deletion mutagenesis studies showed that the C-terminal region of Cft2p/Ydh1p is required for its self association, while the N-terminal region is required for interaction with Pfs2p. In contrast, the full-length protein is required for interactions with the other proteins [118]. Yeast two-hybrid studies showed that the last 245 residues of CPSF-73 can interact with CPSF-100 [117]. Besides protein-protein interactions, Cft2p/Ydh1p can bind the pre-mRNA near the cleavage site [72, 122].

Like CPSF-73, CPSF-100 has a second isoform in humans, known as RC-74 or Int11, also with disrupted zinc binding sites [117, 120].

CPSF-30 (Yth1p). CPSF-30 is required for both cleavage and polyadenylation [123], despite initial experiments showing that it might only be essential for polyadenylation [124]. Interestingly, Yth1p does not appear to co-purify with the other CPSF subunits (Cft1p/Yhh1p, Cft2p/Ydh1p, Brr5p/Ysh1p, Fip1p), even though they are in the larger CPF complex (Fig. 1B) [115, 122]. CPSF-30 has a second isoform in humans (locus XP_945726), sharing 54% sequence identity. However, the function of this close homolog has not been characterized.

CPSF-30 contains five CCCH zinc finger motifs (ZF1–ZF5) in all eukaryotes, and metazoan CPSF-30 has an additional CCHC zinc knuckle at the C terminus (Fig. 2A). CCCH zinc fingers have the consensus sequence $CX_8CX_5CX_3H$, while CCHC zinc knuckles have the consensus sequence $CX_2CX_4HX_4C$ [124]. Both motifs have been shown to bind RNA by UV cross-linking [124], and in vitro RNA binding assays demonstrate that CPSF-30 preferentially binds a poly(U) sequence. Deletion of the zinc knuckle significantly decreases binding but retains specificity. RNase H protection assays in yeast identified that binding of Yth1p occurs at the U-rich sequences (UUE and DUE) that surround the cleavage site [123]. The mutant yth1–1, which lacks the C-terminal 55 amino acids and thus the last zinc finger, still efficiently cleaves pre-mRNA but polyadenylation fails [124].

Sequence alignment indicates that ZF2 is the most conserved zinc finger in CPSF-30 (76% identity and 96% similarity between yeast and mammals). Point mutations made to the conserved Cys residues in ZF2 are lethal [123]. Mutations to other residues within ZF2 reduce the cleavage activity in vitro, but this could be rescued by the addition of wild-type Yth1p. Mutations in ZF2 also disrupt binding to pre-mRNAs [123]. Deletions of ZF1, ZF4 and ZF5 result in lethality or slowed growth, indicating their requirement for fully functioning Yth1p [125]. In contrast, deletion of ZF3 did not alter cell growth, indicating that it is the only zinc finger that is not required for function [125].

CCCH zinc finger motifs are also involved in protein-protein interactions, and CPSF-30 coordinates other proteins in cleavage and polyadenylation. ZF4 of Yth1p is required for interactions with Fip1p, and ZF5 contributes to this binding [123, 124]. hFip1 also interacts with CPSF-30, although the location of interaction has not been mapped [126]. ZF1 and ZF4 of Yth1p interact with the N-terminal region of Brr5p/Ysh1p [125]. Influenza virus attenuates host antiviral response by blocking the function of CPSF-30 with its NS1 protein [127–129].

CPSF-30 has high sequence homology with the *Drosophila* protein clipper (70% identity over the 174 N-terminal residues in bovine CPSF-30). Clipper contains five CCCH zinc finger motifs and two zinc knuckle motifs, and the high sequence conservation covers the five zinc finger motifs. This region of clipper has endoribonuclease activity and cleaves RNA hairpins [130]. Consequently, it has been suggested that CPSF-30 might be the endoribonuclease responsible for the cleavage reaction [131]. However, endoribonucleolytic activity has not been observed for CPSF-30 [125].

hFip1 (Fip1p). Fip1p (factor interacting with Pap1p) was first identified in yeast, through its interaction with Pap1p [132]. hFip1 was subsequently identified

by sequence analysis of the HeLa cell cDNA library and confirmed to be a member of CPSF [126]. hFip1 contains an acidic segment near the N terminus, followed by a highly conserved segment of about 70 residues (48% identity and 72% similarity between hFip1 and Fip1p) (Fig. 2A). The C-terminal region contains a Pro-rich segment, a segment with alternating Arg and Asp residues (RD segment), and an Arg-rich segment. Fip1p, at 35 kDa, is much smaller than hFip1 (66 kDa), because it does not contain the RD and Arg-rich segments. hFip1 runs as a diffuse band in SDS-PAGE, with apparent molecular masses between 65 and 80 kDa.

The acidic segment (residues 1–111) of hFip1 mediates interactions with PAP, and the highly conserved segment (residues 137–243) interacts with CPSF-30. Both of these segments (residues 1–355) are required for binding CPSF-160 and CstF-77, and CPSF-160 may also interact with the C-terminal region of Fip1 (residues 443–594) [126]. A stable ternary complex of hFip1, CPSF-160 and PAP can be identified through purification of HeLa cell extracts. The interaction with PAP may also be aided by the presence of CF I$_m$ [133]. Recombinantly produced RD segment preferentially binds to U-rich RNA sequences, which is present in the auxiliary upstream element of pre-mRNAs [126]. Similar interactions have been observed for Fip1p. Residues 80–105 of Fip1p interact directly with Pap1p, and residues 206–220 interact with Yth1p [134]. There are also weak interactions between Fip1p and Rna14p [124, 132, 134].

The primary function of hFip1 (Fip1p) may be to bring PAP close to the polyadenylation site. Removal of the C-terminal region beyond the Pro-rich segment of Fip1p produces temperature-sensitive growth in yeast. These cells are deficient in polyadenylation, but can be rescued by the addition of wild-type Fip1p. Interestingly, yeast Pap1p is inhibited *in vitro* by the addition of Fip1p, likely mediated by residues 105–206 of Fip1p [134, 135]. In contrast, the addition of recombinant hFip1 stimulates the activity of PAP *in vitro*, which is dependent on the U-rich auxiliary upstream element. The stimulation is abolished by the deletion of this element in the pre-mRNA, or by the deletion of the RD segment or the CPSF-160-binding segment in hFip1 [126].

Cleavage stimulation factor

CstF-64 (Rna15p). CstF-64 was one of the first proteins identified within the 3′-end processing complex because it can be UV cross-linked to RNA [136]. CstF-64 contains a conserved RNA-recognition motif (RRM), the RNA-binding domain (43% identity between CstF-64 and Rna15p), at its N-terminus

(Fig. 3A). The RRM alone is sufficient for RNA binding and shows preference for the G/U-rich DSE (Fig. 1A) [49]. The structure of the RRM of CstF-64 has been determined by NMR (Fig. 3B) [137], which showed the presence of a UU dinucleotide-specific binding site and that the protein:RNA interface is highly mobile [138]. This flexibility may allow CstF-64 to form stable complexes with a wide range of GU-rich sequences.

The RRM in Rna15p has unique properties compared to that in CstF-64 (Table 1). First, the RRM achieves sequence specificity only in the presence of other proteins, especially Rna14p and Hrp1p, but Hrp1p and the EE it recognizes are not present in mammals. Second, the RRM in Rna15p binds specifically to the A-rich PE (Fig. 1B), in contrast to the G/U-rich DSE in mammals. Mutation to conserved amino acid residues in Rna15p eliminates RNA binding, and mutation to the PE prevents Rna15p cross-linking [139]. Therefore, CF IA is located upstream of the CPF (Fig. 1B), whereas its mammalian counterpart CstF is located downstream of CPSF (Fig. 1A).

Immediately after the RRM there is a hinge region of about 100 residues that is highly conserved in CstF-64 but somewhat less conserved in Rna15p (Fig. 3A). The hinge region mediates protein-protein interactions and binds to CstF-77 and symplekin [140, 141]. CstF-64 contains a C-terminal domain (CTD), covering the last 50 amino acids, that is more conserved than the RRM [142]. This domain forms a three-helical bundle (Fig. 3C) [142]. Deletions in the C terminus of Rna15p lead to slow growth or cell death *in vivo* and are defective in cleavage *in vitro*. The C-terminus of Rna15p binds to another member of CF IA, Pcf11p, as truncation of the last 16 residues disrupts the interaction between these two proteins [142]. In addition, this region can interact with transcription factors and may play a role in regulating transcription [4].

Metazoan CstF-64 homologs contain additional sequences between the hinge region and the CTD, including a proline/glycine-rich segment followed by 12 repeats of the MEAR(A/G) pentapeptide motif. The length and composition of this region is variable among metazoans and its function is unknown [1]. The MEAR(A/G) repeats may form a helical structure in solution [143].

CstF-64 has a second isoform in humans, known as τCstF-64. It is expressed in male germ cells and may have important roles in germ-cell-specific polyadenylation and spermatogenesis [144–148].

