The 160-kD subunit of human cleavage–polyadenylation specificity factor coordinates pre-mRNA 3'-end formation

Kanneganti G.K. Murthy and James L. Manley

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

Cleavage–polyadenylation specificity factor (CPSF) is a multisubunit protein that plays a central role in 3' processing of mammalian pre-mRNAs. CPSF recognizes the AAUAAA signal in the pre-mRNA and interacts with other proteins to facilitate both RNA cleavage and poly(A) synthesis. Here we describe the isolation of cDNAs encoding the largest subunit of CPSF (160K) as well as characterization of the protein product. Antibodies raised against the recombinant protein inhibit polyadenylation in vitro, which can be restored by purified CPSF. Extending previous studies, which suggested that 160K contacts the pre-mRNA, we show that purified recombinant 160K can, by itself, bind preferentially to AAUAAA-containing RNAs. While the sequence of 160K reveals similarities to the RNP1 and RNP2 motifs found in many RNA-binding proteins, no clear match to a known RNA-binding domain was found, and RNA recognition is therefore likely mediated by a highly diverged or novel structure. We also show that 160K binds specifically to both the 77K (suppressor of forked) subunit of the cleavage factor CstF and to poly(A) polymerase (PAP). These results provide explanations for previously observed cooperative interactions between CPSF and CstF, which are responsible for poly(A) site specification, and between CPSF and PAP, which are necessary for synthesis of the poly(A) tail. Also supporting a direct role for 160K in these interactions is the fact that 160K by itself retains partial ability to cooperate with CstF in binding pre-mRNA and, unexpectedly, inhibits PAP activity in in vitro assays. We discuss the significance of these multiple functions and also a possible evolutionary link between yeast and mammalian polyadenylation suggested by the properties and sequence of 160K.

[*Key Words:* Pre-mRNA processing; poly(A) synthesis; cleavage–polyadenylation specificity factor]

Received August 9, 1995; revised version accepted September 5, 1995.

Polyadenylation is an important step in the maturation of almost all eukaryotic mRNAs. This reaction involves endonucleolytic cleavage of the pre-mRNA to generate the 3' end, and the subsequent addition of 200–300 adenylate residues (for review, see Manley 1988; Wahle and Keller 1992). The most important and well-defined *cis*-acting element required for polyadenylation is the highly conserved hexanucleotide AAUAAA, located 10–30 nucleotides upstream of the poly(A) addition site in most pre-mRNAs (Proudfoot 1991). Mutations in the hexanucleotide invariably impair correct 3'-end formation and frequently result in reduced mRNA accumulation. Additional sequences that affect polyadenylation efficiency, located both upstream and downstream of AAUAAA, have been identified in several systems. Recently, evidence has been presented in two cases demonstrating that such upstream sequences correspond to binding sites for the U1 small nuclear ribonucleoprotein (snRNP)-specific A protein, which when bound, can affect polyadenylation either negatively (Boelens et al. 1993; Gunderson et al. 1994) or positively (Lutz and Alwine 1994).

Biochemical studies have provided considerable insights into the general, or basal, polyadenylation ma-

chinery in mammalian cells (for review, see Manley 1995). Reconstitution of the reaction has identified six separable factors that are required for efficient polyadenylation in vitro. Two of these, cleavage–polyadenylation specificity factor (CPSF) and cleavage stimulatory factor (CstF), are multisubunit proteins that function together to bind the pre-mRNA, probably as a first step in specifying the poly(A) site. CPSF was suggested to recognize AAUAAA some time ago, because it is the only factor required for both the cleavage and poly(A) addition phases of the reaction, both of which are dependent on AAUAAA (Christofori and Keller 1988; Gilmartin and Nevins 1989; Takagaki et al. 1989). Additional experiments with highly purified CPSF confirmed that the protein can specifically recognize AAUAAA (Bardwell et al. 1991; Keller et al. 1991). Binding is very weak with pure CPSF but can be greatly enhanced by a cooperative interaction with CstF (Wilusz et al. 1990; Gilmartin and Nevins 1991; Murthy and Manley 1992). CstF is a heterotrimer (Takagaki et al. 1990) that also binds RNA, but again very weakly by itself. All three subunits of CstF have been cloned (Takagaki et al. 1991; Takagaki and Manley 1992, 1994), and RNA binding is mediated by a 64-kD polypeptide through the action of an RNP-type

