

# Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition

Krishnan Venkataraman, Kirk M. Brown,<sup>1</sup> and Gregory M. Gilmartin<sup>2</sup>

Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, Vermont 05405, USA

At least half of all human pre-mRNAs are subject to alternative 3' processing that may modulate both the coding capacity of the message and the array of post-transcriptional regulatory elements embedded within the 3' UTR. Vertebrate poly(A) site selection appears to rely primarily on the binding of CPSF to an A(A/U)UAAA hexamer upstream of the cleavage site and CstF to a downstream GU-rich element. At least one-quarter of all human poly(A) sites, however, lack the A(A/U)UAAA motif. We report that sequence-specific RNA binding of the human 3' processing factor CFI<sub>m</sub> can function as a primary determinant of poly(A) site recognition in the absence of the A(A/U)UAAA motif. CFI<sub>m</sub> is sufficient to direct sequence-specific, A(A/U)UAAA-independent poly(A) addition *in vitro* through the recruitment of the CPSF subunit hFip1 and poly(A) polymerase to the RNA substrate. ChIP analysis indicates that CFI<sub>m</sub> is recruited to the transcription unit, along with CPSF and CstF, during the initial stages of transcription, supporting a direct role for CFI<sub>m</sub> in poly(A) site recognition. The recognition of three distinct sequence elements by CFI<sub>m</sub>, CPSF, and CstF suggests that vertebrate poly(A) site definition is mechanistically more similar to that of yeast and plants than anticipated.

[Keywords: mRNA 3' processing; polyadenylation; poly(A) site recognition]

Received January 14, 2005; revised version accepted April 21, 2005.

The process of mRNA 3' end formation is not simply a perfunctory step in eukaryotic gene expression. At least one-half of all human genes are subject to alternative 3' processing (Iseli et al. 2002), the consequences of which may impact the protein coding capacity of the message, as well as its localization, translation efficiency, and stability (Edwards-Gilbert et al. 1997). Moreover, poly(A) site selection may be modulated in a developmental and tissue-specific manner. In addition, pre-mRNA 3' processing contributes directly to transcription termination (Zorio and Bentley 2004), pre-mRNA splicing (Proudfoot et al. 2002), and mRNA export (Hammell et al. 2002; Lei and Silver 2002). While the processing of constitutive poly(A) sites has been examined in considerable detail, the fundamental mechanisms responsible for the regulation of alternative poly(A) site selection have yet to be fully elucidated (Barabino and Keller 1999).

The processing of the majority of human poly(A) sites involves the recognition of an AAUAAA or AUUAAA hexamer by CPSF, coupled with the binding of CstF to a GU-rich downstream element (DSE) (Zhao et al. 1999). The binding of CPSF and CstF appears to be sufficient, at

least *in vitro*, to direct the assembly of a 3' processing complex composed of at least 14 different proteins. *In vivo*, however, the hexamer and DSE alone are unlikely to suffice for poly(A) site definition. The recognition of an authentic poly(A) site within a nascent RNA *in vivo* appears to rely on the "biosynthetic context" provided by the transcription elongation complex (Proudfoot 2004). At least nine 3' processing proteins are recruited to the transcription complex, at least in part through interactions with the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNAPII) (Calvo and Manley 2003). The colocalization of 3' processing factors, along with capping enzymes and spliceosome components, to the transcription elongation complex, allows for the cooperative interaction of these processing machineries within an "mRNA factory" (Zorio and Bentley 2004).

Cotranscriptional recognition of a poly(A) site provides an elegant mechanism for the identification of a processing site demarcated by a limited set of sequence motifs. Yet the mechanisms that regulate the selection of alternative poly(A) sites within a pre-mRNA, or allow for the recognition of poly(A) sites that lack the canonical A(A/U)UAAA motif, are poorly understood. Sequences that function to enhance the efficiency of 3' processing have been identified upstream of several canonical poly(A) sites (Zhao et al. 1999). Such elements might contribute to the regulation of poly(A) site selec-

<sup>1</sup>Present address: Department of Biochemistry and Molecular Pharmacology, University of Massachusetts, Worcester, MA 01605, USA.

<sup>2</sup>Corresponding author.

E-MAIL ggilmart@zoo.uvm.edu; FAX (802) 656-8749.

Article and publication are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1298605>.

tion, although the molecular mechanisms by which they function are largely unclear. In this work, we have identified a mechanism by which the 3' processing factor CFI<sub>m</sub> contributes to poly(A) site recognition and, specifically, to the recognition of a human poly(A) site that lacks the canonical A(A/U)UAAA hexamer.

CFI<sub>m</sub> is an essential, heterodimeric pre-mRNA 3' processing factor, unique to metazoans, composed of a small subunit of 25 kDa and a large subunit of 59, 68, or 72 kDa (Ruegsegger et al. 1996). The human CFI<sub>m</sub> 59- and 68-kDa subunits are encoded by paralogous genes, while the nature of the 72-kDa protein remains unclear. The CFI<sub>m</sub> 59- and 68-kDa proteins possess an N-terminal RNP-type RNA-binding domain (RBD), a central proline-rich domain, and a C-terminal RS-like alternating charge domain—a structure similar to that of the SR family of proteins that function in basal and regulated pre-mRNA splicing (Graveley 2000). The structure of the CFI<sub>m</sub> large subunit, along with the identification of both the 25- and 68-kDa subunits as components of human spliceosomes (Rappsilber et al. 2002; Zhou et al. 2002), suggests a potential role for CFI<sub>m</sub> in the coordination of 3' processing and pre-mRNA splicing. Such a role is supported by the specific interaction of the 25-kDa subunit with the 70-kDa protein of U1 snRNP (Awasthi and Alwine 2003) and the interaction of the 68-kDa subunit with the SR proteins Srp20, 9G8, and hTra-2β (Dettwiler et al. 2004). The CFI<sub>m</sub> 68/25-kDa heterodimer has been shown to be sufficient to reconstitute CFI<sub>m</sub> function in vitro (Ruegsegger et al. 1998), and SELEX analysis has indicated that the 68/25-kDa heterodimer preferentially binds the sequence UGUAN (N = A > U ≥ C/G) (Brown and Gilmartin 2003). Ruegsegger et al. (1998) demonstrated that CFI<sub>m</sub> can function to facilitate pre-mRNA 3' processing complex assembly and to enhance the rate and overall efficiency of poly(A) site cleavage in vitro. In vitro experiments have also shown that CFI<sub>m</sub> can function to autoregulate the 3' processing of the pre-mRNA encoding the CFI<sub>m</sub> 68-kDa subunit through the binding of a set of UGUAA elements that flank and overlap the AAUAAA hexamer (Brown and Gilmartin 2003).

In this report, we demonstrate that CFI<sub>m</sub> has a fundamental role in poly(A) site recognition. Sequence-specific RNA binding of CFI<sub>m</sub> was found to function as the primary determinant for the recognition of a human poly(A) site lacking the A(A/U)UAAA hexamer. Furthermore, we show that CFI<sub>m</sub> can function to direct sequence-specific, A(A/U)UAAA-independent poly(A) addition in vitro, through its ability to recruit the CPSF subunit hFip1 and poly(A) polymerase to the RNA substrate. A direct role for CFI<sub>m</sub> in poly(A) site recognition is supported by chromatin immunoprecipitation (ChIP) analysis indicating that CFI<sub>m</sub> is recruited to the transcription unit, along with CPSF and CstF, during the initial stages of transcription. Taken together, the data indicate that three sequence-specific RNA-binding factors participate in vertebrate poly(A) site recognition, CFI<sub>m</sub>, CPSF, and CstF, and suggest that the mechanisms of poly(A) site recognition in yeast, plants, and verte-

brates are more similar than previously appreciated (Graber et al. 1999; Proudfoot and O'Sullivan 2002).

## Results

### *PAPOLA and PAPOLG gene paralogs as a model system for the analysis of poly(A) site recognition*

EST database analysis has indicated that over a quarter of all human transcripts are processed at noncanonical poly(A) sites that deviate in one or more positions from the A(A/U)AAA consensus (Beaudoing et al. 2000). Mutagenesis studies have indicated that such noncanonical sites are processed poorly, if at all, in vitro (Sheets et al. 1990). To begin to understand the mechanisms that allow for the recognition of noncanonical poly(A) sites, we chose to examine the noncanonical poly(A) site of the human *PAPOLG* gene (also referred to as neo-poly(A) polymerase [Topalian et al. 2001]) in parallel with the canonical poly(A) site of its paralog, *PAPOLA*. The primary poly(A) site of *PAPOLG* (as determined by EST analysis) does not possess a sequence with more than a 4-nt match to the A(A/U)UAAA motif. The *PAPOLA* and *PAPOLG* genes, which encode poly(A) polymerase α (PAPα) and poly(A) polymerase γ (PAPγ), respectively, derive from gene duplication, as evidenced by a common intron/exon structure. PAPα and PAPγ have an overall amino acid sequence similarity of 71%.

As illustrated in Figure 1, the 3' ends of *PAPOLA* and *PAPOLG* are strikingly distinct and highly conserved among vertebrates. The conserved sequences upstream of each poly(A) site encompass multiple copies of potential CFI<sub>m</sub>-binding sites of the form UGUAN (N = A > U ≥ C/G). As denoted by the boxed sequences, the human *PAPOLA* poly(A) site contains six UGUAN elements within 131 nucleotides (nt) upstream of the cleavage site. The *PAPOLG* poly(A) site contains seven UGUAN elements within 119 nt upstream of the cleavage site in a pattern that is clearly distinct from that of the *PAPOLA* poly(A) site. The *PAPOLG* gene contains a single UGUAN element within 130 nt downstream of the cleavage site, while the *PAPOLA* gene has none. The following experiments were undertaken to test the hypothesis that sequence-specific binding of CFI<sub>m</sub> to the UGUAN elements of *PAPOLA* and *PAPOLG* contributes to poly(A) site recognition and processing.