CstF-77 (Rna14p). CstF-77 is required for proper 3′-end cleavage. Mutation of the *Drosophila* homolog of CstF-77, suppressor of forked *su(f)*, results in the utilization of alternative poly(A) sites [149]. This
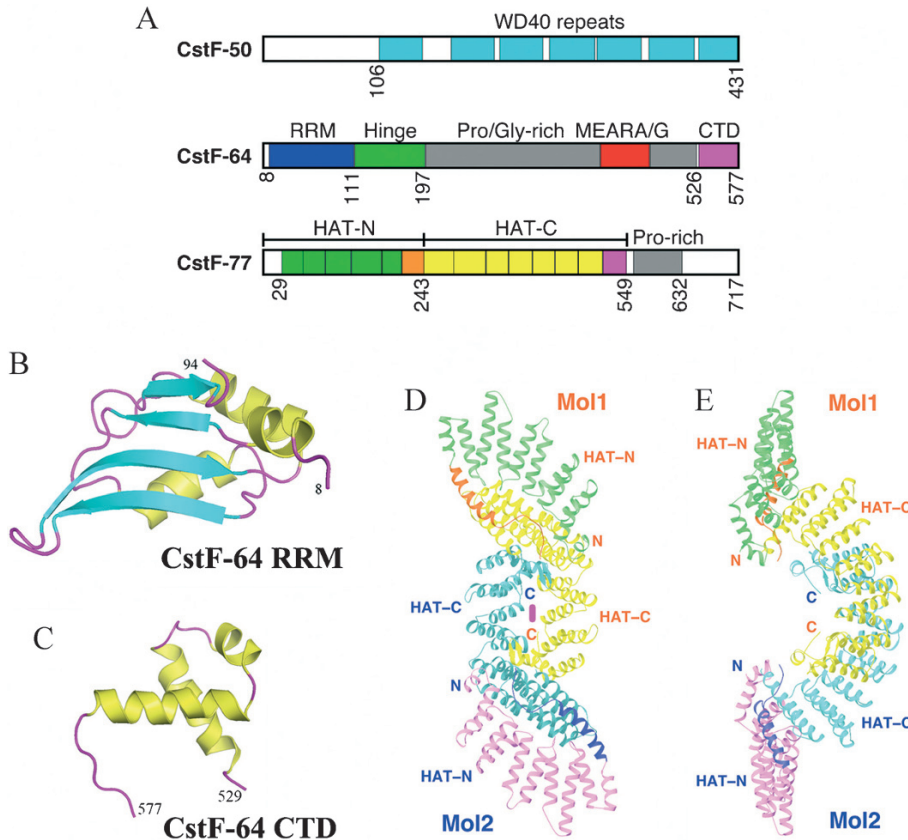
Figure 3. Cleavage stimulation factor (CstF). (*A*) Schematic drawing of the domain organization of the three CstF subunits in humans. (*B*) Structure of the RNA recognition motif (RRM) of CstF-64 [137]. (*C*) Structure of the C-terminal domain (CTD) of CstF-64 [142]. (*D*) Crystal structure of the HAT domain (residues 21–240 and 243–549) of murine CstF-77 [152]. (*E*) Crystal structure of the HAT domain viewed from a different orientation.

defect can be rescued by the addition of human CstF-77 [150].

CstF-77 contains 12 repeated sequences at the N-terminus (Fig. 3A), which are called HAT (half a TPR) motifs for their similarity to tetratricopeptide repeat (TPR) motifs [151]. TPR motifs often mediate protein-protein interactions. Structural and biochemical data suggest that the HAT domain can be further divided into two sub-domains, HAT-N domain (residues 1–240, with HAT motifs 1–5) and HAT-C domain (residues 241–549, HAT motifs 6–12) [152–154]. Most importantly, the structures reveal that the HAT domain is an intimately associated dimer, mediated by the HAT-C domain (Fig. 3D). The overall shape of the dimer is highly elongated, about 45 Å wide but 165 Å long (Fig. 3D). From the side, the dimer looks like a bow (Fig. 3E). This dimerization is supported by studies in solution, yeast two-hybrid assays [152], far Western experiments [141], as well as by solution and electron microscopy studies on Rna14p [155]. In addition, self association of CstF-77 has been suggested based on genetic observations on the *Drosophila* homolog Su(f), as distinct lethal alleles of Su(f) can partially complement each other [150, 156]. Taken together, these data suggest that CstF-77 may function as a dimer at a crucial stage in pre-mRNA 3′-end processing.

The HAT domain is followed by a proline-rich segment in CstF-77 (Fig. 3A). Far Western experiments showed that this segment binds the hinge region of CstF-64 and the WD-40 repeats of CstF-50 [141], and the interaction between CstF-77 and CstF-64 was further confirmed by nickel pull-down assays and analytical ultracentrifugation [152]. Electron microscopy and analytical ultracentrifugation experiments showed that Rna14p and Rna15p can form a heterotetramer [155]. Yeast two-hybrid assays showed that the HAT-C domain mediates interactions with the second β-propeller of CPSF-160, the same region that recognizes the AAUAAA PAS [152]. CstF-77 binds specifically to the CTD of PolII but with less efficiency than CstF-50 [85]. Rna14p binds to unphosphorylated CTD, but the binding increases upon phosphorylation of the CTD [157].

CstF-50. CstF-50 is required for cleavage *in vitro* [158]. It contains seven WD-40 repeats that begin about 90 residues from the N terminus (Fig. 3A). The WD-40 repeats are required for interaction with CstF-77, and deletion of the last repeat reduces binding [141]. CstF-50 also can self-associate and only the N-terminal region is required for this interaction. CstF-50 does not appear to have a sequence homolog in yeast.

Both CstF-50 and CstF-77 bind specifically to the CTD of PolII but CstF-50 binds with a higher efficiency [85]. This binding is significantly reduced upon deletion of the first 91 amino acids of CstF-50, indicating that the WD-40 repeats are not sufficient for interaction. RNAi experiments showed that CstF-50 also interacts with the splicing co-activator SRm160, establishing another link between 3′-end processing and transcription [159]. A yeast two-hybrid screen identified an interaction between the WD-40 repeats of CstF-50 and the protein BARD1, which associates with the tumor suppressor BRCA1 [160, 161]. This interaction inhibits 3′-end cleavage of pre-mRNAs *in vitro*.

Mammalian cleavage factor I

CF I$_m$ is required for cleavage *in vitro* [162], and appears to be unique to higher eukaryotes. Three polypeptides (25, 59 and 68 kDa) as well as a less abundant one (72 kDa) generally copurify with CF I$_m$ activity from HeLa cell nuclear extract. CF I$_m$ functions as a heterodimer that is made up of the 25-kDa subunit and one of the 59-, 68- or 72-kDa subunits. The heterodimer consisting of the 25- and 68-kDa polypeptides can reconstitute cleavage activity to partially purified 3′-end processing factors [163].

The three large subunits have similar amino acid sequences, but are encoded by separate genes [164]. The N-terminal region of the 68-kDa subunit contains an RNP-type RNA-binding domain (RBD), which is necessary for binding to the 25-kDa subunit [164]. RBDs have been observed in protein-protein interactions in other cases, for example the splicing factor U2AF [165, 166] and the *Drosophila* proteins Y14 and Mago [167]. The C-terminal region of the 68-kDa subunit is rich in RS, RD and RE repeats that are similar to pre-mRNA splicing SR proteins. In fact, this region interacts with the spliceosomal SR proteins [167], and CF I$_m$ has been identified as a component of purified spliceosomes [168, 169]. For the 25-kDa subunit, residues 81–160 mediate interactions with the C terminus of PAP (residues 472–739) and PABP [164, 170].

The primary function of CF I$_m$ may be to provide additional recognition of the pre-mRNA substrate and aid the definition of the proper polyadenylation site. The 25-, 59- and 68-kDa subunits of CF I$_m$ can be UV cross-linked to RNA in a sequence-dependent manner [171], and SELEX analysis revealed that CF I$_m$ prefers to bind to the RNA sequences containing UGUAA [170], which is generally found upstream of the PAS. Further studies demonstrated that CF I$_m$ binding enhances the recognition of sequences that contain both the perfect PAS and the noncanonical PAS [133]. The enhancement on the perfect PAS is

aided by binding of hFip1 and the consequent recruitment of PAP [133].

In mammals, dephosphorylation of a protein in CF I$_m$ or CF II$_m$ by a Ser/Thr phosphatase abolished the cleavage reaction [114].

Mammalian cleavage factor II (CF II$_m$)

CF II$_m$ contains two subunits, hPcf11 and hClp1, both of which were originally discovered in yeast (Pcf11p and Clp1p) and belong to CF IA [119]. These proteins are highly conserved among the eukaryotes.

hPcf11 (Pcf11p). hPcf11 can be identified in nuclear extracts with apparent molecular masses between 140 and 200 kDa on SDS gels [119]. Pcf11p is only about half the size of hPcf11, and is equivalent to the N-terminal half of hPcf11. The function of the C-terminal half of hPcf11 is not known.

Pcf11p contains a conserved PolII CTD interacting domain (CID) that covers ~130 residues in the N terminus (Fig. 4A), which prefers the phosphorylated form of the CTD [157, 172, 173]. The crystal structure of Pcf11p in complex with a phosphorylated CTD peptide has been reported (Fig. 4B) [174]. Mutations in the CID reduce or abolish binding to the CTD and result in cell death. Interestingly, these mutations do not affect 3′-end processing, but instead cause incorrect transcriptional termination [157, 174–176]. The CID can also bind RNA to affect transcriptional termination [176]. Studies of the CID indicate that the RNA binding and the protein binding regions overlap, and this competition for binding may play a role in the release of the 3′-end processing factors from PolII [177].

The CID is followed by a 20-residue stretch of glutamines (234–253 in Pcf11p), which is followed by the Rna14p/Rna15p binding domain (Fig. 4A). The binding of Rna14p/Rna15p to Pcf11p is dependent on the binding of Clp1p first [139]. This CstF binding region is followed by the Clp1 interacting segment (477–499, discussed below), which is flanked by a zinc finger on each side (N-terminal C$_2$H$_2$ and C-terminal C$_2$HC type). Pcf11p also interacts with Cft1p/Yhh1p, Cft2p/Ydh1p, Brr5p/Ysh1p, and Pta1p [118], but the location of the binding sites is unknown.

hClp1 (Clp1p). hClp1 is conserved throughout eukaryotes and has 23% identity with Clp1p [119]. Immunodepletion of hClp1 abolished cleavage activity but does not affect polyadenylation. Clp1p binds strongly to Brr5p/Ysh1p and Pcf11p, and weakly to Cft2p/Ydh1p [118]. Similarly, hClp1 interacts with CF I$_m$ and CPSF [119, 178].