RNA-binding domain. CstF has been shown to interact with the G/U-rich downstream sequences that are important elements of many poly(A) sites (Weiss et al. 1991; MacDonald et al. 1994; Y. Takagaki and J.L. Manley, in prep.). Even though CstF does not directly recognize AAUAAA, single-base changes in this element can reduce or eliminate binding (Moore et al. 1988; Wilusz and Shenk 1988; Takagaki et al. 1990; Gilmartin and Nevins 1991; Murthy and Manley 1992), which is indicative of the strong cooperative interaction between CPSF and CstF. Two additional proteins, cleavage factors I and II, are required for the actual pre-mRNA cleavage event (Takagaki et al. 1989), which in many cases also requires the poly(A) polymerase (PAP) (Christofori and Keller 1988, 1989; Takagaki et al. 1988; Gilmartin and Nevins 1989; Ryner et al. 1989a). Following cleavage, CPSF and PAP are necessary and sufficient to catalyze poly(A) addition (e.g., Raabe et al. 1991, 1994; Wahle et al. 1991), and a cooperative RNA-binding interaction involving these two factors has been demonstrated (Murthy and Manley 1992; Bienroth et al. 1993). Following initiation of poly(A) synthesis, another factor, poly(A) binding protein II, enhances the processivity of the elongation reaction (Wahle 1991), although CPSF continues to be required for maximum processivity (Sheets and Wickens 1989; Bienroth et al. 1993).

Given the requirement of CPSF in both the cleavage and poly(A) addition phases of the reaction, and its ability to recognize AAUAAA and interact with other factors, it is clear that CPSF plays a central role in pre-mRNA polyadenylation. CPSF consists of three subunits of ~160, 100, and 73 kD, and possibly a fourth, nonessential polypeptide of 30 kD (Beinroth et al. 1991; Murthy and Manley 1992). The isolation of cDNAs encoding the 100-kD subunit has been reported recently (Jenny et al. 1994). Several earlier studies, employing UV cross-linking assays, suggested that the 160-kD polypeptide may directly contact the RNA (Moore et al. 1988; Gilmartin and Nevins 1989, 1991; Keller et al. 1991). Evidence that the 160-kD subunit interacts with an important upstream polyadenylation signal in the HIV 3' LTR as well as with AAUAAA has also been reported (Gilmartin et al. 1995). In addition, CPSF has been implicated in the cytoplasmic polyadenylation that occurs during oocyte maturation (for review, see Wickens 1992), and this interaction also requires an upstream element, the U-rich cytoplasmic polyadenylation element (Bilger et al. 1994). However, how CPSF interacts both with the pre-mRNA and with other polyadenylation factors is unknown. To begin to understand how CPSF functions, we isolated cDNAs encoding the 160-kD subunit. We show that the protein is essential for polyadenylation in vitro and that it can interact specifically with both AAUAAA-containing pre-mRNA and with other polyadenylation factors.

Results

Sequence of the CPSF 160-kD subunit

Previously, we described purification of CPSF to appar-

ent homogeneity from calf thymus (Murthy and Manley 1992). To isolate cDNAs, the 160-kD species was eluted from preparative gels and subjected to limited proteolysis with V8 protease. The amino-terminal sequence was obtained from several peptides, and a 44-nucleotide degenerate oligonucleotide was designed and used to screen a bovine heart cDNA library (see Materials and methods). A 3.3-kb insert was partially sequenced and found to encode the peptide sequences. This insert was then used to screen a HeLa cell cDNA library, and 20 positive clones were obtained. Sequence analysis of the largest insert (4.5 kb) revealed a 4326-base open reading frame, which is capable of encoding the full-length 160-kD protein (160K). Northern blotting of HeLa poly(A)⁺ RNA detected a single species of ~5.0-kb (data not shown).