### *CFI<sub>m</sub>-binding sites enhance 3' processing at the canonical PAPOLA poly(A) site in vitro*

We first addressed the contribution of the four conserved UGUAN elements immediately upstream of the *PAPOLA* AAUAAA hexamer to poly(A) site function. Single G-to-C point mutations were introduced into the two UGUAN elements distal to the hexamer (PAPαΔ1), the two UGUAN elements proximal to the hexamer (PAPαΔ2), or the combination of all four elements (PAPαΔ1/2) (Fig. 2A). In addition, a single U-to-G point mutation was introduced into the AAUAAA hexamer (PAPαΔhex). As illustrated in Figure 2B, lane 1, the wild-type *PAPOLA* poly(A) site was efficiently cleaved in

## A

## PAPOLA

Hs	<b>TGTAAGT</b> AGTGTCTCATGACTGGAGTTTGCTTTGTTTTATAGTATC <b>TGTAC</b> TCCT <b>TGTAT</b> TTTTCAAGA
Mm	<b>TGTAAGT</b> AGTGTCTCATGACTGGAGTTTGCTTTGTTTTATAGTATC <b>TGTAC</b> TCCT <b>TGTAT</b> TTTTCAAGA
Gg	<b>TGTACGT</b> AGTGTCTCATGACTGGAGTTTGCTTTGTTTTATAGTATC <b>TGTAC</b> TCCT <b>TGTAT</b> TTTTCAAGA
Xl	<b>TGTAAGT</b> AGTGTCTCATGACTGGAGTTTGCTTTGTTTTATAGTATC <b>TGTAC</b> TCCT <b>TGTAT</b> TTTTCAAGA

Hs	GCTATTT <b>TGTA</b> ACAGATGA <b>TGTAT</b> TTCTCCATT GAAAAACAC <b>AA</b> TAAAA <b>AAAA</b> CAGCACAA*
Mm	GCTATTT <b>TGTA</b> ACAGATGA <b>TGTAT</b> TTCTCCATT GAAAAACAC <b>AA</b> TAAAA <b>AAAA</b> CAGCACAA*
Gg	GCTATTT <b>TGTA</b> ACAGATGA <b>TGTAT</b> TTCTCCATT GAAAAACAC <b>AA</b> TAAAA <b>AAAA</b> CAGCACAA*
Xl	TCTATTT <b>TGTA</b> ACAGATGA <b>TGTAT</b> TTCTCCATTGGAAAAA <b>AA</b> TAAAA <b>AA</b> C CAGCACAA*

## B

## PAPOLG

Hs	<b>TGTA</b> AATCTCT <b>TGTA</b> AATGTTGAT TCAG TTG TT <b>TGTA</b> AATG AAGTCGTA <b>TGTAT</b> TTTC AGAG
Mm	<b>TGTA</b> AATCTCT <b>TGTA</b> AATGTTGAC TCCG TTG TT <b>TGTA</b> AATG AAGCCGTA <b>TGTAT</b> TTTC AGAG
Gg	TATGTT <b>TCTAC</b> <b>TGTA</b> AATCTCTTATAATCTGGT <b>TGTAA</b> GTG AAGTTGAA <b>TGTC</b> TTTCCAGAG
Dr	TG CAAAAA <b>A</b> TATATCTACTTG TCAT TCT TT <b>TGTAG</b> GGCAAAGT <b>TGTAT</b> TTGAAAGGG

Hs	TATTTT <b>TGTATGTA</b> CT <b>TGTAA</b> GATACCATCTTTTCAAAGAGAAACGTTTAAAAACCTTTA*
Mm	TATTTT <b>TGTATGTA</b> CT <b>TGTAA</b> GATATCATCTTTTCAAAGAGAAAC TTTAAAAACCTTTA*
Gg	TATTTT <b>TGTATGTA</b> CT <b>TGTAA</b> ATACATCTTT CCAAAGA AATC TAAAAA*
Dr	TATTTT <b>TGTATGTA</b> CT <b>TGTAA</b> ACTGCCCTTCTTTACATTAAGAA C <b>AT</b> AAAAA TTCAAAAACCCCA*

**Figure 1.** Sequence comparison of the 3' ends of vertebrate *PAPOLA* and *PAPOLG* genes. (A) Comparison of the sequences of the *PAPOLA* genes of *Homo sapiens* (Hs), *Mus musculus* (Mm), *Gallus gallus* (Gg), and *Xenopus laevis* (Xl) immediately upstream of the primary poly(A) cleavage site. (B) Comparison of the sequences of the *PAPOLG* genes of *Homo sapiens* (Hs), *Mus musculus* (Mm), *Gallus gallus* (Gg), and *Danio rerio* (Dr) immediately upstream of the primary poly(A) cleavage site. Shaded sequences denote identity to the human sequence. TGTAN elements are highlighted in bold and outlined with solid boxes; AATAAA-related elements are highlighted in bold and outlined with dashed boxes. Asterisk denotes the poly(A) cleavage site.

HeLa cell nuclear extract to yield the expected 5' product. Mutation of the AAUAAA hexamer greatly reduced the overall efficiency of cleavage, although cleavage at multiple adjacent sites was observed (Fig. 2B, lane 2). Mutation of each set of UGUAN elements reduced cleavage and in combination exhibited an additive reduction in cleavage activity (Fig. 2B, lanes 3–5). The efficiency of poly(A) addition to the precleaved PAP $\alpha$ Hex, PAP $\alpha$ Δ2, and PAP $\alpha$ Δ1/2 RNAs (RNAs that extend only to the cleavage site) was also reduced relative to the wild type (Fig. 2C, lanes 2,4,5). As illustrated by the quantitation of four independent sets of experiments (Fig. 2D), the UGUAN elements clearly contribute to the efficiency of both cleavage and poly(A) addition at the *PAPOLA* poly(A) site in vitro.

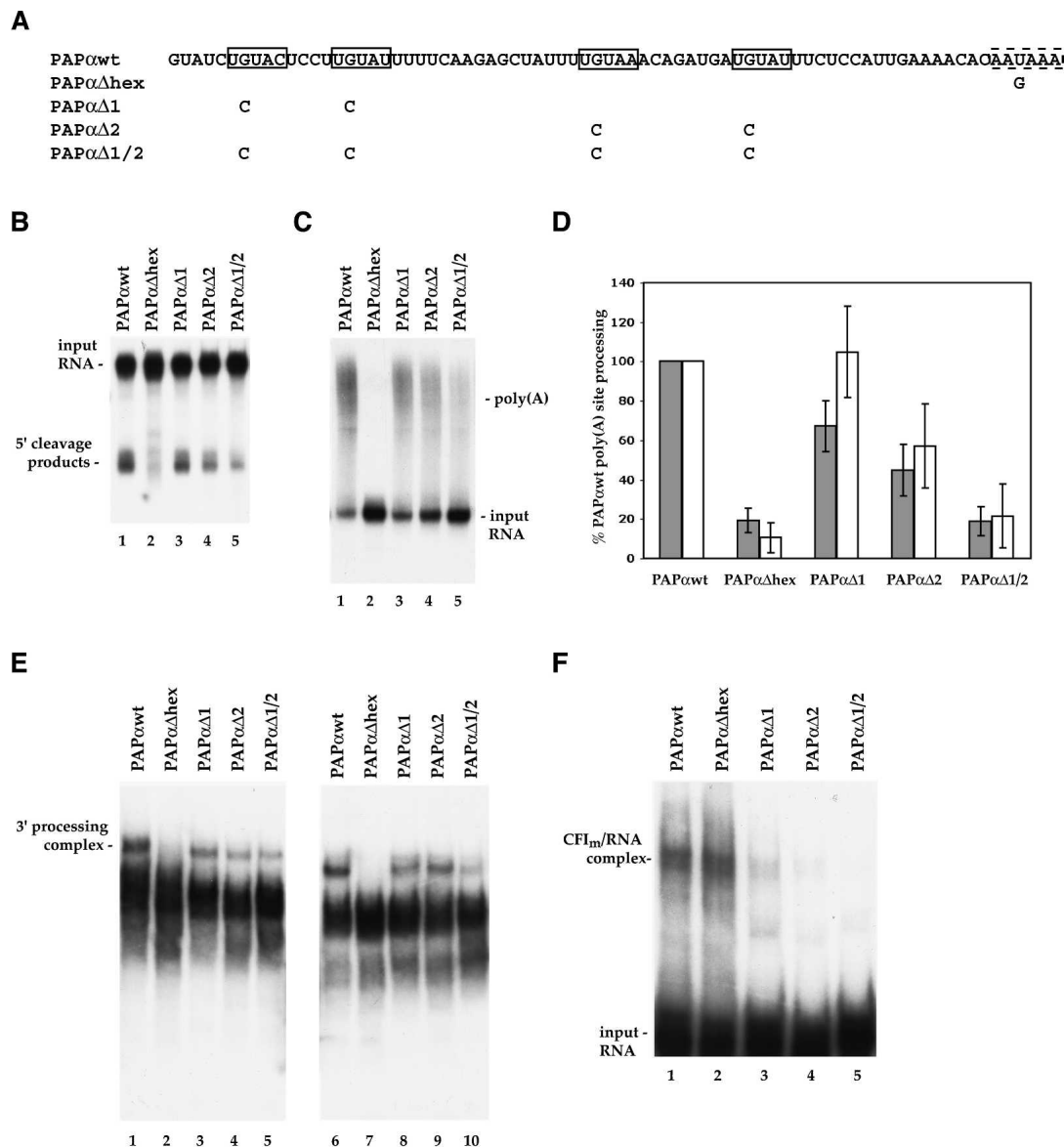
The impact of each of the *PAPOLA* poly(A) site mutations on 3' processing complex assembly and CFI<sub>m</sub> binding is illustrated in Figure 2E and F. As expected, a single point mutation within the AAUAAA hexamer eliminated 3' processing complex formation on both full-length and precleaved *PAPOLA* RNAs in HeLa cell nuclear extract, as assayed by gel mobility shift (Fig. 2E, lanes 2,7). Mutations within the UGUAN elements reduced the efficiency of complex assembly on both full-length (Fig. 2E, lanes 3–5) and precleaved (Fig. 2E, lanes 8–10) RNAs. Point mutations within each set of UGUAN elements also significantly reduced the binding of purified recombinant CFI<sub>m</sub> to the precleaved *PAPOLA* RNA in an additive manner (Fig. 2F, lanes 3,4). Taken together, the data demonstrate that efficient *PAPOLA* 3' processing complex assembly, poly(A) site cleavage, and poly(A) addition are dependent upon a set of UGUAN elements upstream of the AAUAAA hexamer that are targets for CFI<sub>m</sub> binding.