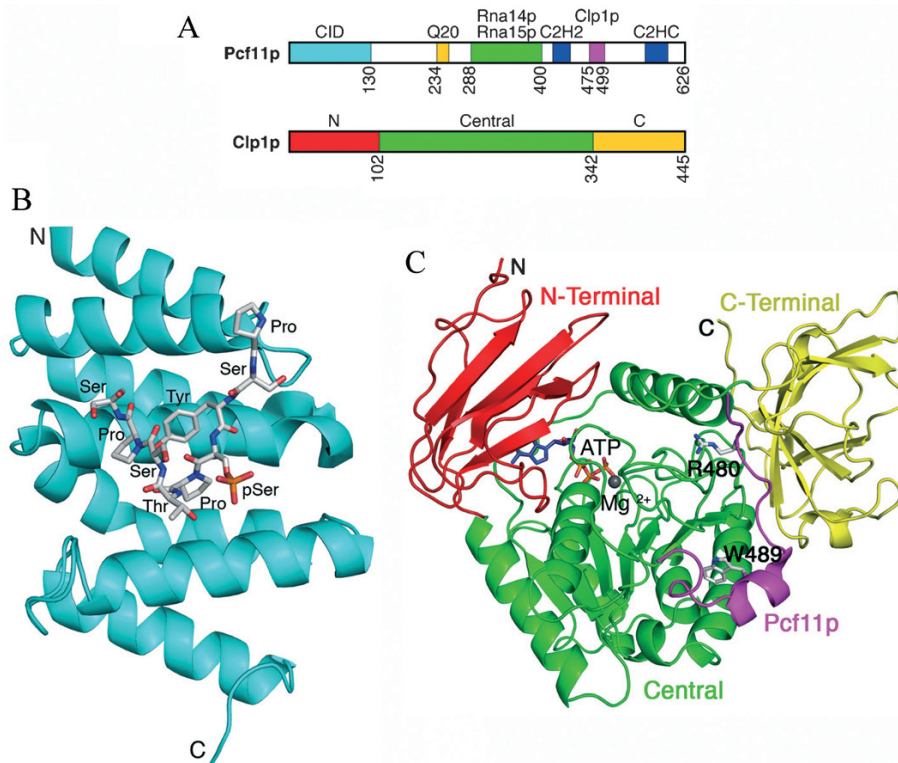hClp1 and its homologs contain a Walker A motif (residues 130–137), which is generally associated with

Figure 4. Cleavage factor II (CF II$_m$). (*A*) Schematic drawing of the domain organization of Pcf11p and hClp1. (*B*) Crystal structure of the PolII CTD interacting domain (CID) of Pcf11p in complex with a CTD peptide [174]. (*C*) Crystal structure of Clp1p in complex with ATP and a peptide from Pcf11p [180].

binding ATP/GTP (Fig. 4A) [179]. The structure of Clp1p confirms the similarity to other ATPases (Fig. 4C) [180]. The bound conformation of ATP suggests that Clp1p may not have ATPase activity, which is confirmed by biochemical assays *in vitro*. Recent studies show however that hClp1 is a RNA 5′-kinase that is important for tRNA splicing and activation of synthetic short interfering RNAs [181]. The structure also reveals the molecular basis for the interactions between Clp1p and Pcf11p (Fig. 4C) [180]. A peptide segment of Pcf11p (residues 475–499) is bound by Clp1p. Two of the residues in this segment, Arg480 and Trp489, are strictly conserved among Pcf11p homologs and make large contributions to the binding interface (Fig. 4C).

Poly(A) polymerase (Pap1p)

In mammals, PAP is required for cleavage and polyadenylation, but in yeast Pap1p is only required for polyadenylation. While PAP interacts with other members of the 3′-end processing complex, it does not require any of these proteins for *in vitro* polyadenylation, although these interactions are important for defining the proper length of the poly(A) tail [14].

The structures of human, bovine and yeast PAP have been reported [182–185]. Overall, the structure of PAP contains three domains, N-terminal, middle and C-terminal domains (Fig. 5A). The active site is located at the bottom of a large cleft between the N-

and C-terminal domains. The N-terminal domain coordinates the two metal ions (Mg^{2+} or Mn^{2+}) that are required for catalysis and is strongly conserved throughout eukaryotes. The C-terminal domain binds hFip1 (Fip1p) and CPSF-160 [126, 132, 183, 186]. The RNA-binding site in this domain is not directly involved in binding the RNA substrate. The structure of Pap1p in a ternary complex with MgATP and an oligo(A) shows that only the last three nucleotides of the pre-mRNA substrate are bound tightly by the enzyme (Fig. 5A) [187].

PAP is an induced-fit enzyme, and this induced-fit behavior has an important role in defining the substrate specificity of the enzyme [187–190]. The N-terminal domain undergoes a large movement upon the formation of the ternary complex with the MgATP and pre-mRNA substrates, whereas MgGTP cannot induce this active site closure. Other polymerases may also use an induced-fit mechanism to help regulate substrate preference [191].

Poly(A)-binding protein

PABP (Pab1p) is required for correct and efficient polyadenylation. The polyadenylation reaction can occur without PABP, but controlling the length of the poly(A) tail requires PABP [192–195]. PABP binds directly to nascent stretches of 11–14 adenylate nucleotides as they become available [196], and this binding continues until the proper poly(A) tail length
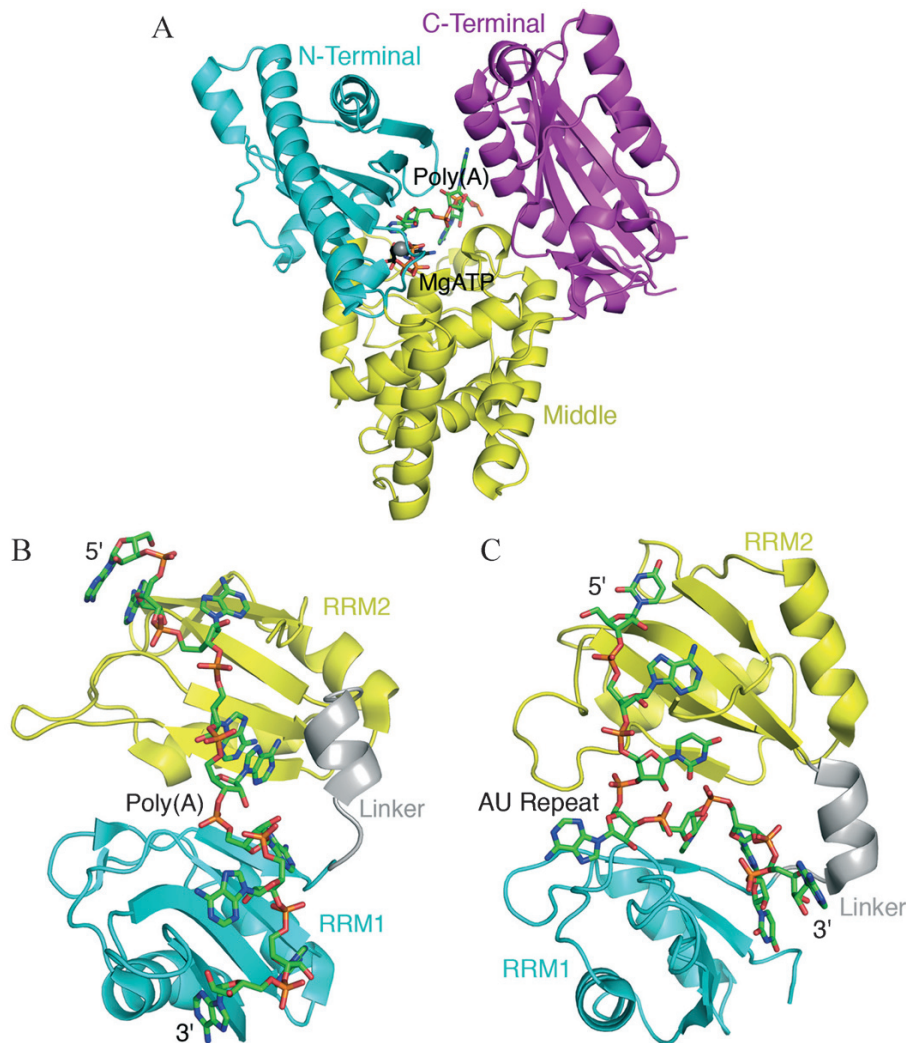
Figure 5. Structural information on other factors in pre-mRNA 3′-end processing. (*A*) Crystal structure of Pap1p (D154A mutant) in a ternary complex with MgATP and oligo-adenylate [187]. (*B*) Crystal structure of the first two RRMs of PABP in complex with an 11-mer adenylate (A_{11}) [199]. (*C*) Structure of the RRMs of Hrp1p in complex with the AU-rich efficiency element (EE) [216].

is reached (~200–300 bases in mammals) [197]. In addition, direct binding of PABP to pre-mRNA, adjacent to PAP, increases the efficiency of polyadenylation 80-fold [198]. Pab1p binds directly to Rna15p, which possibly recruits Pab1p to the polyadenylation site [192]. Interactions between PABP and other members of the mammalian 3′-end processing complex have not been reported.

PABP contains four tandem repeats of the RRMs, and the crystal structure of the first two motifs in complex with an 11-mer adenylate nucleotide (A_{11}) has been reported (Fig. 5B) [199]. The two RRMs form a contiguous RNA-binding site, primarily using the face of the β-sheet in the two RRMs, and the nucleotide adopts an extended conformation in the complex (Fig. 5B). The adenine bases are recognized by conserved residues in the RRMs.

Symplekin (Pta1p)

Pta1p was originally identified as an essential component of pre-tRNA processing [200], while symplekin was discovered in association with tight junctions [201]. Both were later identified in 3′-end processing complexes. Two separate cDNAs of symplekin were isolated, encoding proteins of 1273 and 1058 residues [141]. The first 964 residues of the two forms of symplekin are identical, suggesting that they are derived from alternative splicing. Symplekin and Pta1p share very weak sequence homology, but a putative conserved region between the two proteins have been identified by sequence analysis, with 17 % identity and 31 % similarity [141].

Pta1p is required for both cleavage and polyadenylation [115]. Symplekin/Pta1p probably functions as a scaffolding protein, interacting with and possibly bringing together a large number of proteins in the 3′-end processing complex. Pta1p purifies with CPF and can be further purified into CF II. It interacts

directly with the C terminus of Brr5p/Ysh1p [111], Cft2p/Ydh1p [118], Syc1p, Glc7p, Pti1p, Ssu72p and Swd2p [81, 202, 203]. Symplekin was identified through its interaction with the hinge region of CstF-64 [141], and forms a stable complex with CPSF and CstF.