Figure 1 presents the amino acid sequence of human 160K. The protein consists of 1442 amino acids with a predicted molecular mass (160.8K) that agrees exceptionally well with the apparent size of the native protein, as determined by SDS-gel electrophoresis (160–165 kD). Searches of the data bases did not reveal strong homology with known proteins (although see below). However,

```

1  MYAVYKQAHPPPGLEFSMYCNFFNNSEARNLVVAGTSQLYVYRLNRDAEALTKNDRSTEGK
61  AHREKLELAASFSFFGNVMSMASVQLAGAKRDALLLSFKDAKLSVVEYDPGTHDLKTLSL
121  HYFEPELRDGFVQNVHTPRVRVDPDGRCAAMLVYGTSLVLPFRRESLAEHEGLVGE
181  QRSSFPSYIIDVRALDEKLLNIIDLQFLHGYEPTLLILFENQTPWGRVAVRQDTCIS
241  VAISLNTQKVHPVWISLTLSPFDCTQALAVPKP IGGVVFAVNSLLYNQSVPPYGVGL
301  NSLTGTTTAFPLRTQEGVRITLDCAQATFISYDKMVISLKGGEIYVLTITDGMRSVRVF
361  HFDKAAASVLTTSMTMEIGYLFGLRLGNSLLKYTEKLQEPASAVREAAKEEPPSK
421  KKRVDATAGWSAAGKSVQDEVDIEIVYGSEASGTQLATYSFEVCDLSILNIGPCANAIV
481  GEPAFLEEFQNSPEPDLEIVVCSGHGKNGALSVLQKSIQPVVTTFFELPGCYDMWTIVIA
541  PVRKEEDNPKGEGTEQEPSTTPEADDDGRHRGFLILSREDSTMILQTQGEIMELDTSGF
601  ATQGPTVFACNIGDNRYIVQVSLGIRLLEGVNQLHFIPVDLGAPIVQCAVADPYVIVMS
661  AEGHVTMFLKSDSYGGRHRLALHKPLHQQSKVITCLCLYRDLSGMFTTESRLGGARDE
721  LGGRSGEAEGLGSETSPVDEEEMLYGDSGLSPSPKEARRSSQPPADRPAPFRAE
781  PTHWCLLVRENGTMEIYQLPDWRLVFLVKNFPVQGRVLVDSFGQPTTGQEARREATRO
841  GELPLVKEVLLVALGSRQSRPYLLVHVDPQELLYEAFPHDSQLGQGNLKVREKKVPHNIN
901  FREKKPKPSKKKAEGGGAEGAGARGVRFYFIEDYIGYSGVFICGSPHWLLVTGRGA
961  LRLHPMAIDGPVDSFAPFHNVCPRGFLYFNQGLRISVLPAVLSYDAPWVPRKIPLRC
1021  TAHYVAYHVESKVYAVATSTNTPCARIPRMTGEEKEFETIERDERYIHPQQAESIQILIS
1081  PVSWEAIPNARIELQEWHTCMKTVSLRSEETVSLGKYVAAGTCLMGQEEVTCRGRIL
1141  IMDVIEVPEPGQPLTKNKFVLYEKEQKGPVTALCHCNHLSAIGQKIFLWSLRASEL
1201  TGMAFIDTQLYIHQMISVKNEILAADVMSISLLRYQESKTLVSLRDAKPLEVYSVDF
1261  MVDNAQLGLFVSDRDRNLVMVYMLPEAKESFGGMRLLRADFHVGAHVNTFWRTPCRATE
1321  GLSKKSVVWENKHITWFATLDGGIGLLPMQEKTYRRLMLQNALTTMLPHHAGLNPRAF
1381  RMLHVDRTLQNAVRNVLDGELLNRYLYLSTMERSELAKKIGTTPDIILDDLLETDRVTA
1441  HF

```

Figure 1. Amino acid sequence of the human CPSF 160-kD subunit. The deduced amino acid sequence of 160K is shown, using the single-letter code for amino acid residues. The similarities to the RNP1 and RNP2 motifs found in RNA-binding proteins are boxed, the possible bipartite nuclear localization signal is underlined, and the sequence of the peptide used in cloning is indicated by broken underlining. The GenBank accession number of the sequence is (U37012).

two possibly significant motifs were apparent from visual inspection of the sequence. One is a likely bipartite nuclear localization signal (NLS), between residues 891 and 912 (Fig. 1, underlined). From previous studies of other proteins, it is not clear whether the upstream basic cluster is required or whether the long downstream cluster is sufficient to direct nuclear targeting [for review, see Dingwall and Lasky 1991]. Consistent with both the presence of an NLS and the known function of 160K in polyadenylation, indirect immunofluorescence experiments indicate that the protein in HeLa cells is exclusively nuclear (data not shown). Second, 160K contains matches to the RNP 1 (residues 379–386) and RNP 2 (residues 344–349) consensus sequences (Fig. 1, boxes) that are hallmarks of the RNP-type RNA-binding domain [(RBD); for review, see Kenan et al. 1991; Burd et al. 1994]. These two motifs are reasonable matches to the established consensus sequences, and the separation between them (29 residues) agrees very well with that found in other RBD proteins. However, certain other conserved positions in the ~80-residue RBD consensus are not present in 160K, and the significance of this similarity thus remains to be determined.