*CFI<sub>m</sub>-binding sites are required for efficient 3' processing at the noncanonical PAPOLG poly(A) in vitro*

We next addressed the role of sequence-specific RNA binding by CFI<sub>m</sub> in the processing of the noncanonical

*PAPOLG* poly(A) site. Single G-to-C point mutations were introduced into the two UGUAN elements distal to the cleavage site (PAP $\gamma$ Δ1), the three UGUAN elements proximal to the cleavage site (PAP $\gamma$ Δ2), or the combination of all five elements (PAP $\gamma$ Δ1/2) (Fig. 3A). A complementary set of RNAs was produced in which a dinucleotide change was introduced to create a canonical AAUAAA sequence 15 nt upstream of the cleavage site (PAP $\gamma$ cswt, PAP $\gamma$ csΔ1, PAP $\gamma$ csΔ2, and PAP $\gamma$ cs Δ1/2). The A-rich sequence centered 20 nt upstream of the cleavage site (AAAGAGAAA) was chosen for mutagenesis based on the observation that the poly(A) sites of both yeast and plants possess similar A-rich sequences that function as positioning elements (Graber et al. 1999). The PAP $\gamma$ ΔA RNA contained two A-to-C mutations within the A-rich element (AAAGAGAAA to AACGAGCAA).

As illustrated in Figure 3B, cleavage at the *PAPOLG* poly(A) site in HeLa cell nuclear extract was significantly reduced by point mutations within each set of UGUAN elements (lanes 3,4), as well as within the A-rich element (lane 2). The combination of both sets of UGUAN element mutations nearly eliminated poly(A) site cleavage in vitro (Fig. 3B, lane 5). The introduction of an AAUAAA sequence significantly enhanced the efficiency of *PAPOLG* poly(A) site cleavage (Fig. 3B, cf. lanes 1 and 6). Even in the context of the canonical hexamer, however, mutation of the three proximal UGUAN elements (Fig. 3B, lane 8) or all five UGUAN elements (Fig. 3B, lane 9) reduced the efficiency of cleavage. Point mutations within the UGUAN elements had a comparable impact on poly(A) addition to precleaved RNA substrates (Fig. 3C). Strikingly, the poly(A) tails added to the PAP $\gamma$ Δ1/2 RNA substrate were consistently longer than those of the other PAP $\gamma$  RNAs (Fig. 3C, cf. lanes 5 and 1–4). As poly(A) tail length control is dependent upon the binding of both CPSF and PABPN1 (Wahle 1995), the extended poly(A) tails of PAP $\gamma$ Δ1/2 may result from the inability of CPSF to interact stably with the RNA as a result of the loss of CFI<sub>m</sub> binding (see below). In HeLa cell nuclear extract, short oligo(A) tails produced by the

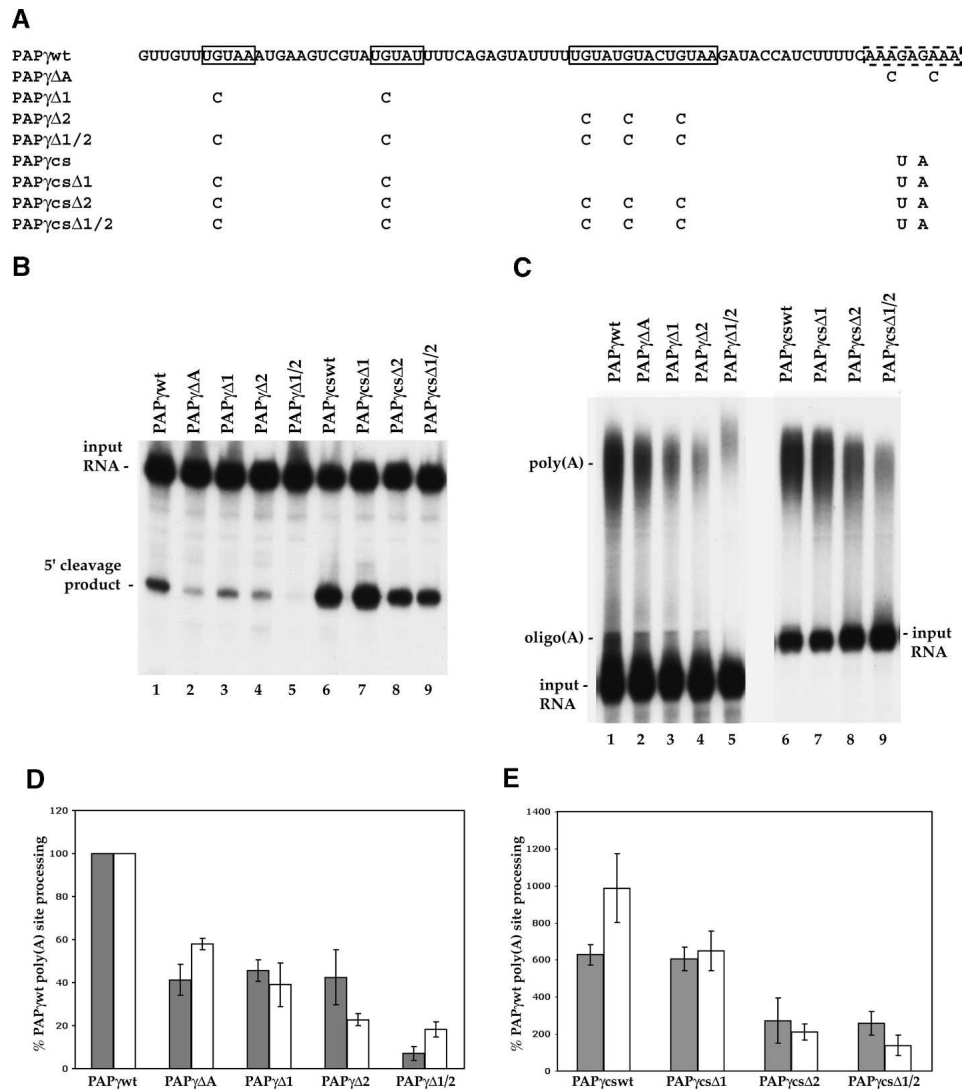


**Figure 2.** Conserved UGUAN elements within the *PAPOLA* pre-mRNA bind CFI<sub>m</sub> and enhance 3' processing. (A) Sequence of the 5' end of the *PAPOLA* RNAs used for in vitro analysis. Each of the RNAs is identical to the wild-type RNA (PAP<sup>α</sup>wt) except at the indicated positions. (B) Poly(A) site cleavage. Uniformly <sup>32</sup>P-labeled RNA substrates were incubated in HeLa cell nuclear extract with 3'dATP for 30 min at 30°C, and the RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. (C) Poly(A) addition. Uniformly <sup>32</sup>P-labeled precleaved RNA substrates were incubated in HeLa cell nuclear extract with ATP for 15 min at 30°C, and the RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. (D) The results of four independent poly(A) site cleavage experiments (shaded bars) and four independent poly(A) addition experiments (open bars) are shown as an average, with the SD shown as error bars. Within each experiment, the cleavage and poly(A) addition efficiencies of each of the RNAs is plotted relative to the efficiency of the PAP<sup>α</sup>wt RNA, which is arbitrarily set to 100%. (E) Gel mobility shift analysis of 3' processing complexes assembled in HeLa cell nuclear extract. Uniformly <sup>32</sup>P-labeled full-length (lanes 1–5) or precleaved (lanes 6–10) RNAs were incubated with HeLa cell nuclear extract for 5 min at 30°C, followed by the addition of heparin to 5 mg/mL, and the RNA/protein complexes were resolved on a nondenaturing 3% polyacrylamide gel. (F) Gel mobility shift analysis of CFI<sub>m</sub>/RNA complexes. Two picomoles of recombinant CFI<sub>m</sub> was incubated with uniformly <sup>32</sup>P-labeled precleaved RNAs for 5 min at 30°C and the RNA/protein complexes were resolved on a nondenaturing 3% polyacrylamide gel.

transient interaction of CPSF with the RNA substrate would be subject to further elongation through the cooperative interaction of PABPN1 and poly(A) polymerase (Kuhn and Wahle 2004). Consistent with this hypothesis, the level of oligoadenylated PAP<sup>γ</sup>Δ1/2 is clearly

lower than that of oligoadenylated PAP<sup>γ</sup>Δ2 RNA (Fig. 3C, cf. lanes 4 and 5). The quantitation of four independent sets of experiments is presented in Figure 4D and E.

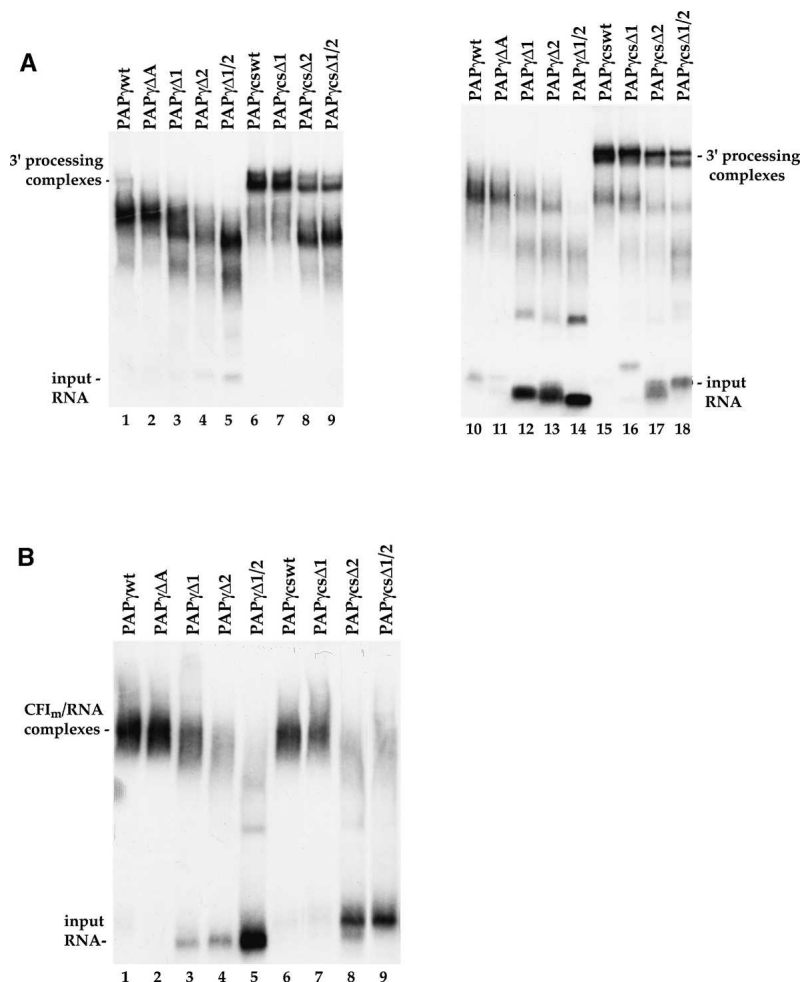
The impact of each of the *PAPOLG* poly(A) site mutations on 3' processing complex assembly and CFI<sub>m</sub>