Depletion of the phosphatase Glc7p leads to accumulation of phosphorylated Pta1p and shortened poly(A) tails [204]. Restoration of normal polyadenylation requires either Glc7p or unphosphorylated Pta1p, indicating that Pta1p is likely a target for the regulation of polyadenylation through phosphorylation. The mechanism of this regulation is currently not known.

Symplekin may also function as a scaffold protein in the 3′-end processing of histone pre-mRNAs [205] as well as polyadenylation in the cytoplasm. Symplekin co-localizes with CPSF-100 in Cajal bodies during oocyte maturation, and is required for cytoplasmic polyadenylation in these oocytes [206]. Symplekin mediates protein-protein interactions in cytoplasmic polyadenylation by directly contacting CPSF and the RNA-binding protein CPEB [33].

The CTD of PolII

The CTD is made up of 52 heptapeptide repeats in humans and 26 repeats in yeast with a consensus sequence of YSPTSPS for each repeat. The repeated serine residues within the CTD are susceptible to phosphorylation [207]. The CTD is not observed in the structure of yeast PolII, suggesting that it may be flexible [208]. Sequence analysis and early NMR studies proposed that the CTD forms a β-turn. The structure of the CID in Pcf11p in complex with a CTD peptide containing a phosphorylated serine showed that the CTD peptide assumed the conformation of a β-turn in this complex (Fig. 4B) [174]. CTD peptides have also been observed in more extended conformations [209, 210]. The variety of CTD conformations suggests that the CTD may not form a single basal structure [211].

PolII was identified in 3′-end processing because transiently transfected cells with CTD truncations exhibit inefficient polyadenylation [85]. Purified phosphorylated and unphosphorylated CTD activated a reconstituted cleavage reaction *in vitro* [212]. While the CTD is necessary for cleavage in mammals, it does not appear to be necessary for cleavage in yeast [213]. Deletion of the CTD in yeast produces unstable mRNAs that are degraded, but the degradation can be rescued by blocking the 5′-to-3′ exonuclease XRN1, indicating a problem with the 5′-end cap as opposed to 3′-end polyadenylation [214].

Both mammalian and yeast CTD bind CPSF-160 (Cft1p/Yhh1p) [85, 93]. Additionally, yeast CTD interacts with Cft2p/Ydh1p and Pcf11p [118, 172, 173, 213] and the mammalian CTD binds CstF-50 [85]. Protein binding generally increases upon phosphorylation of the CTD, and 3′-end processing is stimulated *in vitro* by phosphorylated CTD [215]. Disruption of these interactions often leads to poor transcriptional termination.

Other protein factors in 3′-end processing

Several other 3′-end processing proteins have been identified in yeast. While some of these proteins have mammalian homologs, none of these homologs have yet been identified as necessary for 3′-end processing in mammals. Additional yeast proteins were identified through their co-purification with Pta1p [77] or Ref2p [78]. Most of the proteins identified are associated with snRNA and snoRNA 3′-end formation, strengthening the connection between pre-mRNA processing and other cellular processes [4]. The proteins that have been recognized as playing a role in 3′-end processing are described below.

Hrp1p. Hrp1p is the only member of the processing factor CF IB (Fig. 1B). It is required for proper cleavage in yeast, although it does not have a homolog in mammals. Hrp1p contains two centrally located consecutive RRM-type RNA-binding domains (residues 160–290). Hrp1p cross-links to RNA with specificity for the AU-rich EE, and the molecular basis of this recognition has been determined by NMR [216]. Both RRMs specifically recognize the hexamer AU-rich site (Fig. 5C). Besides binding the EE, genetic analysis indicates that Hrp1p interacts directly with the CF IA components Rna14 and Rna15 [217].

Pfs2p. Pfs2p is a subunit of CPF and can be further purified into PFI. It contains seven WD-40 repeats from residue 90 to 380. Disruption of Pfs2p, like most proteins in the 3′-end processing complex, results in cell death [218]. Mutation or depletion of Pfs2p impairs both cleavage and polyadenylation, indicating that Pfs2p plays an essential role in both processes. The N-terminal domain prior to the WD-40 repeats is not necessary for 3′-end processing activity. Pfs2p directly interacts with proteins from CPF (Fip1p, Brr5p/Ysh1p and Swd2p) and CF IA (Rna14p), suggesting that Pfs2p plays a role in tethering the two factors for proper 3′-end processing [202, 219].

Pfs2p has recognizable sequence homology with the N-terminal domain of the human protein WDR33, which also contains a large, collagen-like domain in the C terminus [220]. It is not known whether WDR33 has any role in pre-mRNA 3′-end processing.

Ssu72p. Ssu72p was originally identified by its genetic interaction with the transcription factor TFIIB, as a mutation in Ssu72p disrupted this interaction and affected the accuracy of start site selection [221]. Further studies identified Ssu72p within the CPF [77–79]. It is essential for the cleavage reaction but is not required for polyadenylation [81]. Ssu72p interacts with Pta1p, Cft2p/Ydh1p [79] and the Rpb2p subunit of PolII [222]. Upon binding to Pta1p in CPF, Ssu72p functions as a phosphatase with specificity for the fifth serine of the heptapeptide repeat in the CTD [223, 224]. However, this activity is not required for 3′-end processing, but it may affect its role in 5′-end capping [224].

The human homolog of Ssu72p (hSsu72) has been identified as an interacting protein of the retinoblastoma tumor suppressor, Rb [225]. Similar to its yeast homolog, hSsu72 binds directly to TFIIB and Pta1p, and exhibits phosphatase activity, but its role in 3′-end processing is still unknown. *hSsu72* cannot rescue a lethal mutation in *Ssu72* in yeast [225]. A close homolog of hSsu72 is found in humans [222], sharing 72 % sequence identity. It has been suggested that this second isoform is encoded by a pseudogene [225].

Pti1p. Pti1p is part of the snoRNA associated complex from the CPF [78]. It is a sequence homolog of Rna15p and CstF-64, and contains a RRM at the N terminus (residues 20–100). The hinge region (residues 185–260) is more conserved with CstF-64 than Rna15p, and interacts with Rna14p, Pcf11p and Pta1p [203]. In addition, the N terminus and the hinge region of Pti1p interact with Glc7p [204]. Defects in Pti1p affect the cleavage site selection and the polyadenylation length of some pre-mRNAs, but the presence of Pti1p is not essential for 3′-end processing [203]. Over-expression of Pti1p prevents polyadenylation, consistent with its association with snoRNA, which are not polyadenylated. The more likely role for Pti1p may be to coordinate the CPF with snoRNAs, and this has minimal effect on 3′-end processing [82].

Swd2p. Swd2p was identified by its interaction with Pta1p and belongs to the CPF [77]. Additionally, Swd2p associates with the histone methylation complex Set1 [226, 227], indicating a role in transcription as well as 3′-end processing. Sequence analysis of Swd2p and its homologs predicts that they contain seven WD-40 repeats [202]. Depletion of Swd2p does not have an effect on 3′-end cleavage *in vitro* [226]. However, mutations or depletion of Swd2p disrupt proper transcriptional termination [202, 226]. Recombinantly purified GST-Swd2p interacts with many proteins of the CPF, including Cft1p/Yhh1p, Brr5p/Ysh1p, Pta1p, Pfs2p, Ref2p, Ssu72p, Glc7p and Pti1p,

as well as Pcf11p in CF IA [202]. Swd2p may not be essential for pre-mRNA 3′-end processing, but it does tether the complex to other proteins and establishes another connection to transcription.

Swd2p has a close homolog in humans, known as WDR82 (SwissProt entry Q6UXN9), with 35 % sequence identity. The function of this protein is currently not known.

Mpe1p. Mpe1p was identified because it suppressed a temperature-sensitive mutation in Pcf11p [228]. It is required for cleavage and polyadenylation. Affinity-tagged Mpe1p can pull down the entire CPF, indicating that it is a functional member of the CPF. Mpe1p is conserved throughout eukaryotes from yeast to humans, but it is unknown whether the homologs are also involved in 3′-end processing. Mpe1p contains a zinc knuckle and a possible ring finger motif. The zinc knuckle might mimic that in CPSF-30, which is absent in Yth1p [228].

Syc1p. Syc1p is a component of CPF that is highly homologous (38 % identity) to the C-terminal residues of Brr5p/Ysh1p. Syc1p is not required for either of the 3′-end processing functions and deletion of Syc1p does not disrupt cell viability [78]. In fact, deletion of Syc1p partially rescues *in vivo* phenotypes of mutated Brr5p/Ysh1p and Pta1p and alleviates *in vitro* processing defects of mutated Brr5p/Ysh1p. GST pull-down assays identify that Syc1p interacts with the C-terminal domain of Brr5p/Ysh1p and with Pta1p [111]. Therefore, Syc1p may negatively regulate 3′-end processing, possibly by competing with the C-terminal domain of Brr5p/Ysh1p.

Glc7p. Glc7p is a type 1 protein phosphatase that is required for cell viability and participates in a variety of cellular processes, including 3′-end processing [77, 229]. Depletion, inhibition or removal of Glc7p results in mRNAs with shortened poly(A) tails, indicating that Glc7p plays a role in polyadenylation but not cleavage [204, 230]. The level of phosphorylated Pta1p increases upon depletion of Glc7p *in vivo*, identifying Pta1p as a specific target for Glc7p phosphatase activity. The effects of depleted Glc7p can be rescued upon the addition of unphosphorylated Pta1p and to a lesser degree Fip1p [204]. GST-pull-down assays show that Glc7p interacts with residues 101–200 of Pta1p [204]. Additionally, Glc7p can pull-down Pti1p and to a weaker extent Cft1p/Yhh1p, Cft2p/Ydh1p and Brr5p/Ysh1p.