160K is required for CPSF-dependent polyadenylation in vitro

To begin to characterize 160K, we first prepared antisera against the recombinant protein. Despite its large size, full-length recombinant 160K was expressed well in *Escherichia coli* (see below). However, for production of antisera an amino-terminal fragment of 972 residues was expressed, purified, and used to inject rabbits (see Materials and methods). The antisera obtained (but not the preimmune sera) recognized a single species of 160 kD in Western blots of HeLa cell extracts as well as purified preparations of calf thymus CPSF (data not shown). The preimmune and immune sera were then used in immunodepletion experiments (Takagaki et al. 1990; see Materials and methods). Antibodies were prebound to protein A agarose beads, which were then mixed with either HeLa nuclear extract or a partially purified fraction of HeLa CPSF (Mono Q fraction; Takagaki et al. 1990). Untreated and treated preparations were then used in AAUAAA-dependent polyadenylation reactions with an SV40 late pre-mRNA substrate. Treatment of nuclear extract with preimmune antibodies had little, if any, effect on polyadenylation (Fig. 2, cf. lanes 2 and 3), whereas the immune antibodies resulted in a significant decrease in polyadenylated product and an increase in unprocessed pre-mRNA (lane 4). Similar results were obtained when the partially purified CPSF and PAP were utilized in an AAUAAA-dependent poly(A) addition reaction. In this case, the CPSF fraction was depleted with preimmune or immune antibodies prior to the reaction. As described above, the CPSF fraction depleted with the preimmune antibodies was fully active, whereas the fraction treated with the anti-160K antibodies was inactivated almost completely (Fig. 2, lanes 5–7). However, when highly purified calf thymus CPSF (see Materials and methods) was

Figure 2. Immunodepletion of CPSF activity with anti-160K antibodies. Antibodies from rabbit anti-160K antisera and from the preimmune sera were conjugated to protein A–Sepharose. The resins were used to deplete aliquots of HeLa nuclear extract (lanes 2–4) or a crude HeLa CPSF fraction lacking other polyadenylation factors (lanes 5–8). In all samples, an SV40 late pre-mRNA was used. In lanes 5–8, reaction mixtures also contained HeLa PAP. The sample shown in lane 8 was supplemented with ~100 ng of CPSF extensively purified from calf thymus (see Materials and methods for details). RNA samples were resolved on a denaturing polyacrylamide gel. Numbers at left indicate sizes of DNA markers, in nucleotides.

added to the depleted HeLa CPSF fraction, activity was completely restored (lane 8). These results provide strong evidence that the 160K cDNA indeed encodes the functional large subunit of CPSF.

160K binds RNA selectively

As mentioned above, CPSF recognizes the AAUAAA signal sequence, and it is possible that binding is mediated by 160K. We therefore wished to determine whether recombinant 160K could bind RNA, and if so, whether such binding was sequence specific. For these experiments we used both *E. coli*-produced 160K (Fig. 3A, lane 1) and also 160K purified from recombinant baculovirus-infected sf21 cells (similar results were obtained with each; results with the bacterial protein are not shown). The experimental approach was to mix an sf21 cell nuclear extract highly enriched in recombinant 160K (Fig. 3A, lane 2) with a suspension of anti-160K antibodies bound to protein A–agarose. After washing, the beads were used in RNA-binding assays. (Fig. 3A; lane 3 displays the protein profile of the bound proteins, which indicates that apart from the antibodies, 160K is the only protein detectable, and relative to the antibodies is a small fraction of the total proteins.)

Immobilized 160K was incubated with several differ-

Figure 3. Recombinant 160K binds selectively to AAUAAA-containing RNAs. (A) Protein profiles. Aliquots of recombinant protein preparations utilized in this and subsequent experiments were resolved on an 8% SDS-polyacrylamide gel and stained with Coomassie blue. (Lane 1) 1.5 μ g of 160K expressed in and purified from *E. coli*; (lane 2) 10 μ g of 160K expressed from a recombinant baculovirus and partially purified (prior to Ni^{2+} agarose chromatography); (lane 3) 4 μ l of baculovirus-produced 160K bound to anti-160K-protein A resin (~ 1 μ g of 160K); (lane 4) 4 μ l of anti-160K-protein A resin. (B) 160K binds AAUAAA-containing RNAs. The indicated ^{32}P -labeled RNAs were prepared by *in vitro* transcription, mixed with baculovirus 160K bound to anti-160K-protein A-Sepharose, and the bound RNAs eluted and analyzed on a 5% denaturing polyacrylamide gel. (Lanes 1,5) Adenovirus L3 poly (A) site-containing RNA (wt); (lanes 2,6) L3 RNA with a U $\rightarrow$ C mutation in the AAUAAA hexanucleotide (mt); (lanes 3,4,7,8) RNAs derived from the PIP7A splicing substrate. (Lanes 1–4) RNAs prior to binding (input); (lanes 5–8) bound RNAs. (M) DNA size markers. (C) L3 RNAs are not bound by control resins lacking 160K. Binding reactions were performed as in B. (Lanes 1,3,5) wild-type L3 RNA; (lanes 2,4,6) mt L3 RNA; (lane 1,2) input RNAs; (lanes 3,4) RNA retained by anti-160K-protein A resin. (Lanes 5,6) RNAs retained by 160K conjugated to anti-160K-protein A resin.