**Figure 3.** Conserved UGUAN elements within the *PAPOLG* pre-mRNA are required for efficient 3' processing in vitro. (A) Sequence of the 5' end of the RNAs used for in vitro analysis. Each of the RNAs is identical to the wild-type RNA (PAP $\gamma$ wt) except at the indicated positions. (B) Poly(A) site cleavage. Uniformly  $^{32}$ P-labeled RNA substrates were incubated in HeLa cell nuclear extract with 3'dATP for 30 min at 30°C, and the RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. (C) Poly(A) addition. Uniformly  $^{32}$ P-labeled precleaved RNA substrates were incubated in HeLa cell nuclear extract with ATP for 30 min at 30°C, and the RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. Lanes 1–5 were subjected to autoradiography for 15 h, lanes 6–9 for 5 h. (D,E) The results of four independent poly(A) site cleavage experiments (shaded bars) and four independent poly(A) addition experiments (open bars) are shown as an average, with the SD shown as error bars. Within each experiment, the cleavage and poly(A) addition efficiencies of each of the RNAs is plotted relative to the efficiency of the PAP $\gamma$ wt RNA, which is arbitrarily set to 100%.

binding is illustrated in Figure 4. In contrast to the canonical poly(A) site of *PAPOLA*, only a low level of 3' processing complex formation was detected upon incubation of the full-length PAP $\gamma$ wt RNA in HeLa cell nuclear extract (Fig. 4A, lane 1). Point mutations within either the A-rich element (Fig. 4A, lane 2) or the UGUAN elements (Fig. 4A, lanes 3–5) eliminated this complex. Mutations within the UGUAN elements also increased the mobility of the lower complex. Introduction of an AAUAAA hexamer dramatically enhanced 3' processing complex formation in HeLa cell nuclear ex-

tract (Fig. 4A, lanes 6–9), although 3' processing complex formation was reduced by point mutations within the UGUAN elements. A similar pattern was observed in the assembly of complexes on precleaved *PAPOLG* RNAs in HeLa cell nuclear extract (Fig. 4A, lanes 10–18). The differences between the complexes assembled on full-length RNAs and those assembled on precleaved RNAs are most likely attributable to the binding of CstF to the DSE of the full-length RNAs, which stabilizes the interaction of CPSF with the RNA (Gilmartin and Nevins 1991).



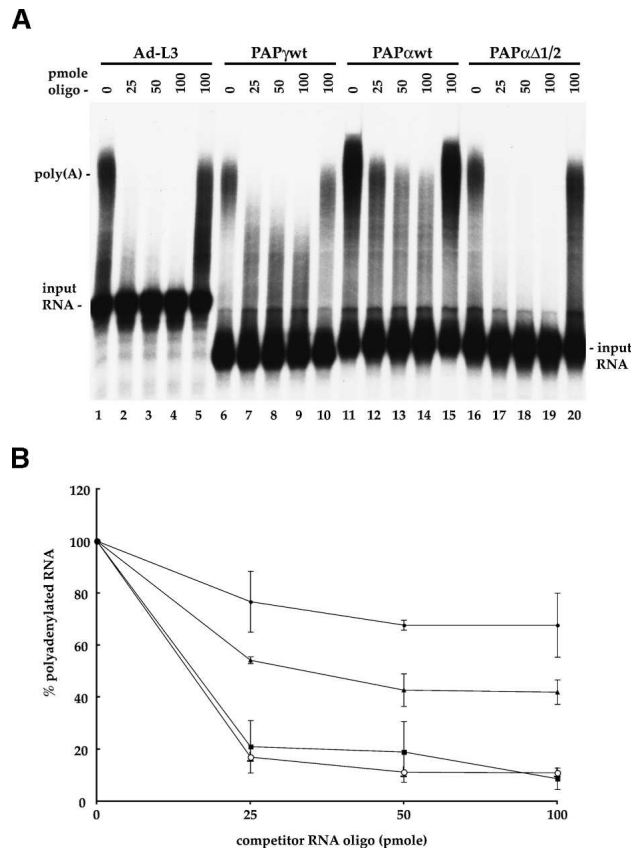
**Figure 4.** UGUAN elements within the *PAPOLG* pre-mRNA bind  $CFI_m$  and contribute to 3' processing complex assembly. (A) Gel mobility shift analysis of 3' processing complexes assembled in HeLa cell nuclear extract. Uniformly  $^{32}P$ -labeled full-length (lanes 1–9) or precleaved (lanes 10–18) RNAs were incubated with HeLa cell nuclear extract for 5 min at 30°C, followed by the addition of heparin to 5 mg/mL, and the RNA/protein complexes were resolved on a nondenaturing 3% polyacrylamide gel. (B) Gel mobility shift analysis of  $CFI_m$ /RNA complexes. Two picomoles of recombinant  $CFI_m$  was incubated with each uniformly  $^{32}P$ -labeled precleaved RNA for 5 min at 30°C and the RNA/protein complexes were resolved on a nondenaturing 3% polyacrylamide gel.

The impact of each of the *PAPOLG* poly(A) site mutations on the binding of recombinant  $CFI_m$  is illustrated in Figure 4B.  $CFI_m$  binding was unaltered by mutations within the A-rich element (Fig. 4B, lanes 2,6), whereas the introduction of point mutations within both sets of UGUAN elements greatly reduced  $CFI_m$  binding (Fig. 4B, lanes 5,9). It should be noted, however, that the combination of single point mutations within each of the five UGUAN elements (Fig. 4B, lanes 5,9) did not completely eliminate the binding of  $CFI_m$ . In the same manner, these mutations greatly reduced, but did not completely eliminate, cleavage and poly(A) addition (Fig. 3D). Taken together, the data demonstrate that *PAPOLG* 3' processing complex assembly, poly(A) site cleavage, and poly(A) addition are dependent upon a set of UGUAN elements that are targets for  $CFI_m$  binding.

*CFI<sub>m</sub> enhances the recruitment of CPSF to the noncanonical PAPOLG poly(A) site*

The observation that sequence-specific RNA binding of  $CFI_m$  was required for efficient *PAPOLG* poly(A) site cleavage and poly(A) addition in the absence of an A(A/U)UAAA hexamer prompted us to ask if CPSF was es-

sential for 3' processing at this poly(A) site. To address this question, we used a 23-nt RNA oligonucleotide encompassing the AAUAAA hexamer of the SV40 late poly(A) site (5'-CUGCAAUAAACAAGUUAACAACA-3') to sequester CPSF in HeLa cell nuclear extract. This RNA oligo is fully competent to bind CPSF and direct specific poly(A) addition in vitro (Wigley et al. 1990). CPSF has previously been shown to be essential for AAUAAA-dependent poly(A) addition in vitro (Bienroth et al. 1993). As illustrated in Figure 5A, addition of the AAUAAA-containing RNA oligo reduced poly(A) addition to the  $PAP^wt$  RNA, as well as to RNAs containing a canonical AAUAAA hexamer ( $PAP^wt$  and Ad-L3). In contrast, the addition of the same RNA oligo containing two mutations within the AAUAAA motif (AAACA) that eliminated the binding of CPSF had little impact on poly(A) addition (Fig. 5A, lanes 5,10,15,20). Comparable results were observed for poly(A) site cleavage (data not shown). As quantitated in Figure 5B, poly(A) addition to the precleaved Ad-L3 RNA, which contains a single UGUAN element (UGUAC, 13 nt upstream of the AAUAAA hexamer) and binds  $CFI_m$  relatively poorly (Brown and Gilmartin 2003), was found to be significantly more sensitive to the RNA competitor than either



**Figure 5.** CFI<sub>m</sub> contributes to the recruitment of CPSF. (A) The impact of the sequestration of CPSF on poly(A) addition at the Ad-L3, *PAPOLG*, and *PAPOLA* poly(A) sites. Uniformly <sup>32</sup>P-labeled precleaved RNA substrates were incubated in HeLa cell nuclear extract in the presence of ATP for 30 min at 30°C. An RNA oligo containing a 23-nt segment of the SV40 late poly(A) site (5'-CUGCAAUAAAACAAGUUAACAACA-3') was added at the indicated concentrations prior to incubation. The reactions in lanes 5, 10, 15, and 20 were incubated with 100 pmol of an RNA oligo of the same sequence, except for two hexamer mutations (AACACA). The RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. (B) The results of four independent sets of poly(A) addition experiments are shown, with the SD at each RNA competitor concentration shown as error bars. Within each experiment, the efficiency of poly(A) addition in the absence of the RNA competitor is arbitrarily set to 100%. (Filled circles) PAP $\gamma$ wt; (filled triangles) PAP $\alpha$ wt; (filled squares) PAP $\alpha\Delta 1/2$ ; (open circles) Ad-L3.

the PAP $\gamma$ wt or PAP $\alpha$ wt RNAs. These results suggested that CFI<sub>m</sub> enhanced the recruitment of CPSF to the *PAPOLG* and *PAPOLA* poly(A) sites. To test this hypothesis, we assayed the precleaved PAP $\alpha\Delta 1/2$  RNA that retained the AAUAAA hexamer but lacked the four UGUAN elements. As illustrated in Figure 5A, lanes 16–20, and quantitated in Figure 5B, poly(A) addition to the PAP $\alpha\Delta 1/2$  RNA was significantly more sensitive to the RNA competitor than the PAP $\alpha$ wt RNA. Taken together, these results indicate that CFI<sub>m</sub> binding at the noncanonical poly(A) site of *PAPOLG* does not obviate the requirement for CPSF in poly(A) addition, but rather