Ref2p. Ref2p belongs to the CPF [78] and is required for efficient 3′-end processing of poorly identified poly(A) sites [231] and snoRNAs [82]. Currently the
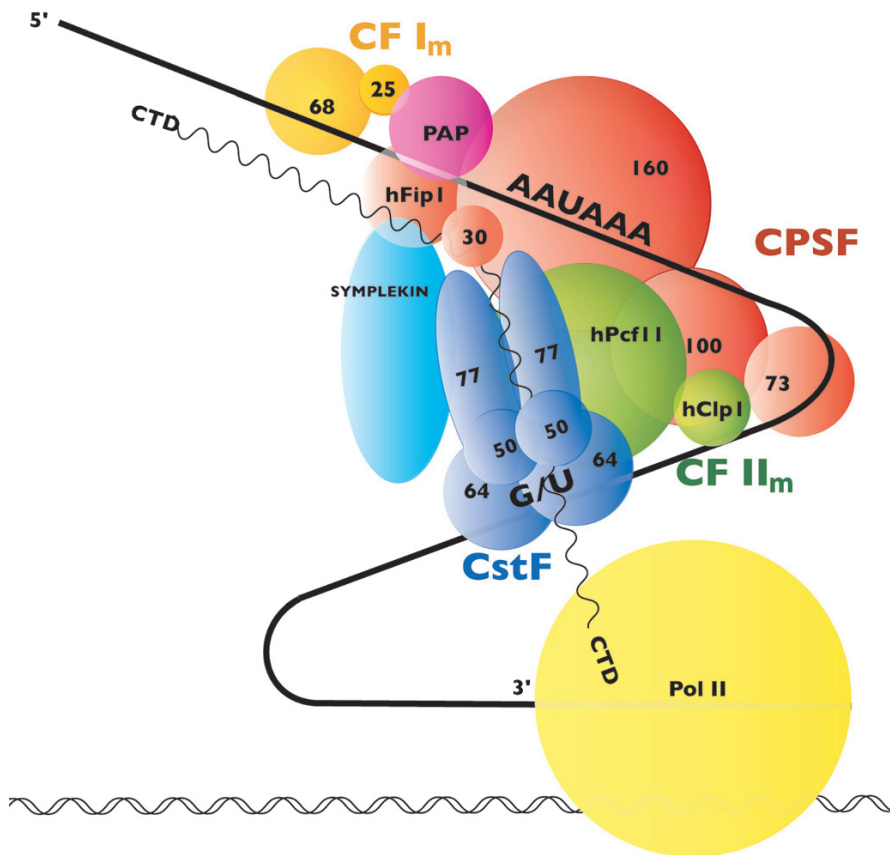
Figure 6. A schematic model for the mammalian pre-mRNA 3′-end processing machinery. The CstF complex is shown in its dimeric form based on current structural and biochemical information.

only known interaction partner of Ref2p is Swd2p [202].

Summary and perspectives

Pre-mRNA 3′-end processing is a fundamental event in most eukaryotes. Studies over the past 20 years have identified a large number of protein factors that are involved in this processing (Table 1), and a schematic model for the mammalian machinery based on the current structural and biochemical information is shown in Figure 6. Many of the protein factors are directly involved in pre-mRNA recognition, cleavage, or polyadenylation. It is likely that the recognition of the upstream PAS by CPSF-160 and the G/U-rich DSE by CstF-64 help position the 3′-end processing machinery on the pre-mRNA, bringing the endoribonuclease (CPSF-73) close to the cleavage site (Fig. 6). This correct placement may also help to activate CPSF-73, as CPSF-73 on its own appears to have very weak activity. This may ensure that the nuclease is only functional when it is in the correct location on the pre-mRNA, preventing nonspecific activity on other cellular RNAs. The current evidence suggests that CstF may be dimeric in this machinery

(Fig. 6). The presence of two CstF-64 subunits (and their RRMs) may be important for recognizing the DSE, whereas the dimeric HAT domains of CstF-77 may provide a platform for interacting with other protein factors, including CPSF-160 (Fig. 6). In addition to binding and processing the pre-mRNA, some of the protein factors in this machinery are also needed to coordinate with transcriptional initiation and termination, 5′-end capping, as well as splicing. This may explain why such a large machinery is involved in pre-mRNA 3′-end processing.

Future studies will focus on defining the functional roles of this large number of protein factors in more detail, as well as how this 3′-end processing machinery is constructed at the molecular level. Structural information on the protein factors themselves is already revealing significant, and sometimes unexpected, molecular insights, for example the dimeric association of CstF in the pre-mRNA 3′-end processing machinery (Fig. 6). The greatest molecular and functional insights will be derived from structural information on the sub-complexes of these protein factors (CPSF, CstF and others) and even the entire pre-mRNA 3′-end processing machinery.

1   Colgan, D. F. and Manley, J. L. (1997) Mechanism and regulation of mRNA polyadenylation. Genes Dev. 11, 2755–2766.

2   Zhao, J., Hyman, L. and Moore, C. L. (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.

3   Proudfoot, N. J. and O'Sullivan, J. (2002) Polyadenylation: A tail of two complexes. Curr. Biol. 12, R855-R857.

4   Proudfoot, N. J. (2004) New perspectives on connecting messenger RNA 3' end formation to transcription. Curr. Opin. Cell Biol. 16, 272–278.

5   Edmonds, M. and Abrams, R. (1960) Polynucleotide biosynthesis: Formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. J. Biol. Chem. 235, 1142–1149.

6   Edmonds, M., Vaughan Jr., M. H. and Nakazato, H. (1971) Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: Possible evidence for a precursor relationship. Proc. Natl. Acad. Sci. USA 68, 1336–1340.

7   Darnell, J. E., Wall, R. and Tushinski, R. J. (1971) An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. Proc. Natl. Acad. Sci. USA 68, 1321–1325.

8   Lee, S. Y., Mendecki, J. and Brawerman, G. (1971) A polynucleotide segment rich in adenylic acid in the rapidly-labeled polyribosomal RNA component of mouse sarcoma 180 ascites cells. Proc. Natl. Acad. Sci. USA 68, 1331–1335.

9   Nevins, J. R. and Darnell, J. E. Jr. (1978) Steps in the processing of Ad2 mRNA: Poly(A)⁺ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell 15, 1477–1493.

10 Ford, J. P. and Hsu, M. T. (1978) Transcription pattern of *in vivo*-labeled late simian virus 40 RNA: Equimolar transcription beyond the mRNA 3' terminus. J. Virol. 28, 795–801.

11 Manley, J. L., Sharp, P. A. and Gefter, M. L. (1982) RNA synthesis in isolated nuclei processing of adenovirus serotype 2 late messenger RNA precursors. J. Mol. Biol. 159, 581–599.

12 Shatkin, A. J. and Manley, J. L. (2000) The ends of the affair: Capping and polyadenylation. Nat. Struct. Biol. 7, 838–842.

13 Calvo, O. and Manley, J. L. (2003) Strange bedfellows: Polyadenylation factors at the promoter. Genes Dev. 17, 1321–1327.

14 Edmonds, M. (2002) A history of poly A sequences: From formation to factors to function. Prog. Nucleic Acid Res. Mol. Biol. 71, 285–389.

15 Gilmartin, G. M. (2005) Eukaryotic mRNA 3' processing: A common means to different ends. Genes Dev. 19, 2517–2521.

16 Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. Nature 416, 499–506.

17 Proudfoot, N. J., Furger, A. and Dye, M. J. (2002) Integrating mRNA processing with transcription. Cell 108, 501–512.

18 Vinciguerra, P. and Stutz, F. (2004) mRNA export: An assembly line from genes to nuclear pores. Curr. Opin. Cell Biol. 16, 285–292.

19 Wahle, E. and Ruegsegger, U. (1999) 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. 23, 277–295.

20 Wickens, M., Anderson, P. and Jackson, R. J. (1997) Life and death in the cytoplasm: Messages from the 3' end. Curr. Opin. Genet. Dev. 7, 220–232.

21 Wilusz, C. J. and Wilusz, J. (2004) Bringing the role of mRNA decay in the control of gene expression into focus. Trends Genet. 20, 491–497.

22 Wilusz, C. J., Wormington, M. and Peltz, S. W. (2001) The cap-to-tail guide to mRNA turnover. Nat. Rev. Mol. Cell Biol. 2, 237–246.

23 Zorio, D. A. R. and Bentley, D. (2004) The link between mRNA processing and transcription: Communication works both ways. Exp. Cell Res. 296, 91–97.

24 Huang, Y. and Carmichael, G. G. (1996) Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol. Cell. Biol. 16, 1534–1542.

25 Ford, L. P., Bagga, P. S. and Wilusz, J. (1997) The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an *in vitro* RNA stability system. Mol. Cell. Biol. 17, 398–406.

26 Wormington, M., Searfoss, A. M. and Hurney, C. A. (1996) Overexpression of poly(A) binding protein prevents maturation-specific deadenylation and translational inactivation in Xenopus oocytes. EMBO J. 15, 900–909.

27 Coller, J. M., Gray, N. K. and Wickens, M. P. (1998) mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes Dev. 12, 3226–3235.

28 Chekanova, J. A. and Belostotsky, D. A. (2006) MicroRNAs and messenger RNA turnover. Methods Mol. Biol. 342, 73–85.

29 Sachs, A. B., Sarnow, P. and Hentze, M. W. (1997) Starting at the beginning, middle, and end: Translation initiation in eukaryotes. Cell 89, 831–838.

30 Preiss, T. and Hentze, M. W. (1998) Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. Nature 392, 516–520.

31 Hirose, Y. and Manley, J. L. (2000) RNA polymerase II and the integration of nuclear events. Genes Dev. 14, 1415–1429.

32 Bilger, A., Fox, C. A., Wahle, E. and Wickens, M. (1994) Nuclear polyadenylation factors recognize cytoplasmic polyadenylation elements. Genes Dev. 8, 1106–1116.

33 Barnard, D. C., Ryan, K., Manley, J. L. and Richter, J. D. (2004) Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. Cell 119, 641–651.

34 Stevenson, A. L. and Norbury, C. J. (2006) The CidI family of non-canonical poly(A) polymerases. Yeast 23, 991–1000.

35 Tian, B., Hu, J., Zhang, H. and Lutz, C. S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acid Res. 33, 201–212.

36 Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M. and Gautheret, D. (2000) Patterns of variant polyadenylation signal usage in human genes. Genome Res. 10, 1001–1010.

37 Proudfoot, N. J. and Brownlee, G. G. (1976) 3' Non-coding region sequences in eukaryotic messenger RNA. Nature 263, 211–214.