ent ^{32}P -labeled RNAs in the presence of a large excess of tRNA. After washing, bound RNAs were eluted and displayed on denaturing polyacrylamide gels (Fig. 3B,C). A 165-base polyadenylation substrate RNA from the adenovirus L3 poly(A) site (Ryner et al. 1989b) was bound efficiently by the 160K immunobeads (cf. lanes 1 and 5). A mutant derivative of the L3 RNA containing a single-base change in the AAUAAA hexanucleotide (A $\rightarrow$ C) was also bound by 160K, but the binding efficiency was reduced by approximately a factor of 2. Although the effect of the point mutation was small, it was observed in multiple independent experiments. For example, Figure 3C presents a control experiment indicating that neither wild-type nor mutant L3 RNA bound detectably to immunobeads lacking 160K (lanes 3,4), whereas binding was again observed to 160K-containing beads. In this case, binding to the wild-type RNA was about three-fold more efficient than to the mutant RNA (lanes 5,6). More dramatic decreases in binding were seen with RNAs containing pre-mRNA splicing signals (PIP7A; Kohtz et al. 1994) as opposed to a poly(A) signal. Neither of two PIP7A RNAs was bound significantly by 160K (Fig. 3B, cf. lanes 3 and 4 with 7 and 8). Similar results were obtained with RNAs derived from pGEM plasmid se-

quence, whereas an RNA containing the SV40 late poly(A) site was bound as efficiently as the L3 substrate (data not shown). Together, these results indicate that 160K is an RNA-binding protein with a preference for AAUAAA-containing RNAs. However, the limited effect of the U $\rightarrow$ C mutation on binding suggests that sequence specificity of intact native CPSF is likely greater than that of the purified recombinant protein.

160K interacts functionally with CstF and PAP

As mentioned in the introductory section, CPSF functions in both the cleavage and poly(A) synthesis phases of 3'-end formation. In addition to binding the pre-mRNA, this involves cooperative interactions with CstF and PAP, respectively (Murthy and Manley 1992 and references therein). We therefore wished to determine whether 160K contributes to these interactions.

The CPSF-CstF-RNA interaction can be detected by gel shift or UV cross-linking assays. The latter measures label transfer from a ^{32}P -labeled RNA to the 64-kD RNA-binding subunit of CstF (64K). CPSF and CstF by themselves give rise to weak cross-linking, at most, but mixtures of the two factors result in strong and specific

cross-linking of 64K (Fig. 4A, lanes 1–3; see also Wilusz et al. 1990; Gilmartin and Nevins 1991; Murthy and Manley 1992). To test whether 160K is able to enhance 64K cross-linking by itself, cross-linking experiments were performed with the adenovirus L3 pre-mRNA and human CstF (Takagaki et al. 1990) plus either calf thymus CPSF (Murthy and Manley 1992) or recombinant 160K containing an amino-terminal histidine tag expressed in and purified from *E. coli* [Fig. 3A, lane 1; essentially identical data (not shown) were obtained with the baculovirus-encoded 160K]. Figure 4A shows that increasing concentrations of 160K resulted in significant cross-linking of 64K when mixed with a constant amount of CstF (lanes 5,7,9). In contrast, essentially no cross-linking was detected when CstF (lane 1), CPSF (lane 2), or 160K (lanes 4,6,8) was incubated alone with the RNA. As mentioned above, point mutations in AAUAAA reduce or eliminate 64K cross-linking. This is shown in Figure 4B, lanes 3 and 4, which indicate a significant decrease in CPSF-dependent cross-linking of 64K when the L3 AAUAAA was changed to AACAAA. [Our previous experiments (Murthy and Manley 1992) showed a more complete inhibition of cross-linking by this mutation. The reason for the difference is not known.] However, when the mutant RNA was used with recombinant 160K plus CstF, only a very slight reduction in cross-linking was detected (Fig. 4B, lanes 6,7). This may reflect the relatively small effect of this mutation on RNA binding by 160K (see above) and further suggests that additional CPSF subunits are required for optimal specificity.