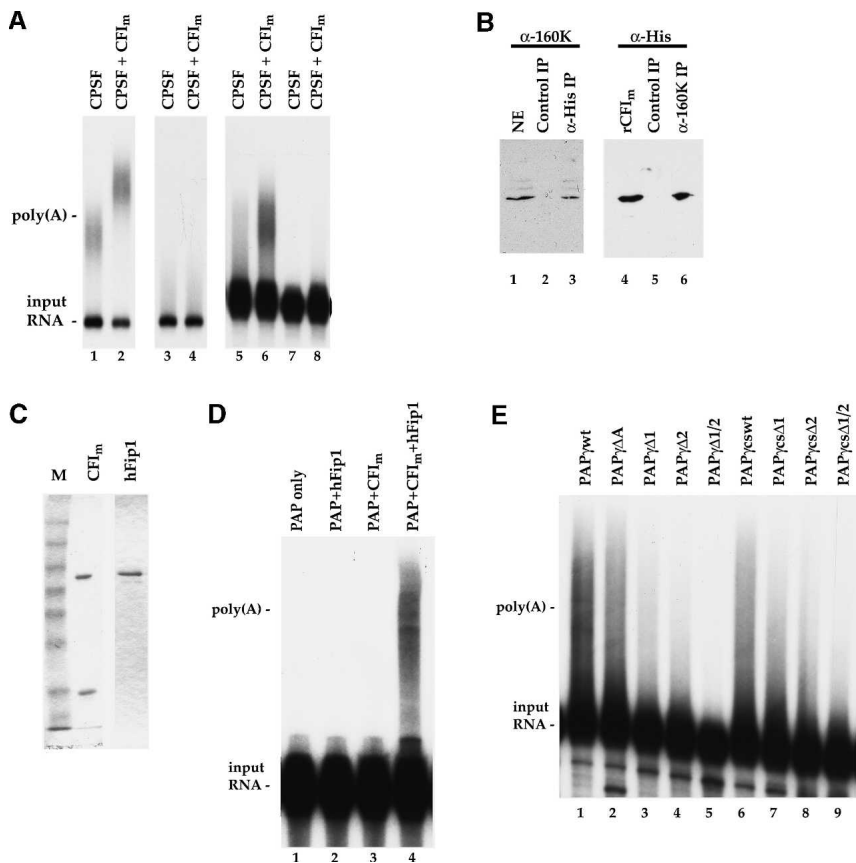
contributes to the recruitment of CPSF to the RNA. The recruitment of CPSF by CFI<sub>m</sub> is consistent with the ability of CFI<sub>m</sub> to enhance the binding of CPSF to a pre-mRNA (Rueggsegger et al. 1996) and the ability of CPSF to bind to a CFI<sub>m</sub>/RNA aptamer complex (Brown and Gil-martin 2003).

*CFI<sub>m</sub> can direct RNA sequence-specific, A(A/U)UAAA-independent poly(A) addition through its interaction with the CPSF subunit hFip1 and poly(A) polymerase*

To address the mechanism by which CFI<sub>m</sub> contributes to the recruitment of CPSF and enhances the efficiency of 3' processing, we first determined that CFI<sub>m</sub> was sufficient to enhance both complex formation and poly(A) addition in reactions reconstituted with purified HeLa cell CPSF. As illustrated in Figure 6A, CFI<sub>m</sub> enhanced the extent and efficiency of polyadenylation of both the PAP $\alpha$  and PAP $\gamma$  RNA substrates in reactions reconstituted with purified HeLa cell CPSF, recombinant CFI<sub>m</sub>, and recombinant poly(A) polymerase. Furthermore, the impact of CFI<sub>m</sub> was dependent upon the UGUAN elements of both poly(A) sites. It should be noted that the HeLa cell CPSF contained a very low, but detectable, level of CFI<sub>m</sub> (see Discussion). To address the potential interaction of CFI<sub>m</sub> and CPSF in the absence of RNA, coimmunoprecipitation experiments were carried out. Recombinant CFI<sub>m</sub> was added to HeLa cell nuclear extract in the presence of RNase and immunoprecipitated with an anti-hexahistidine antibody. As illustrated by Western analysis of the immunoprecipitated material (Fig. 6B, lanes 1–3), CPSF was coimmunoprecipitated with CFI<sub>m</sub>. The converse experiment demonstrated that recombinant CFI<sub>m</sub> could be coimmunoprecipitated with an anti-CPSF antibody following the addition of CFI<sub>m</sub> to HeLa cell nuclear extract in the presence of RNase (Fig. 6B, lanes 4–6).

We next asked if we could detect a direct interaction between CFI<sub>m</sub> and a specific CPSF subunit. CPSF is composed of five subunits: CPSF160, CPSF100, CPSF73, CPSF30, and hFip1 (Kaufman, et al. 2004). We had previously cloned the human CPSF subunit hFip1 (K. Venkataraman, unpubl.), which is identical in sequence to that recently reported by Kaufmann et al. (2004). We therefore investigated the interaction of CFI<sub>m</sub> with hFip1, both of which were expressed and purified from Sf9 cells (Fig. 6C). In our initial experiments, we did not observe a stable CFI<sub>m</sub>/hFip1/PAP $\gamma$ wt RNA complex by gel mobility shift analysis. We therefore asked if a functional interaction between CFI<sub>m</sub> and hFip1 could be observed in a poly(A) addition assay. Kaufmann et al. (2004) have shown that hFip1 interacts directly with poly(A) polymerase and that the binding of hFip1 to U-rich sequences enabled it to recruit poly(A) polymerase and stimulate poly(A) addition. As illustrated in Figure 6D, no poly(A) addition was observed upon incubation of the precleaved PAP $\gamma$ wt RNA substrate with poly(A) polymerase and either hFip1 (lane 2) or CFI<sub>m</sub> (lane 3) alone. Upon incubation of both hFip1 and CFI<sub>m</sub> with poly(A)

**Figure 6.** CFI<sub>m</sub> directs sequence-specific, A(A/U)UAAA-independent poly(A) addition through its interaction with hFip1 and poly(A) polymerase. (A) Impact of the addition of recombinant CFI<sub>m</sub> on poly(A) addition to the PAP $\alpha$ wt (lanes 1,2), PAP $\alpha$  $\Delta$ 1/2 (lanes 3,4), PAP $\gamma$ wt (lanes 5,6), and PAP $\gamma$  $\Delta$ 1/2 (lanes 7,8) RNA substrates in reactions reconstituted with purified HeLa cell CPSF and recombinant poly(A) polymerase. Twenty femtomoles of uniformly <sup>32</sup>P-labeled precleaved RNA substrates was incubated with purified HeLa cell CPSF and recombinant poly(A) polymerase in the presence or absence of 1 pmol of CFI<sub>m</sub> and ATP for 30 min at 30°C, and the RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. (B) Coimmunoprecipitation of CFI<sub>m</sub> and CPSF. Recombinant hexahis-tagged CFI<sub>m</sub> was mixed with HeLa cell nuclear extract and RNase A and subjected to immunoprecipitation with a mouse anti-hexahis antibody (lane 3), an affinity-purified CPSF160K rabbit anti-peptide antibody (lane 6), or control antibodies [mouse IgG1 [lane 2], and rabbit IgG [lane 5]]. Aliquots of the immunoprecipitates were analyzed by Western analysis using the anti-CPSF160K antibody (lanes 1–3) or the anti-hexahis antibody (lanes 4–6). (Lane 1) Five percent of the input HeLa cell nuclear extract. (Lane 6) Five percent of the input recombinant CFI<sub>m</sub>. (C) Purified baculovirus-expressed recombinant CFI<sub>m</sub> and hFip1. Proteins were resolved on an SDS–polyacrylamide gel and stained with Coomassie Blue. (M) Protein standards. (D) Poly(A) addition. Twenty femtomoles of uniformly <sup>32</sup>P-labeled precleaved PAP $\gamma$ wt RNA substrate was incubated with the indicated proteins in the presence of ATP for 60 min at 30°C. Each reaction contained 10 fmol of *E. coli*-expressed poly(A) polymerase. Lanes 2 and 4 contained 0.5 pmol of hFip1, and lanes 3 and 4 contained 1 pmol of CFI<sub>m</sub>. The RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel. (E) Twenty femtomoles of the indicated uniformly <sup>32</sup>P-labeled precleaved RNA substrates was incubated as in B with 10 fmol of poly(A) polymerase, 0.5 pmol of hFip1, and 1 pmol of CFI<sub>m</sub>. The RNA products were isolated and resolved on a denaturing 10% polyacrylamide gel.



polymerase, however, poly(A) addition was readily detected (Fig. 6D, lane 4). To determine whether sequence-specific RNA binding of CFI<sub>m</sub> was required for poly(A) addition in the presence of hFip1 and poly(A) polymerase, we assayed each of the precleaved *PAPOLG* mutant RNA substrates. The efficiency of poly(A) addition directly correlated with the ability of CFI<sub>m</sub> to bind the RNA substrates (Fig. 6E). CFI<sub>m</sub> is therefore able to direct sequence-specific, A(A/U)UAAA-independent poly(A) addition through the recruitment of hFip1 and poly(A) polymerase to the RNA substrate. These results suggest that the sequence-specific binding of CFI<sub>m</sub> to sequences upstream of the poly(A) site facilitates the recruitment of CPSF, at least in part, through CFI<sub>m</sub> interactions with the CPSF subunit hFip1.

*ChIP analysis indicates that CFI<sub>m</sub> is recruited along with CPSF and CstF at an early stage in transcription*

The ability of CFI<sub>m</sub> to participate in the recruitment of CPSF in an RNA sequence-specific manner strongly sug-

gests that CFI<sub>m</sub> participates in poly(A) site recognition along with CPSF and CstF. Both CPSF and CstF have been shown to colocalize with elongating RNAPII along the entire transcription unit on lampbrush chromosomes (Gall et al. 1999). In addition, Dantonel et al. (1997) demonstrated that CPSF is recruited to the promoter by TFIID and transferred to RNAPII during transcription initiation. The association of CPSF and CstF with the transcription elongation complex, through their interaction with the RNAPII CTD (McCracken et al. 1997), likely facilitates the cotranscriptional recognition of the poly(A) site as it emerges from RNAPII. We hypothesized that if CFI<sub>m</sub> participates in poly(A) site recognition, it may colocalize with CPSF and CstF along the length of the transcription unit. To test this hypothesis we used ChIP analysis. ChIP analysis has previously been used to localize several components of the 3' processing complex within the transcription unit (Licatalosi et al. 2002; Ahn et al. 2004; Kim et al. 2004). The 16.3-kb human housekeeping gene *G6PD* was chosen for ChIP analysis because of its high level of mRNA expression,



the use of a single poly(A) site, and an extensive non-transcribed region downstream of the gene (as determined by EST analysis). Four regions of the human G6PD gene were subjected to ChIP analysis (described in the legend for Fig. 7). Cross-linked chromatin was prepared from HeLa cells and processed as described in Materials and Methods.