38 Manley, J. L. (1988) Polyadenylation of mRNA precursors. Biochim. Biophys. Acta 950, 1–12.

39 Higgs, D. R., Goodbourn, S. E., Lamb, J., Clegg, J. B., Weatherall, D. J. and Proudfoot, N. J. (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation. Nature 306, 398–400.

40 Wickens, M. and Stephenson, P. (1984) Role of the conserved AAUAAA sequence: Four AAUAAA point mutations prevent messenger RNA 3' end formation. Science 226, 1045–1051.

41 Fitzgerald, M. and Shenk, T. (1981) The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24, 251–260.

42 Simonsen, C. C. and Levinson, A. D. (1983) Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Mol. Cell. Biol. 3, 2250–2258.

43 Gil, A. and Proudfoot, N. J. (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. Cell 49, 399–406.

44 McLauchlan, J., Gaffney, D., Whitton, J. L. and Clements, J. B. (1985) The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for

efficient formation of mRNA 3' termini. Nucleic Acid Res. 13, 1347–1368.

45 Chou, Z. F., Chen, F. and Wilusz, J. (1994) Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals. Nucleic Acid Res. 22, 2525–2531.

46 Sittler, A., Gallinaro, H. and Jacob, M. (1994) Upstream and downstream cis-acting elements for cleavage at the L4 polyadenylation site of adenovirus-2. Nucleic Acid Res. 22, 222–231.

47 McDevitt, M. A., Hart, R. P., Wong, W. W. and Nevins, J. R. (1986) Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO J. 5, 2907–2913.

48 Zarkower, D. and Wickens, M. (1988) A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex *in vitro*. J. Biol. Chem. 263, 5780–5788.

49 Takagaki, Y. and Manley, J. L. (1997) RNA recognition by the human polyadenylation factor CstF. Mol. Cell. Biol. 17, 3907–3914.

50 Chen, F., MacDonald, C. C. and Wilusz, J. (1995) Cleavage site determinants in the mammalian polyadenylation signal. Nucleic Acid Res. 23, 2614–2620.

51 Gilmartin, G. M., Fleming, E. S., Oetjen, J. and Graveley, B. R. (1995) CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: Multiple sequence contacts involved in poly(A) site definition. Genes Dev. 9, 72–83.

52 MacDonald, C. C., Wilusz, J. and Shenk, T. (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol. Cell. Biol. 14, 6647–6654.

53 Sheets, M. D., Ogg, S. C. and Wickens, M. P. (1990) Point mutations in AAUAAA and the poly(A) addition site: Effects on the accuracy and efficiency of cleavage and polyadenylation *in vitro*. Nucleic Acid Res. 18, 5799–5805.

54 Hu, J., Lutz, C. S., Wilusz, J. and Tian, B. (2005) Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. RNA 11, 1485–1493.

55 Brackenridge, S. and Proudfoot, N. J. (2000) Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. Mol. Cell. Biol. 20, 2660–2669.

56 Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J. L. and Proudfoot, N. J. (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. Genes Dev. 12, 2522–2534.

57 Moreira, A., Wollerton, M., Monks, J. and Proudfoot, N. J. (1995) Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. EMBO J. 14, 3809–3819.

58 Huang, Y., Wimler, K. M. and Carmichael, G. G. (1999) Intronless mRNA transport elements may affect multiple steps of pre-mRNA processing. EMBO J. 18, 1642–1652.

59 Le Hir, H., Nott, A. and Moore, M. J. (2003) How introns influence and enhance eukaryotic gene expression. Trends Biochem. Sci. 28, 215–220.

60 Arhin, G. K., Boots, M., Bagga, P. S., Milcarek, C. and Wilusz, J. (2002) Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. Nucleic Acid Res. 30, 1842–1850.

61 Bagga, P. S., Ford, L. P., Chen, F. and Wilusz, J. (1995) The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor. Nucleic Acid Res. 23, 1625–1631.

62 Oberg, D., Fay, J., Lambkin, H. and Schwartz, S. (2005) A downstream polyadenylation element in human papillomavirus type 16 L2 encodes multiple GGG motifs and interacts with hnRNP H. J. Virol. 79, 9254–9269.

63 Dalziel, M., Nunes, N. M. and Furger, A. (2007) Two G-rich regulatory elements located adjacent to and 440 nucleotides downstream of the core poly(A) site of the intronless melanocortin receptor 1 gene are critical for efficient 3' end processing. Mol. Cell. Biol. 27, 1568–1580.

64 Gromak, N., West, S. and Proudfoot, N. J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol. Cell. Biol. 26, 3986–3996.

65 Irniger, S. and Braus, G. H. (1994) Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 91, 257–261.

66 Guo, Z., Russo, P., Yun, D. F., Butler, J. S. and Sherman, F. (1995) Redundant 3' end-forming signals for the yeast CYC1 mRNA. Proc. Natl. Acad. Sci. USA 92, 4211–4214.

67 Guo, Z. and Sherman, F. (1996) 3'-end-forming signals of yeast mRNA. Trends Biochem. Sci. 21, 477–481.

68 Guo, Z. and Sherman, F. (1995) 3'-end-forming signals of yeast mRNA. Mol. Cell. Biol. 15, 5983–5990.

69 Duvel, K., Egli, C. M. and Braus, G. H. (1999) A single point mutation in the yeast TRP4 gene affects efficiency of mRNA 3' end processing and alters selection of the poly(A) site. Nucleic Acid Res. 27, 1289–1295.

70 Heidmann, S., Obermaier, B., Vogel, K. and Domdey, H. (1992) Identification of pre-mRNA polyadenylation sites in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12, 4215–4229.

71 Heidmann, S., Schindewolf, C., Stumpf, G. and Domdey, H. (1994) Flexibility and interchangeability of polyadenylation signals in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 14, 4633–4642.

72 Dichtl, B. and Keller, W. (2001) Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. EMBO J. 20, 3197–3209.

73 Graber, J. H., Cantor, C. R., Mohr, S. C. and Smith, T. F. (1999) Genomic detection of new yeast pre-mRNA 3'-end-processing signals. Nucleic Acid Res. 27, 888–894.

74 Graber, J. H., Cantor, C. R., Mohr, S. C. and Smith, T. F. (1999) *In silico* detection of control signals: mRNA 3'-end-processing sequences in diverse species. Proc. Natl. Acad. Sci. USA 96, 14055–14060.

75 Moore, C. L. and Sharp, P. A. (1984) Site-specific polyadenylation in a cell-free reaction. Cell 36, 581–591.

76 Butler, J. S. and Platt, T. (1988) RNA processing generates the mature 3' end of yeast CYC1 messenger RNA *in vitro*. Science 242, 1270–1274.

77 Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelman, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.

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

79 Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H. and Keller, W. (2002) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. Mol. Cell 10, 1139–1150.

80 Walsh, E. P., Lamont, D. J., Beattie, K. A. and Stark, M. J. (2002) Novel interactions of *Saccharomyces cerevisiae* type 1 protein phosphatase identified by single-step affinity purification and mass spectrometry. Biochemistry 41, 2409–2420.

81 He, X., Khan, A. U., Cheng, H., Pappas Jr., D. L., Hampsey, M. and Moore, C. L. (2003) Functional interactions between

the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. Genes Dev. 17, 1030–1042.

82 Dheur, S., Vo le, T. A., Voisinet-Hakil, F., Minet, M., Schmitter, J. M., Lacroute, F., Wyers, F. and Minvielle-Sebastia, L. (2003) Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. EMBO J. 22, 2831–2840.

83 Chen, J. and Moore, C. (1992) Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. Mol. Cell. Biol. 12, 3470–3481.

84 Dantonel, J.-C., Murthy, K. G. K., Manley, J. L. and Tora, L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature 389, 399–402.

85 McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M. and Bentley, D. L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385, 357–361.

86 Christofori, G. and Keller, W. (1988) 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP. Cell 54, 875–889.

87 Gilmartin, G. M., McDevitt, M. A. and Nevins, J. R. (1988) Multiple factors are required for specific RNA cleavage at a poly(A) addition site. Genes Dev. 2, 578–587.

88 Murthy, K. G. K. and Manley, J. L. (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. Genes Dev. 9, 2672–2683.

89 Moore, C. L., Chen, J. and Whoriskey, J. (1988) Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. EMBO J. 7, 3159–3169.

90 Gilmartin, G. M., Fleming, E. S. and Oetjen, J. (1992) Activation of HIV-1 pre-mRNA 3' processing in vitro requires both an upstream element and TAR. EMBO J. 11, 4419–4428.

91 Murthy, K. G. K. and Manley, J. L. (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. J. Biol. Chem. 267, 14804–14811.

92 Stumpf, G. and Domdey, H. (1996) Dependence of yeast pre-mRNA 3'-end processing on CFT1: A sequence homolog of the mammalian AAUAAA binding factor. Science 274, 1517–1520.

93 Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S. and Keller, W. (2002) Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. EMBO J. 21, 4125–4135.

94 Das, B. K., Xia, L., Palandjian, L., Gozani, O., Chyung, Y. and Reed, R. (1999) Chracterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. Mol. Cell. Biol. 19, 6796–6802.

95 Takao, M., Abramic, M., Moos Jr., M., Otrin, V. R., Wootton, J. C., McLenigan, M., Levine, A. S. and Protic, M. (1993) A 127 kDa component of a UV-damaged DNA-binding complex, which is defective in some xeroderma pigmentosum group E patients, is homologous to a slime mold protein. Nucleic Acid Res. 21, 4111–4118.

96 Neuwald, A. F. and Poleksic, A. (2000) PSI-BLAST searches using hidden Markov models of structural repeats: Prediction of an unusual sliding DNA clamp and of beta-propellers in UV-damaged DNA-binding protein. Nucleic Acid Res. 28, 3570–3580.

97 Li, T., Chen, X., Garbutt, K. C., Zhou, P. and Zheng, N. (2006) Structure of DDB1 in complex with a paramyxovirus V protein: Viral hijack of a propeller cluster in ubiquitin ligase. Cell 124, 105–117.

98 Angers, S., Li, T., Yi, X., MacCoss, M. J., Moon, R. T. and Zheng, N. (2006) Molecular architecture and assembly of the DDB1-CUL4A ubiquitin ligase machinery. Nature 443, 590–593.