160K also enhanced cross-linking of 64K to two additional poly(A) site-containing RNAs, SV40 late (Fig. 4C) and HIV (data not shown). In the case of SV40, a low level of cross-linking was detected when CstF was incubated alone with the RNA (lane 1), but this was increased greatly when 160K was added (lane 3). The effect of an AAUAAA mutation (U → A) was again slight, although with this RNA two- to threefold reductions were reproducibly detected (e.g., cf. lanes 3 and 4). The faint band indicated by the asterisk corresponds to possible cross-linking of 160K. The reason for the weak 160K cross-linking is unclear. However, we note that although efficient cross-linking of 64K has been observed consistently in a number of previous studies, cross-linking of species that appear to correspond to 160K has been more variable. This has been commented on previously (Murthy and Manley 1992) and is discussed further below in light of our current findings that 160K does bind RNA directly. In any event, our results indicate that recombinant 160K can partially reconstitute the cooperative CPSF–CstF interaction that specifies the poly(A) site, suggesting that 160K binds one (or more) subunit of CstF.

We next wished to examine the role of 160K in the poly(A) synthesis stage of the reaction. We first tested whether 160K could substitute for intact CPSF in AAUAAA-dependent poly(A) synthesis, using a pre-cleaved SV40 late pre-mRNA as substrate and PAP isolated from HeLa cells or from recombinant baculovirus-infected sf21 cells (see Materials and methods). Under conditions where native CPSF gave rise to efficient

Figure 4. Recombinant 160K cooperates with CstF in binding pre-mRNA. The indicated proteins and RNAs were incubated, exposed to UV light, treated with RNase and the proteins resolved by SDS-gel electrophoresis. Samples contained the appropriate 32 P-labeled RNA, ~100 ng of HeLa CstF, ~100 ng of calf thymus CPSF, and/or the indicated amount of recombinant 160K purified from *E. coli*. In all cases, the predominant cross-linked protein detected was the 64-kD subunit of CstF (64K). (A) 64K cross-linking is induced by recombinant 160K. Cross-linking assays were performed with L3 pre-mRNA. (Lane 1) CstF alone; (lane 2) CPSF alone; (lane 3) CstF + CPSF; (lanes 4,6,8) 160K alone; (lanes 5,7,9) 160K + CstF; (lanes 4,5) 100 ng of 160K; (lanes 6,7) 200 ng of 160K; (lanes 8,9) 400 ng of 160K. (B) Cross-linking to wild-type and mutant L3 pre-mRNAs. The indicated combinations of proteins were incubated and subjected to cross-linking as in A. (wt) L3 wild-type pre-mRNA; (mt) L3 pre-mRNA containing a U → C mutation in AAUAAA. (C) Cross-linking to wild-type and mutant SV40 late pre-mRNAs. The indicated proteins were used in cross-linking assays as in A. (wt) Wild-type SV40 late pre-mRNA; (mt) SV40 late pre-mRNA containing a U → A mutation in the AAUAAA hexanucleotide. The asterisk indicates a faint band that may correspond to 160K cross-linking.

poly(A) synthesis, recombinant 160K, isolated from either bacteria or insect cells, was inactive (data not shown), suggesting that one or more additional CPSF subunit is required for this activity. To address the possible involvement of 160K in poly(A) synthesis in another way, we asked whether addition of the recombinant protein to reaction mixtures could influence poly(A) synthesis. (As above, *E. coli* and baculovirus-produced 160K proteins behaved nearly indistinguishably; results with the bacterial protein are not shown.) We first tested the effect of 160K on AAUAAA (i.e., CPSF)-dependent polyadenylation, using calf thymus CPSF and HeLa PAP. The results (Fig. 5A) show that increasing concentrations of 160K caused a severe reduction in accumulation of polyadenylated RNA (lanes 3–5). Inhibition was not observed when 160K was preheated (to 65°C or 100°C; lanes 8,9) for 5 min, arguing against the involvement of a nonprotein contaminant, nor was an effect observed when 160K was added following completion of poly(A) synthesis, indicating that inhibition was not attributable to a contaminating poly(A) nuclease (data not shown; see below). Fivefold more BSA than the maximum amount of 160K also had no effect (lane 6), and the inhibition was reversible, as it could be largely overcome by addition of excess CPSF (lane 7). These findings indicate that free 160K can have a negative effect on the activity of CPSF, PAP, or both.

To investigate further the inhibition of polyadenylation caused by 160K, we tested the effect of recombinant 160K on nonspecific poly(A) synthesis. In the presence of Mn^{2+} , PAP can synthesize poly(A) independently of CPSF and can utilize essentially any RNA as a primer. Figure 5B shows that the nonspecific reaction was also very effectively inhibited by 160K. Inhibition was more complete and required less 160K than observed in the specific reaction. Again, no significant inhibition was detected when 160K was preheated (lane 6) or added at the end of the reaction (lane 7), nor was inhibition observed with unrelated proteins [BSA (lane 8) or recombi-

nant His-tagged TFIIB; data not shown]. Inhibition was also fully overcome by addition of excess PAP (data not shown). Nonspecific polyadenylation of an unrelated RNA [a 140-nucleotide PIP7A RNA; see above] was also strongly inhibited by 160K (lanes 9–13). This finding suggests that inhibition was likely not attributable to competitive RNA binding by 160K, as the results in Figure 3B showed that 160K did not bind the PIP7A RNA whereas the SV40 RNA was efficiently bound. Together, our results indicate that the 160K subunit of CPSF can inhibit polyadenylation and that this likely reflects an interaction involving 160K and PAP.