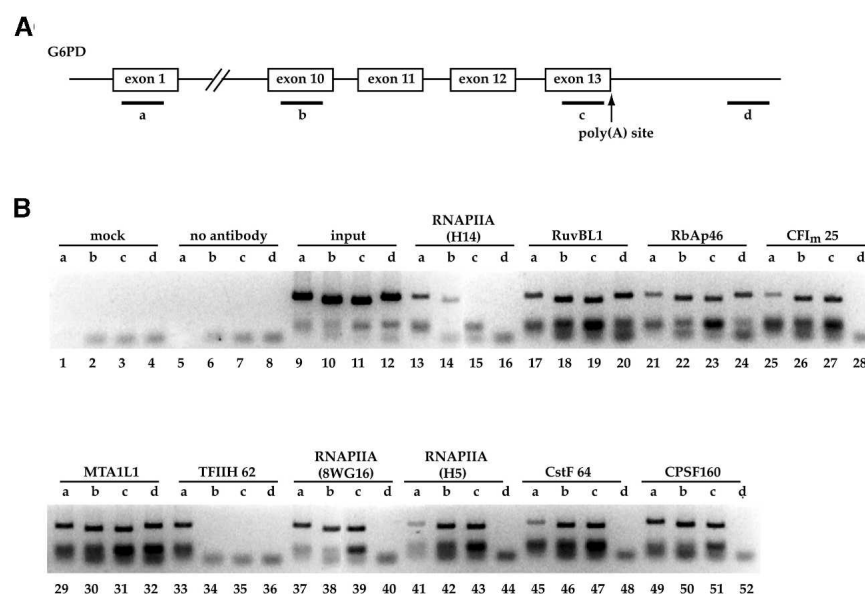
As illustrated in Figure 7, RNAPII antibodies directed against phospho-Ser 5 (H14, lanes 13–16), phospho-Ser 2 (H5, lanes 41–44), or a nonphosphorylated epitope (8WG16, lanes 37–40) of the CTD detected RNAPII associated with the transcribed region of the human G6PD gene, but not with sequences ~800 bp downstream of the poly(A) site. As expected, cross-linking of Ser 5-phosphorylated RNAPII was enriched at the 5' end of the gene, Ser 2-phosphorylated RNAPII appeared to be enriched downstream of the first exon (Komarnitsky et al. 2000), and cross-linking of the TFIIH 62-kDa subunit was observed only at the 5' end of the gene (Fig. 7, lanes 33–36) (Cheng and Sharp 2003). All three 3' processing proteins assayed, CFI<sub>m</sub> 25 (Fig. 7, lanes 25–28), CstF 64 (Fig. 7, lanes 45–48), and CPSF 160 (Fig. 7, lanes 49–52) cross-linked to the entire transcription unit. CFI<sub>m</sub> 25 and CstF 64 cross-linked in a manner comparable to that of the Ser 2-phosphorylated form of RNAPII, consistent with the requirement for Ser 2 phosphorylation for the cotranscriptional recruitment of 3' processing factors observed in yeast (Ahn et al. 2004). CPSF160 appeared to cross-link more uniformly throughout the transcription unit, consistent with the recruitment of CPSF at the promoter (Dantonel et al. 1997). Whereas neither RNAPII nor any of the 3' processing proteins cross-linked to sequences downstream of the G6PD gene, the cross-linking of several chromatin remodeling proteins (RuvBL1, RbAp46, and MTA1L1) was observed in this region, as

well as throughout the transcription unit. These results indicate that CFI<sub>m</sub> is indeed present along with CPSF and CstF throughout the transcription unit and is therefore ideally localized to participate in cotranscriptional poly(A) site recognition.

## Discussion

Yeast, plants, and vertebrates share a common set of at least a dozen well-conserved proteins that function in pre-mRNA 3' processing. Within each phylogenetic group, the assembly of a 3' processing complex is directed by a surprisingly diverse array of RNA sequences, suggesting that poly(A) site recognition is accomplished through a network of weak RNA:protein and protein:protein interactions. The work presented in this report indicates that vertebrate poly(A) site recognition is not restricted to the well-characterized sequence-specific interactions of CPSF and CstF. Rather, sequence-specific RNA binding of CFI<sub>m</sub> upstream of the poly(A) site plays a direct role in the recognition of both canonical and noncanonical poly(A) sites. Sequence-specific binding of CFI<sub>m</sub> to sequences upstream of the PAPOLA and PAPOLG poly(A) sites enhanced the efficiency of 3' processing complex assembly, as well as both poly(A) site cleavage and poly(A) addition. Furthermore, CFI<sub>m</sub> was shown to direct sequence-specific, A(A/U)UAAA-independent poly(A) addition through the recruitment of the CPSF subunit hFip1 and poly(A) polymerase to the RNA substrate. These results, along with the observation that CFI<sub>m</sub> is recruited near the 5' end of the transcription unit, indicate that CFI<sub>m</sub> functions along with CPSF and CstF in cotranscriptional poly(A) site recognition.

In their initial characterization of human CFI<sub>m</sub>, Ruegger et al. (1996) noted that CFI<sub>m</sub> exhibited an RNA



**Figure 7.** ChIP analysis of the G6PD gene in HeLa cells. Four regions of the human G6PD gene were subjected to ChIP analysis: (a) a 273-bp sequence located within exon 1, 159 bp downstream of the transcription start site and 15,433 bp upstream of the poly(A) site; (b) a 232-bp sequence located within exon 10, 14,455 bp downstream of the transcription start site and 1177 bp upstream of the poly(A) site; (c) a 227-bp segment located within exon 13, 15,628 bp downstream of the transcription start site and 10 bp upstream of the poly(A) site; and (d) a 265-bp segment located 796 bp downstream of the poly(A) site. (A) A schematic representation of the relative positions of segments of the G6PD gene analyzed by ChIP. (Note that the diagram is not drawn to scale.) (B) ChIP analysis of the human G6PD gene. (mock) Negative control in which buffer was substituted for chromatin. The H14 monoclonal antibody recognizes the RNAPII CTD phosphorylated at Ser 5, the H5 monoclonal antibody recognizes the RNAPII CTD phosphorylated at Ser 2, and the 8WG16 monoclonal antibody recognizes an unphosphorylated RNAPII CTD epitope. The antibodies are described in detail in Materials and Methods. Following agarose gel electrophoresis, PCR products were visualized by ethidium bromide staining.

lated at Ser 5, the H5 monoclonal antibody recognizes the RNAPII CTD phosphorylated at Ser 2, and the 8WG16 monoclonal antibody recognizes an unphosphorylated RNAPII CTD epitope. The antibodies are described in detail in Materials and Methods. Following agarose gel electrophoresis, PCR products were visualized by ethidium bromide staining.

binding activity with a preference for a poly(A) site-containing RNA, and that CFI<sub>m</sub> enhanced the binding of CPSF to the pre-mRNA. CFI<sub>m</sub> was subsequently found to function at an early step in pre-mRNA 3' processing complex assembly and to enhance the rate and overall efficiency of poly(A) site cleavage in vitro (Ruegsegger et al. 1998). Together, these observations suggested that CFI<sub>m</sub> might serve a role in poly(A) site recognition. SELEX analysis indicated a CFI<sub>m</sub> binding preference for the sequence UGUAN (N = A > U > C/G), and CFI<sub>m</sub> was found to auto-regulate the 3' processing of the CFI<sub>m</sub> 68-kDa subunit pre-mRNA through its interaction with a set of UGUAA elements that flank and overlap the AAUAAA hexamer (Brown and Gilmartin 2003). The data presented in this report indicate that the RNA binding activity of CFI<sub>m</sub> not only serves a stimulatory or auto-regulatory role, but rather, directly contributes to the processing of both canonical and noncanonical poly(A) sites. The functional interaction of CFI<sub>m</sub> and hFip1 that we observed reveals a mechanism by which CFI<sub>m</sub> and CPSF cooperate to define a poly(A) site. The ability of CFI<sub>m</sub> and hFip1 to function together in the recruitment of poly(A) polymerase to the RNA substrate is consistent with previously identified interactions of poly(A) polymerase with the CFI<sub>m</sub> 25-kDa subunit (Kim and Lee 2001; Dettwiler et al. 2004) and hFip1 (Kaufmann et al. 2004). Whereas hFip1 interacts with the N terminus and RNA-binding domain of poly(A) polymerase, the CFI<sub>m</sub> 25-kDa subunit interacts with the CTD. The CTD has previously been shown to be the target of proteins that regulate poly(A) polymerase activity, including U1A (Gunderson et al. 1997), U1 70K (Gunderson et al. 1998), U2AF 65 (Vagner et al. 2000), Srp75 (Ko and Gunderson 2002), and 14-3-3 ε (Kim et al. 2003).