99 Chanfreau, G., Noble, S. M. and Guthrie, C. (1996) Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). Science 274, 1511–1514.

100 Callebaut, I., Moshous, D., Mornon, J.-P. and de Villartay, J.-P. (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: The beta-CASP family. Nucleic Acid Res. 30, 3592–3601.

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

102 Dominski, Z. (2007) Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. Crit. Rev. Biochem. Mol. Biol. 42, 67–93.

103 Jenny, A., Minvielle-Sebastia, L., Preker, P. J. and Keller, W. (1996) Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. Science 274, 1514–1517.

104 Ma, Y., Pannicke, U., Schwarz, K. and Lieber, M. R. (2002) Hairpin opening and overhand processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108, 781–794.

105 Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A. and de Villartay, J.-P. (2001) Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell 105, 177–186.

106 Mathy, N., Benard, L., Pellegrini, O., Daou, R., Wen, T. and Condon, C. (2007) 5'-to-3' exoribonuclease activity in bacteria: Role of RNase J1 in rRNA maturation and 5' stability of mRNA. Cell 129, 681–692.

107 de la Sierra-Gallay, I. L., Mathy, N., Pellegrini, O. and Condon, C. (2006) Structure of the ubiquitous 3' processing enzyme RNase Z bound to transfer RNA. Nat. Struct. Mol. Biol. 13, 376–377.

108 Ishii, R., Minagawa, A., Takaku, H., Takagi, M., Nashimoto, M. and Yokoyama, S. (2005) Crystal structure of the tRNA 3' processing endoribonuclease tRNase Z from Thermotoga maritima. J. Biol. Chem. 280, 14138–14144.

109 Ryan, K., Calvo, O. and Manley, J. L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA 10, 565–573.

110 Dominski, Z., Yang, X.-C. and Marzluff, W. F. (2005) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell 123, 37–48.

111 Zhelkovsky, A. M., Tacahashi, Y., Nasser, T., He, X., Sterzer, U., Jensen, T. H., Domdey, H. and Moore, C. L. (2006) The role of Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in Saccharomyces cerevisiae. RNA 12, 435–445.

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

113 Ishikawa, H., Nakagawa, N., Kuramitsu, S. and Masui, R. (2006) Crystal structure of TTHA0252 from Thermus thermophilus HB8, a RNA degradation protein of the metallo-beta-lactamase superfamily. J. Biochem. 140, 535–542.

114 Ryan, K. (2007) Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. RNA Biol. 4, 26–33.

115 Zhao, J., Kessler, M. M., Helmling, S., O'Connor, J. P. and Moore, C. L. (1999) Pta1, a component of yeast CFII, is required for both cleavage and poly(A) addition of mRNA precursor. Mol. Cell. Biol. 19, 7733–7740.

116 Karkashon, S., Hopkinson, A. and Levinger, L. (2007) tRNase Z catalysis and conserved residues on the carboxy side of the His cluster. Biochemistry 46, 9380–9387.

117 Dominski, Z., Yang, X.-C., Purdy, M., Wagner, E. J. and Marzluff, W. F. (2005) 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. 25, 1489–1500.

118 Kyburz, A., Sadowski, M., Dichtl, B. and Keller, W. (2003) The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. Nucleic Acid Res. 31, 3936–3945.

119 de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H. and Keller, W. (2000) Human pre-mRNA cleavage factor IIm contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J. 19, 5895–5904.

120 Baillat, D., Hakimi, M.-A., Naar, A. M., Shilatifard, A., Cooch, N. and Shiekhattar, R. (2005) Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell 123, 265–276.

121 Preker, P. J., Ohnacker, M., Minvielle-Sebastia, L. and Keller, W. (1997) A multisubunit 3' end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. EMBO J. 16, 4727–4737.

122 Zhao, J., Kessler, M. M. and Moore, C. L. (1997) Cleavage factor II of Saccharomyces cerevisiae contains homologues to subunits of the mammalian cleavage/polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. J. Biol. Chem. 272, 10831–10838.

123 Barabino, S. M. L., Ohnacker, M. and Keller, W. (2000) Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. EMBO J. 19, 3778–3787.

124 Barabino, S. M. L., Hubner, W., Jenny, A., Minvielle-Sebastia, L. and 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.

125 Tacahashi, Y., Helmling, S. and Moore, C. L. (2003) Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. Nucleic Acid Res. 31, 1744–1752.

126 Kaufmann, I., Martin, G., Friedlein, A., Langen, H. and 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.

127 Noah, D. L., Twu, K. Y. and Krug, R. M. (2003) Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3' end processing of cellular pre-mRNAs. Virology 307, 386–395.

128 Nemeroff, M. E., Barabino, S. M. L., Li, Y., Keller, W. and Krug, R. M. (1998) Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. Mol. Cell 1, 991–1000.

129 Li, Y., Chen, Z.-Y., Wang, W., Baker, C. C. and Krug, R. M. (2001) The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. RNA 7, 920–931.

130 Bai, C. and Tolias, P. P. (1996) Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger protein. Mol. Cell. Biol. 16, 6661–6667.

131 Zarudnaya, M. I., Kolomiets, I. M. and Hovorun, D. M. (2002) What nuclease cleaves pre-mRNA in the process of polyadenylation? IUBMB Life 54, 27–31.

132 Preker, P. J., Lingner, J., Minvielle-Sebastia, L. and Keller, W. (1995) The FIP1 gene encodes a component of a yeast premRNA polyadenylation factor that directly interacts with poly(A) polymerase. Cell 81, 379–389.

133 Venkataraman, K., Brown, K. M. and Gilmartin, G. M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. 19, 1315–1327.

134 Helmling, S., Zhelkovsky, A. M. and Moore, C. L. (2001) Fip1 regulates the activity of poly(A) polymerase through multiple interactions. Mol. Cell. Biol. 21, 2026–2037.

135 Zhelkovsky, A. M., Helmling, S. and Moore, C. L. (1998) Processivity of the Saccharomyces cerevisiae poly(A) polymerase requires interactions at the carboxyl-terminal RNA binding domain. Mol. Cell. Biol. 18, 5942–5951.

136 Wilusz, J. and Shenk, T. (1988) A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell 52, 221–228.

137 Canadillas, J. M. P. and Varani, G. (2003) Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. EMBO J. 22, 2821–2830.

138 Deka, P., Rajan, P. K., Canadillas, J. M. P. and Varani, G. (2005) 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. 347, 719–733.

139 Gross, S. and Moore, C. L. (2001) Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. Mol. Cell. Biol. 21, 8045–8055.

140 Hatton, L. S., Eloranta, J. J., Figueiredo, L. M., Takagaki, Y., Manley, J. L. and O'Hare, K. (2000) 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 Acid Res. 28, 520–526.

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

142 Qu, X., Perez-Canadillas, J.-M., Agrawal, S., de Baecke, J., Cheng, H., Varani, G. and Moore, C. L. (2007) 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. 282, 2101–2115.

143 Richardson, J. M., McMahon, K. W., MacDonald, C. C. and Makhatadze, G. I. (1999) MEARA sequence repeat of human CstF-64 polyadenylation factor is helical in solution. A spectroscopic and calorimetric study. Biochemistry 38, 12869–12875.

144 Wallace, A. M., Dass, B., Ravnik, S. E., Tonk, V., Jenkins, N. A., Gilbert, D. J., Copeland, N. G. and MacDonald, C. C. (1999) Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells. Proc. Natl. Acad. Sci. USA 96, 6763–6768.

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

146 Dass, B., McDaniel, L., Schultz, R. A., Attaya, E. N. and MacDonald, C. C. (2002) The Gene CSTF2T, encoding the human variant CstF-64 polyadenylation protein tCstF-64, lacks introns and may be associated with male sterility. Genomics 80, 509–514.

147 Wallace, A. M., Denison, T. L., Attaya, E. N. and MacDonald, C. C. (2004) Developmental distribution of the polyadenylation protein CstF-64 and the variant tCstF-64 in mouse and rat testis. Biol. Reprod. 70, 1080–1087.

148 Monarez, R. R., MacDonald, C. C. and Dass, B. (2007) Polyadenylation protein CstF-64 and tauCstF-64 exhibit different binding affinities for RNA polymers. Biochem. J. 401, 651–658.

149 Juge, F., Audibert, A., Benoit, B. and Simonelig, M. (2000) Tissue-specific autoregulation of Drosophila suppressor of forked by alternative poly(A) site utilization leads to accumulation of the suppressor of forked protein in mitotically active cells. RNA 6, 1529–1538.

150 Benoit, B., Juge, F., Iral, F., Audibert, A. and Simonelig, M. (2002) Chimeric human CstF-77/Drosophila suppressor of forked proteins rescue suppressor of forked mutant lethality and mRNA 3' end processing in Drosophila. Proc. Natl. Acad. Sci. USA 99, 10593–10598.

151 Preker, P. J. and Keller, W. (1998) The HAT helix, a repetitive motif implicated in RNA processing. Trends Biochem. Sci. 23, 15–16.

152 Bai, Y., Auperin, T. C., Chou, C.-Y., Chang, G.-G., Manley, J. L. and Tong, L. (2007) Crystal structure of murine CstF-77: Dimeric association and implications for polyadenylation of mRNA precursors. Mol. Cell 25, 863–875.

153 Legrand, P., Pinaud, N., Minvielle-Sebastia, L. and Fribourg, S. (2007) The structure of CstF-77 homodimer provides insights into CstF assembly. Nucleic Acid Res. 35, 4515–4522.

154 Bai, Y., Auperin, T. C. and Tong, L. (2007) The use of in situ proteolysis in the crystallization of murine CstF-77. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 63, 135–138.

155 Noble, C. G., Walker, P. A., Calder, L. J. and Taylor, I. A. (2004) Rna14-Rna15 assembly mediates the RNA-binding capability of Saccharomyces cerevisiae cleavage factor IA. Nucleic Acid Res. 32, 3364–3375.

156 Simonelig, M., Elliott, K., Mitchelson, A. and O'Hare, K. (1996) Interallelic complementation at the suppressor of forked locus of Drosophila reveals complementation between Suppressor of forked proteins mutated in different regions. Genetics 142, 1225–1235.