160K binds to the 77-kD subunit of CstF and to PAP

The experiments just described are consistent with the idea that 160K is responsible, at least in part, for the previously documented interactions between CPSF and CstF and between CPSF and PAP. The following coimmunoprecipitation experiments were performed to address this more directly. With respect to the interaction with CstF, a recent structure–function analysis of the three CstF subunits, 77K, 64K, and 50K, provided evidence that only one, 77K, could interact with CPSF (Y. Takagaki and J.L. Manley, in prep.). To test whether this reflects the postulated interaction with 160K, we carried out experiments similar to those used to examine RNA binding by 160K. Specifically, anti-160K antibodies were bound to protein A–agarose beads, which were then bound to recombinant baculovirus-encoded 160K (see Fig. 3A for protein profiles). An aliquot of this suspension was mixed with in vitro-translated (IVT) 77K (or other proteins, see below) and, after extensive washing (see Materials and methods), bound proteins were eluted and analyzed by SDS–gel electrophoresis. The results (Fig. 6A) indicate that 77K bound tightly to the resin containing 160K (lane 3), but not to beads lacking 160K (lane 2). Binding was quite efficient, as 30%–50% of the input 77K was retained by the 160K resin. Essentially

50, 100, or 200 ng of 160K, respectively; (lane 6) 200 ng of 160K preheated to 100°C for 5 min; (lane 7) 200 ng of 160K added after 20 min and incubation continued for an additional 20 min; (lane 8) 1 μg of BSA instead of 160K.

Figure 5. Recombinant 160K inhibits specific and nonspecific poly(A) synthesis. Specific (A) and nonspecific (B) polyadenylation reactions were performed as described in Materials and methods. RNA products were resolved by denaturing polyacrylamide gel electrophoresis. (A) Reaction mixtures contained 50 ng of calf thymus CPSF, 10 ng of HeLa PAP, and the SV40 late precleaved pre-mRNA. Recombinant 160K was added as follows: (Lanes 3–5) 100, 200, or 300 ng of 160K, respectively; (lane 7) 1.5 μg of BSA instead of 160K; (lane 6) 300 ng of 160K plus 60 ng of additional CPSF; (lanes 8–9) 300 ng of 160K preheated to 65°C or 100°C for 5 min, respectively. (B) Reaction mixtures contained 20 ng of HeLa PAP and either SV40 late (lanes 1–8) or PIP7A (lanes 9–13) pre-mRNA. Recombinant 160K was added as follows: (Lanes 3–5, 11–13)

identical results (data not shown) were obtained when His-tagged 160K produced in *E. coli* was purified, adsorbed to Ni^{2+} resin, and used in binding assays with IVT 77K. These experiments demonstrate a strong interaction between 160K and the CstF 77K subunit, which offers an explanation for the previously observed CPSF–CstF cooperative RNA binding.

Next, we wished to address the possible existence of a direct interaction between 160K and PAP. We first used a protocol similar to that just described to detect 77K binding to 160K, except for the substitution of IVT PAP for IVT 77K. Two other IVT proteins were used as controls for specificity. The U1 snRNP-specific 70kD protein (U1 70K) was shown previously to interact with the splicing factor ASF/SF2 in an assay nearly identical to that employed here (Kohtz et al. 1994), but there is no evidence that U1 70K functions in polyadenylation. PAP3 is a truncated form of PAP that can arise from alternative splicing (Wahle et al. 1991; Raabe et al. 1994). However, the recombinant protein, which is inactive by itself, is unable to block AAUAAA-dependent poly(A) addition, suggesting that it cannot interact with CPSF (W. Zhao and J.L. Manley, unpubl.). As with 77K, binding of full-length PAP to immuno-beads containing but not lacking 160K was detected (Fig. 6B, lanes 1, 4 and 7; in lane 7, there is some distortion because of overloading of the antibody H-chain). Identical results (data not shown) were obtained with the *E. coli*-produced 160K immobilized on Ni^{2+} resin. Binding was specific, and likely to reflect the functional interaction between PAP and CPSF, as little if any PAP3 (lanes 2,5,8) and no U1 70K (lanes 3,6,9) bound 160K. However, in contrast to the efficient binding of 77K, <5% of the input PAP was retained by the 160K resin. Also suggesting a weak interaction, significant amounts of PAP were found in the wash fractions, whereas PAP3 and U1 70K were exclusively in the unbound supernatant (data not shown).