Our analysis of the noncanonical poly(A) site of *PAPOLG* revealed a set of 3' processing elements strikingly similar to those of yeast and plants. Yeast and plant poly(A) sites possess a tripartite structure characterized by (1) an A-rich positioning element centered approximately on position -20, (2) an upstream efficiency element that is usually UA rich in the case of yeast (the best sequence being UAUUA), and most commonly of the form UUGUAAU or UUGUAA in plants, and (3) downstream U-rich sequences (Graber et al. 1999). The human *PAPOLG* poly(A) site displays a comparable tripartite set of sequence elements: an A-rich element (AAAGAGAAA) centered at position -20, multiple upstream UGUAN elements, and a U-rich downstream element. While plants possess clear CFI<sub>m</sub> homologs, neither CFI<sub>m</sub> subunit appears to have a yeast homolog. Zhao et al. (1999) initially noted that, based on the presence of an N-terminal RBD and a role in cleavage, the closest counterpart of CFI<sub>m</sub> in *Saccharomyces cerevisiae* was the 3' processing protein Hrp1p. This work supports the hypothesis that CFI<sub>m</sub> is the functional counterpart of Hrp1p. Both Hrp1p (Gross and Moore 2001) and CFI<sub>m</sub> (this work) contribute to poly(A) site recognition and 3' processing complex assembly by binding sequences upstream of the A-rich element, and each function in poly(A) site cleavage as well as poly(A) addition. Consis-

tent with a role in poly(A) site recognition, ChIP analysis indicated that both Hrp1p (Komarnitsky et al. 2000) and CFI<sub>m</sub> (this work) are associated with the entire transcription unit. In addition, both Hrp1p (Shen et al. 1998) and CFI<sub>m</sub> (Boisvert et al. 2003) are subject to arginine methylation. Arginine methylation is essential for the shuttling of Hrp1p between the nucleus and cytoplasm and appears to function in the coupling of transcription, mRNA processing, and export (Yu et al. 2004). As yet, is not known whether CFI<sub>m</sub> shuttles. Taken together, the functional and structural similarities of human CFI<sub>m</sub> and *S. cerevisiae* Hrp1p strongly suggest that they share comparable roles in pre-mRNA 3' processing.

The enhancement of PAPOLA and PAPOLG 3' processing by the specific binding of CFI<sub>m</sub> to sequences upstream of the poly(A) site raises the possibility that CFI<sub>m</sub> may also function in the recognition of previously identified upstream sequence elements (USEs). USEs, which act to enhance the efficiency of 3' processing, have been identified upstream of a number of canonical viral and cellular poly(A) sites and have generally been characterized as "U-rich sequences" (Zhao et al. 1999). Strikingly, most USEs identified to date encompass one or more UGUAN sequences. These include the USEs of the SV40 late (Schek et al. 1992), GSHV (Rusnak 1991), HIV-1 (Valsamakis et al. 1991), Lamin B2 (Brackenridge and Proudfoot 2000), 2/5'-oligoadenylate synthase (Aissouni et al. 2002), and the collagen 1A2 (Natalizio et al. 2002) poly(A) sites. A preliminary analysis of the collagen 1A2 poly(A) site indicated that mutations within the USE that reduced the efficiency of 3' processing also reduced the binding of CFI<sub>m</sub> (K. Venkataraman, unpubl.). A bioinformatic analysis of human 3' UTRs identified a TGT element, usually positioned slightly upstream of the AATAAA hexamer, as the most abundant family of motifs in human 3' UTRs (Louie et al. 2003). Based on this observation, and the similarity of the TGT motif to the upstream elements of yeast and plant poly(A) sites, Louie et al. (2003) proposed that this motif may constitute a general class of human USEs. The evidence presented in this report supports this proposal, and furthermore suggests that CFI<sub>m</sub> is the 3' processing factor responsible for the recognition of these elements.

A previous analysis of the upstream element of the HIV-1 poly(A) indicated that this sequence was contacted by the 160-kDa subunit of CPSF (Gilmartin et al. 1995). Strong UV-cross-linking products of 160, 68, and 25 kDa were observed upon incubation of an HIV-1 RNA, specifically <sup>32</sup>P-labeled within the USE, with partially purified CPSF. When a highly purified CPSF fraction was used, low levels of both the 160- and 68-kDa UV-cross-linking products were observed. These results are consistent with the copurification of CFI<sub>m</sub> with CPSF and the interaction of both proteins with the HIV-1 USE. It is intriguing that Kaufmann et al. (2004) found that CPSF reconstituted with recombinant subunits (including hFip1) was incapable of directing AAUAAA-dependent poly(A) addition in vitro. They suggested that additional proteins might be required for specific poly(A) addition. We therefore raise the possibility that CFI<sub>m</sub>

might represent the missing component. In a manner comparable to that of the long-overlooked CPSF subunit hFip1 (Kaufmann et al. 2004), the presence of CFI<sub>m</sub> in purified active mammalian CPSF preparations may have gone unnoticed due to its presence at sub-stoichiometric levels.

Experiments based on the analysis of CPSF purified from human or bovine tissue have indicated that CPSF is both necessary and sufficient to direct AAUAAA-dependent poly(A) addition in vitro. Through the use of purified recombinant proteins we have demonstrated that sequence-specific poly(A) addition to an authentic human poly(A) site lacking an A(A/U)UAAA motif can be directed by CFI<sub>m</sub>. Through its ability to bind the PAPOLG RNA substrate and recruit both hFip1 and poly(A) polymerase CFI<sub>m</sub> was able to confer sequence-specificity to the poly(A) addition reaction.

Although the *PAPOLA* and *PAPOLG* genes were initially chosen for analysis because of the distinct composition of their poly(A) sites, this work highlights a potential mechanism for the differential regulation of these two paralogs. *PAPOLA* and *PAPOLG* appear to be expressed in most cell types, although EST expression profiles (UniGene) indicate that *PAPOLG* mRNA is generally expressed at a level one-tenth that of *PAPOLA* mRNA. Topalian et al. (2001) and Kyriakopoulou et al. (2001) found PAP $\alpha$  and PAP $\gamma$  to be biochemically indistinguishable in vitro. Evolutionary theory, supported by recent bioinformatics analysis, predicts that the fates of duplicated gene pairs is determined in the initial phases of duplicate gene evolution and that positive selection plays a prominent role in the evolutionary dynamics of the very early histories of duplicate genes (Moore and Purugganan 2003). This suggests that the distinct sequences of the *PAPOLA* and *PAPOLG* poly(A) sites were acquired soon after duplication and that these sequence differences have been maintained by selective pressure, most likely due to their contribution to the regulation of the expression of each paralog. In vivo data indicate that PAP $\alpha$  activity is tightly regulated (Zhao and Manley 1998). The activity of PAP $\alpha$  has been found to be regulated through cdc2-cyclin B phosphorylation (Colgan et al. 1996, 1998), as well as through the interaction of several splicing factors with its CTD—the domain that exhibits the greatest divergence between PAP $\alpha$  and PAP $\gamma$ . In addition, both *PAPOLA* and *PAPOLG* have been reported to be significantly overexpressed in human tumors (Topalian et al. 2001), and poly(A) polymerase activity has been proposed as a prognostic marker in primary breast cancer (Scorilas et al. 2000). The importance of the regulation of poly(A) polymerase activity is further suggested by the identification of overlapping sets of potential miRNA targets within the 3' UTRs of both human *PAPOLA* (seven predicted targets) and *PAPOLG* (eight predicted targets) (John et al. 2004).

The work presented in this report supports a model in which three distinct sequence elements contribute to vertebrate poly(A) site definition: an upstream UGUAN element recognized by CFI<sub>m</sub>; an A-rich element, most often AAUAAA, recognized by CPSF; and a U-rich

downstream element recognized by CstF. Each of these RNA:protein interactions are likely to contribute to the establishment of the network of interactions responsible for poly(A) site choice in vivo. The observation by Takagaki et al. (1996) that poly(A) site choice within the IgM heavy chain gene was dependent upon the concentration of the CstF 64-kDa subunit demonstrated the ability of a basal mRNA 3' processing factor to influence alternative poly(A) site choice. The modulation of the RNA binding affinity and the cooperativity of the interactions of CPSF, CstF, and CFI<sub>m</sub>, along with their effective concentrations, is likely to make a decisive contribution to the regulation of alternative poly(A) site selection.

## Materials and methods

### *Protein expression and purification*

Human hFip1 (FIP1L1) was cloned from human cDNA, and sequencing confirmed it to be identical to that reported by Kaufmann et al. (2004). hFip1 was subcloned into a Bac-to-Bac pFast-Bac Dual baculovirus expression vector (Invitrogen) with an N-terminal HA-tag, expressed in Sf9 cells, and purified by immunoaffinity chromatography. The CFI<sub>m</sub> 68/25-kDa heterodimer was also expressed in Sf9 cells and purified as described previously (Brown and Gilmartin 2003). The bovine poly(A) polymerase cDNA was a kind gift of E. Wahle (Institut für Biochemie, Martin Luther Univesitat Halle Wittenberg, Halle, Germany). Bovine poly(A) polymerase  $\alpha$  was expressed in *Escherichia coli* with an N-terminal hexahistidine tag and purified by Ni<sup>2+</sup>-NTA affinity chromatography.

### *In vitro 3' processing assays*

Capped, uniformly <sup>32</sup>P-labeled *PAPOLA* and *PAPOLG* RNA substrates were prepared by SP6 RNA polymerase transcription of PCR-amplified human DNA. The sequence of the templates was confirmed by DNA sequencing. The oligonucleotide primer sequences for PCR amplification are available upon request. The Ad-L3 poly(A) site-containing RNA substrate is described by Gilmartin et al. (1995). In vitro poly(A) cleavage, poly(A) addition, and gel mobility shift assays were carried out as described by Brown and Gilmartin (2003). HeLa cell CPSF was purified as described by Gilmartin et al. (1995). Quantitation of the RNA products was done on a Bio-Rad Personal Molecular Imager FX.

### *Coimmunoprecipitation analysis*

Ten micrograms of recombinant hexahistidine-tagged CFI<sub>m</sub> was added to 100  $\mu$ L of HeLa cell nuclear extract, followed by the addition of 50  $\mu$ L of protein A Sepharose (Pharmacia), 1  $\mu$ g of antibody, 880  $\mu$ L of IP buffer (25 mM KCl, 50 mM Tris-HCl at pH 7.8, 0.2 mM EDTA, 10% glycerol, 1% BSA), protease inhibitor (Calbiochem, Set III at 1:1000), and 10  $\mu$ g of RNase A. The mixture was incubated for 10 min at 20°C, and then placed on a rotating wheel for 12 h at 4°C. The binding reactions were centrifuged and the protein A Sepharose beads were washed five times with PBS. The pellet was resuspended in 100  $\mu$ L of Laemmli sample buffer. The following antibodies were used for immunoprecipitation: (1) His-probe (H3) (Santa Cruz Biotechnology), (2) an affinity-purified rabbit anti-peptide antibody directed against the peptide DKEEPPSKKKRVDAT of human CPSF160K, and (3) control antibodies: rabbit IgG or mouse IgG1 (Southern Biotechnology). Aliquots of each immunoprecipitate

were resolved on a 10% polyacrylamide-SDS gel and subjected to Western analysis with either the His-probe antibody or the CPSF160K anti-peptide antibody. As a positive control for each Western, a 5% aliquot of the starting material was also run. HRP-conjugated secondary antibodies were detected by ECL [Amersham].