157 Sadowski, M., Dichtl, B., Hubner, W. and Keller, W. (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. EMBO J. 22, 2167–2177.

158 Takagaki, Y. and Manley, J. L. (1994) A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. Nature 372, 471–474.

159 McCracken, S., Longman, D., Johnstone, I. L., Caceres, J. F. and Blencowe, B. J. (2003) An evolutionarily conserved role for SRm160 in 3'-end processing that functions independently of exon junction complex formation. J. Biol. Chem. 278, 44153–44160.

160 Kleiman, F. E. and Manley, J. L. (1999) Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. Science 285, 1576–1579.

161 Kleiman, F. E. and Manley, J. L. (2001) The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. Cell 104, 743–753.

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

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

164 Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W. and Barabino, S. M. L. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. J. Biol. Chem. 279, 35788–35797.

165 Kielkopf, C. L., Rodionova, N. A., Green, M. R. and Burley, S. K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. Cell 106, 595–605.

166 Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol. Cell 11, 965–976.

167 Fribourg, S., Gatfield, D., Izaurralde, E. and Conti, E. (2003) A novel mode of RBD-protein recognition in the Y14-Mago complex. Nat. Struct. Biol. 10, 433–439.

168 Rappsilber, J., Ryder, U., Lamond, A. I. and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. Genome Res. 12, 1231–1245.

169 Zhou, Z., Sim, J., Griffith, J. and Reed, R. (2002) Purification and electron microscopic visualization of functional human spliceosomes. Proc. Natl. Acad. Sci. USA 99, 12203–12207.

170 Brown, K. M. and Gilmartin, G. M. (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol. Cell 12, 1467–1476.

171 Ruegsegger, U., Beyer, K. and Keller, W. (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. J. Biol. Chem. 271, 6107–6113.

172 Barilla, D., Lee, B. A. and Proudfoot, N. J. (2001) Cleavage/polyadenylation factor IA associates with the carboxy-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 98, 445–450.

173 Zhang, Z. and Gilmour, D. S. (2006) Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol. Cell 21, 65–74.

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

175 Noble, C. G., Hollingworth, D., Martin, S. R., Ennis-Adeniran, V., Smerdon, S. J., Kelly, G., Taylor, I. A. and Ramos, A. (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. Nat. Struct. Mol. Biol. 12, 144–151.

176 Zhang, Z., Fu, J. and Gilmour, D. S. (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. Genes Dev. 19, 1572–1580.

177 Hollingworth, D., Noble, C. G., Taylor, I. A. and Ramos, A. (2006) RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. RNA 12, 555–560.

178 Gross, S. and Moore, C. (2001) Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I. Proc. Natl. Acad. Sci. USA 98, 6080–6085.

179 Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

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

181 Weitzer, S. and Martinez, J. (2007) The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. Nature 447, 222–226.

182 Bard, J., Zhelkovsky, A. M., Helmling, S., Earnest, T. N., Moore, C. L. and Bohm, A. (2000) Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. Science 289, 1346–1349.

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

184 Kyriakopoulou, C. B., Nordvarg, H. and Virtanen, A. (2001) A novel nuclear human poly(A) polymerase (PAP), PAP gamma. J. Biol. Chem. 276, 33504–33511.

185 Martin, G., Moglich, A., Keller, W. and Doublie, S. (2004) Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase. J. Mol. Biol. 341, 911–925.

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

187 Balbo, P. B. and Bohm, A. (2007) Mechanism of poly(A) polymerase: Structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. Structure 15, 1117–1131.

188 Balbo, P. B., Meinke, G. and Bohm, A. (2005) Kinetic studies of yeast polyA polymerase indicate an induced fit mechanism for nucleotide specificity. Biochemistry 44, 7777–7786.

189 Balbo, P. B., Toth, J. and Bohm, A. (2007) X-ray crystallographic and steady state fluorescence characterization of the protein dynamics of yeast polyadenylate polymerase. J. Mol. Biol. 366, 1401–1415.

190 Mandel, C. R. and Tong, L. (2007) How to get all "A"s in polyadenylation. Structure 15, 1024–1026.

191 Doublie, S., Sawaya, M. R. and Ellenberger, T. (1999) An open and closed case for all polymerases. Structure 7, R31-R35.

192 Amrani, N., Minet, M., Le Gouar, M., Lacroute, F. and Wyers, F. (1997) Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length *in vitro*. Mol. Cell. Biol. 17, 3694–3701.

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

194 Minvielle-Sebastia, L., Preker, P. J., Wiederkehr, T., Strahm, Y. and Keller, W. (1997) The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. Proc. Natl. Acad. Sci. USA 94, 7897–7902.

195 Sachs, A. B. and Davis, R. W. (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58, 857–867.

196 Meyer, S., Urbanke, C. and Wahle, E. (2002) Equilibrium studies on the association of the nuclear poly(A) binding protein with poly(A) of different lengths. Biochemistry 41, 6082–6089.

197 Keller, R. W., Kuhn, U., Aragon, M., Bornikova, L., Wahle, E. and Bear, D. G. (2000) The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J. Mol. Biol. 297, 569–583.

198 Kerwitz, Y., Kuhn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E. and Wahle, E. (2003) Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J. 22, 3705–3714.

199 Deo, R. C., Bonanno, J. B., Sonenberg, N. and Burley, S. K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell 98, 835–845.

200 O'Connor, J. P. and Peebles, C. L. (1992) PTA1, an essential gene of *Saccharomyces cerevisiae* affecting pre-tRNA processing. Mol. Cell. Biol. 12, 3843–3856.

201 Keon, B. H., Schafer, S., Kuhn, C., Grund, C. and Franke, W. W. (1996) Symplekin, a novel type of tight junction plaque protein. J. Cell. Biol. 134, 1003–1018.

202 Dichtl, B., Aasland, R. and Keller, W. (2004) Functions for *S. cerevisiae* Swd2p in 3' end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. RNA 10, 965–977.

203 Skaar, D. A. and Greenleaf, A. L. (2002) The RNA polymerase II CTD kinase CTDK-1 affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. Mol. Cell 10, 1429–1439.

204 He, X. and Moore, C. L. (2005) Regulation of yeast mRNA 3' end processing by phosphorylation. Mol. Cell 19, 619–629.

205 Kolev, N. G. and Steitz, J. A. (2005) Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. Genes Dev. 19, 2583–2592.

206 Hofmann, I., Schnolzer, M., Kaufmann, I. and Franke, W. W. (2002) Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in *Xenopus laevis* oocytes. Mol. Biol. Cell 13, 1665–1676.

207 Phatnani, H. P. and Greenleaf, A. L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 20, 2922–2936.

208 Cramer, P., Bushnell, D. A. and Kornberg, R. D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863–1876.

209 Fabrega, C., Shen, V., Shuman, S. and Lima, C. D. (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell 11, 1549–1561.

210 Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. and Noel, J. P. (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat. Struct. Biol. 7, 639–643.

211 Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. and Cramer, P. (2005) A structural perspective of CTD function. Genes Dev. 19, 1401–1415.

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

213 Licatalosi, D. D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B. and Bentley, D. L. (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol. Cell 9, 1101–1111.

214 McNeil, J. B., Agah, H. and Bentley, D. (1998) Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast. Genes Dev. 12, 2510–2521.

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

216 Perez-Canadillas, J.-M. (2006) Grabbing the message: Structural basis of mRNA 3' UTR recognition by Hrp1. EMBO J. 25, 3167–3178.

217 Kessler, M. M., Henry, M. F., Shen, E., Zhao, J., Gross, S., Silver, P. A. and Moore, C. L. (1997) 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. 11, 2545–2556.

218 Wang, S. W., Asakawa, K., Win, T. Z., Toda, T. and Norbury, C. J. (2005) Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects. Mol. Cell. Biol. 25, 2288–2296.

219 Ohnacker, M., Barabino, S. M. L., Preker, P. J. and Keller, W. (2000) The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. EMBO J. 19, 37–47.

220 Ito, S., Sakai, A., Nomura, T., Miki, Y., Ouchida, M., Sasaki, J. and Shimizu, K. (2001) A novel WD40 repeat protein, WDC146, highly expressed during spermatogenesis in a stage-specific manner. Biochem. Biophys. Res. Commun. 280, 656–663.

221 Sun, Z. W. and Hampsey, M. (1996) Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection *in vivo*. Mol. Cell. Biol. 16, 1557–1566.

222 Pappas Jr., D. L. and Hampsey, M. (2000) Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 20, 8343–8351.

223 Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M. and Shuman, S. (2005) Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. J. Biol. Chem. 280, 37681–37688.

224 Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C. L. and Hampsey, M. (2004) Ssu72 is an RNA polymerase II CTD phosphatase. Mol. Cell 14, 387–394.

225 St-Pierre, B., Liu, X., Kha, L. C., Zhu, X., Ryan, O., Jiang, Z. and Zacksenhaus, E. (2005) Conserved and specific functions of mammalian ssu72. Nucleic Acid Res. 33, 464–477.

226 Cheng, H., He, X. and Moore, C. L. (2004) The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. Mol. Cell. Biol. 24, 2932–2943.

227 Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F. and Shilatifard, A. (2001) COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. Proc. Natl. Acad. Sci. USA 98, 12902–12907.

228 Vo, L. T., Minet, M., Schmitter, J. M., Lacroute, F. and Wyers, F. (2001) Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Mol. Cell. Biol. 21, 8346–8356.

229 Feng, Z. H., Wilson, S. E., Peng, Z. Y., Schlender, K. K., Reimann, E. M. and Trumbly, R. J. (1991) The yeast GLC7 gene required for glycogen accumulation encodes a type 1 protein phosphatase. J. Biol. Chem. 266, 23796–23801.

230 Garcia-Gimeno, M. A., Munoz, I., Arino, J. and Sanz, P. (2003) Molecular characterization of Ypi1, a novel *Saccha-*

romyces cerevisiae type 1 protein phosphatase inhibitor. J. Biol. Chem. 278, 47744–47752.

231 Russnak, R., Nehrke, K. W. and Platt, T. (1995) REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation. Mol. Cell. Biol. 15, 1689–1697.

To access this journal online:
http://www.birkhauser.ch/CMLS