#### ChIP analysis

ChIPs were performed according to the Farnham lab protocol (<http://genomecenter.ucdavis.edu/farnham/farnham/protocols/chips.html>). Chromatin was prepared from four confluent T75 flasks of HeLa cells. The antibodies used were as follows: RNAPII CTD monoclonal antibodies H5, H14, and 8WG16 were obtained from BABCO; RuvBL1 [NMP 238 (N-15)], RbAp46 [N-19], Mta1L1 (C-20), TFIIF 62 (Q-19), and CPSF 160 [CPSF1 (E-20)] were obtained from Santa Cruz Biotechnology. The CFI<sub>m</sub>25 antibody was an affinity-purified rabbit anti-peptide antibody raised against the sequence YTFGTKEPLYEK DSS. Anti-CstF 64 (3A7) was a kind gift of C. MacDonald (Department of Cell Biology and Biochemistry, Texas Tech University, Health Sciences Center, Lubbock, TX). Oligonucleotide primer sequences for PCR amplification are available upon request.

#### Acknowledgments

We thank the Vermont Cancer Center DNA Analysis Facility for support in DNA sequencing and molecular imaging.

#### References

- Ahn, S.H., Kim, M., and Buratowski, S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell* **13**: 67–76.
- Aissouni, Y., Perez, C., Calmels, B., and Benech, P.D. 2002. The cleavage/polyadenylation activity triggered by a U-rich motif sequence is differently required depending on the poly(A) site location at either the first or last 3'-terminal exon of the 2'-5' oligo(A) synthetase gene. *J. Biol. Chem.* **277**: 35808–35814.
- Awasthi, S. and Alwine, J.C. 2003. Association of polyadenylation cleavage factor I with U1 snRNP. *RNA* **9**: 1400–1409.
- Barabino, S.M.L. and Keller, W. 1999. Last but not least: Regulated poly(A) tail formation. *Cell* **99**: 9–11.
- 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.
- Bienroth, S., Keller, W., and Wahle, E. 1993. Assembly of a processive messenger RNA polyadenylation complex. *EMBO J.* **12**: 585–594.
- Boisvert, F.M., Cote, J., Boulanger, M.C., and Richard, S. 2003. A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* **2**: 1319–1330.
- 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.
- Brown, K.M. and Gilmartin, G.M. 2003. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor I<sub>m</sub>. *Mol. Cell* **12**: 1467–1476.
- Calvo, O. and Manley, J.L. 2003. Strange bedfellows: Polyadenylation factors at the promoter. *Genes & Dev.* **17**: 1321–1327.
- Cheng, C.H. and Sharp, P.A. 2003. RNA polymerase II accumulation in the promoter-proximal region of the dihydrofolate reductase and  $\gamma$ -actin genes. *Mol. Cell. Biol.* **23**: 1961–1967.
- Colgan, D.F., Murthy, K.G., Prives, C., and Manley, J.L. 1996. Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature* **384**: 282–285.
- Colgan, D.F., Murthy, K.G., Zhao, W., Prives, C., and Manley, J.L. 1998. Inhibition of poly(A) polymerase requires p34cdc2/cyclin B phosphorylation of multiple consensus and non-consensus sites. *EMBO J.* **17**: 1053–1062.
- Dantone, 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.
- Dettwiler, S., Aringhieri, C., Cardinale, S., and Keller, W. 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.
- Edwards-Gilbert, G., Veraldi, K.L., and Milcarek, C. 1997. Alternative poly(A) site selection in complex transcription units: Means to an end? *Nucleic Acids Res.* **25**: 2547–2561.
- Gall, J.G., Bellini, M., Wu, Z., and Murphy, C. 1999. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol. Biol. Cell* **10**: 4385–4402.
- Gilmartin, G.M. and Nevins, J.R. 1991. Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. *Mol. Cell. Biol.* **11**: 2432–2438.
- 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.
- 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.* **96**: 14055–14060.
- Graveley, B.R. 2000. Sorting out the complexity of SR protein functions. *RNA* **6**: 1197–1211.
- 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.
- Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I.W. 1997. Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation. *Genes & Dev.* **11**: 761–773.
- Gunderson, S.I., Polycarpou-Schwarz, M., and Mattaj, I.W. 1998. U1 snRNP inhibits polyadenylation through a direct interaction between U1 70K and polyA polymerase. *Mol. Cell* **1**: 255–264.
- Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C., and Cole, C.N. 2002. Coupling of termination, 3' processing, and mRNA export. *Mol. Cell. Biol.* **22**: 6441–6457.
- Iseli, C., Stevenson, B.J., deSouza, S.J., Samaia, H.B., Camargo, A.A., Buetow, K.H., Strausberg, R.L., Simpson, A.J.G., Bucher, P., and Jongeneel, C.V. 2002. Long-range heterogeneity at the 3' ends of human mRNAs. *Genome Res.* **12**: 1068–1074.
- John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. 2004. Human microRNA targets. *PLoS Biol.* **2**: e363.
- 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.
- Kim, H. and Lee, Y. 2001. Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I. *Biochem. Biophys. Res. Commun.* **289**: 513–518.
- Kim, H., Lee, J.H., and Lee, Y. 2003. Regulation of poly(A) polymerase by 14–3–3 $\epsilon$ . *EMBO J.* **22**: 5208–5219.
- Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. 2004. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J.* **23**: 354–364.
- Ko, B. and Gunderson, S.I. 2002. Identification of new poly(A) polymerase-inhibitory proteins capable of regulating pre-mRNA polyadenylation. *J. Mol. Biol.* **318**: 1189–1206.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Dev.* **14**: 2452–2460.
- Kuhn, U. and Wahle, E. 2004. Structure and function of poly(A) binding proteins. *Biochim. Biophys. Acta* **1678**: 67–84.
- Kyriakopoulou, C.B., Nordvang, H., and Virtanen, A. 2001. A novel human poly(A) polymerase (PAP), PAP $\gamma$ . *J. Biol. Chem.* **276**: 33504–33511.
- Lei, E.P. and Silver, P.A. 2002. Intron status and 3'-end formation control cotranscriptional export of mRNA. *Genes & Dev.* **16**: 2761–2766.
- Licalosi, 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.
- Louie, E., Ott, J., and Majewski, J. 2003. Nucleotide frequency variation across human genes. *Genome Res.* **13**: 2594–2601.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. 1997. The carboxy-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**: 357–361.
- Moore, R.C. and Purugganan, M.D. 2003. The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci.* **100**: 15682–15687.
- Natalizio, B.J., Muniz, L.C., Arhin, G.K., Wilusz, J., and Lutz, C.S. 2002. Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals. *J. Biol. Chem.* **277**: 42733–42740.
- Proudfoot, N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr. Opin. Cell Biol.* **16**: 272–278.
- Proudfoot, N.J. and O'Sullivan, J. 2002. Polyadenylation: A tail of two complexes. *Curr. Biol.* **12**: R855–R857.
- Proudfoot, N.J., Furger, A., and Dye, M. J. 2002. Integrating mRNA processing with transcription. *Cell* **108**: 501–512.
- Rappilber, J., Ryder, U., Lamond, A.I., and Mann, M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**: 1231–1245.
- Ruegsegger, U., Beyer, K., and Keller, W. 1996. Purification and characterization of human cleavage factor I-m, involved in the 3' end processing of messenger RNA precursors. *J. Biol. Chem.* **271**: 6107–6113.
- 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.
- Russnak, R.H. 1991. Regulation of polyadenylation in hepatitis B viruses: Stimulation by the upstream activating signal PS1 is orientation-dependent, distance dependent, and additive. *Nucleic Acids Res.* **19**: 6449–6456.
- Schek, N., Cooke, C., and Alwine, J.C. 1992. Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses. *Mol. Cell. Biol.* **12**: 5386–5393.
- Scorilas, A., Talieri, M., Alexandros, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsiapalis, C.M., and Trangas, T. 2000. Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: A new independent prognostic marker in primary breast cancer. *Cancer Res.* **60**: 5427–5433.
- 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 of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* **18**: 5799–5805.
- Shen, E.C., Henry, M.F., Weiss, V.H., Valentini, S.R., Silver, P.A., and Lee, M.S. 1998. Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes & Dev.* **12**: 679–691.
- Takagaki, Y., Seipelt, R.L., Peterson, M.L., and Manley, J.L. 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**: 941–952.
- Topalian, S.L., Kaneko, S., Gonzales, M.I., Bond, G.L., Ward, Y., and Manley, J.L. 2001. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. *Mol. Cell. Biol.* **21**: 5614–5623.
- Vagner, S., Vagner, C., and Mattaj, I.W. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes & Dev.* **14**: 403–413.
- Valsamakis, A., Zeichner, S., Carswell, S., and Alwine, J.C. 1991. The human immunodeficiency virus type 1 polyadenylation signal: A 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. *Proc. Natl. Acad. Sci.* **88**: 2108–2112.
- Wahle, E. 1995. Poly(A) tail length control is caused by termination of processive synthesis. *J. Biol. Chem.* **270**: 2800–2808.
- Wigley, P.L., Sheets, M.D., Zarkower, D.A., Whitmer, M.E., and Wickens, M. 1990. Polyadenylation of mRNA: Minimal substrates and a requirement for the 3' hydroxyl of the U in AAUAAA. *Mol. Cell. Biol.* **10**: 1705–1713.
- Yu, M.C., Bachand, F., McBride, A.E., Komili, S., Casolari, J.M., and Silver, P.A. 2004. Arginine methyltransferase affects interactions and recruitment of mRNA processing and export factors. *Genes & Dev.* **18**: 2024–2035.
- Zhao, W.Q. and Manley, J.L. 1998. Deregulation of poly(A) polymerase interferes with cell growth. *Mol. Cell. Biol.* **18**: 5010–5020.
- Zhao, J., Hyman, L., and Moore, C. 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.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**: 182–185.
- Zorio, D.A.R. and Bentley, D.L. 2004. The link between mRNA processing and transcription: Communication works both ways. *Exp. Cell Res.* **296**: 91–97.