Although the above findings are consistent with previous observations that PAP–CPSF–RNA ternary complexes are less stable than CstF–CPSF–RNA complexes (Murthy and Manley 1992; see also Bienroth et al. 1993)—and it is not unreasonable that the CPSF–PAP interaction should be relatively weak—we nonetheless wished to obtain further evidence that the observed interaction was significant. Therefore, as a complementary approach, we asked whether IVT 160K could be coimmunoprecipitated with recombinant PAP. For this, His-tagged PAP purified from baculovirus-infected sf21 cells was bound to an anti-PAP antibody–protein A resin, which was then tested for its ability to coprecipitate IVT 160K. The results (Fig. 6C, lanes 1–3), indicate that 160K was retained by the resin but not significantly by control beads containing the anti-PAP antibodies but not PAP. Taken together, our results provide strong support for the existence of a specific interaction between 160K and PAP.

Discussion

Previous studies have indicated that CPSF plays a key role in pre-mRNA 3'-end formation, recognizing the AAUAAA signal sequence and interacting with other factors to bring about cleavage and poly(A) addition. The data described here indicate that the 160-kD subunit plays a central role in all these processes. It recognizes RNAs containing AAUAAA with specificity; it cooperates with CstF in RNA binding, very likely through a direct interaction with the CstF 77K subunit; and it binds to and influences the activity of PAP. However, in none of these three situations can 160K fully substitute for CPSF: The RNA binding specificity of 160K appears less than that of CPSF, 160K cooperates less efficiently with CstF, and 160K inhibits rather than enhances poly(A) synthesis. An intriguing scenario, then, is that

Figure 6. 160K interacts specifically with the 77-kD subunit of CstF and with PAP. In vitro-translated (IVT) [^{35}S]-methionine-labeled proteins were mixed with either recombinant 160K (A,B) or PAP (C) that had been immobilized on the appropriate antibody–protein A resin. After washing, precipitated proteins were analyzed by SDS–gel electrophoresis. Numbers to the left of each panel indicate the sizes of marker proteins (in kD). (A) IVT 77K binds recombinant 160K. IVT 77K (lane 1) was mixed with either 160K prebound to anti-160–protein A resin (lane 3) or the same resin lacking 160K (lane 2). (B) Full-length IVT PAP specifically binds recombinant 160K. The indicated IVT proteins (PAP, lanes 1,4,7; PAP3, lanes 2,5,8; and U1 70K, lanes 3,6,9) were mixed with either 160K prebound to anti-160–protein A resin (lanes 7–9) or the same resin lacking 160K (lanes 4–6). Equal counts per minute of each IVT protein were used in binding reactions. (C) IVT 160K binds recombinant PAP. IVT 160K (lane 1) was mixed with either PAP prebound to anti-PAP–protein A resin (lane 3) or the same resin lacking PAP (lane 2).

the function of the other CPSF subunits is to enhance or modify activities that are intrinsic to 160K. Below we discuss the significance of these activities, and conclude by speculating how these properties of 160K, and its sequence, suggest a possible mechanistic and evolutionary link between yeast and mammalian polyadenylation.

The idea that the 160-kD subunit of CPSF interacts with RNA has been suggested previously based on the results of UV cross-linking experiments. The first such indication was provided from studies employing unfractionated nuclear extracts, before the existence of separable polyadenylation factors had even been uncovered (Moore et al. 1988). However, several other studies (e.g., Wilusz and Shenk 1988; Takagaki et al. 1990; Wilusz et al. 1990; Murthy and Manley 1992) failed to observe significant cross-linking of a 160-kD protein, and the experiments described here detected, at most, very weak 160K cross-linking. The relevance of this cross-linking has been unclear. However, our demonstration that 160K binds directly and selectively to AAUAAA-containing pre-mRNAs strongly supports the significance of the cross-linking, as has recent data indicating that a cross-linked species of 160 kD can be immunoprecipitated from nuclear extracts by anti-CPSF monoclonal antibodies (Jenny et al. 1994). The failure of some studies to detect 160K cross-linking probably reflects subtle technical differences in experimental protocols and/or a relatively low "cross-linkability" of 160K. In contrast, efficient cross-linking of the 64K subunit of CstF has been detected consistently in these and other related studies. Whether this reflects stronger binding by 64K, or probably more likely, differences in cross-linking efficiency, is not clear.

The RNA-binding properties of 160K described here strongly support the idea that this polypeptide plays a key role in the recognition of AAUAAA. However, our data also suggest that it is not sufficient, as the specificity of binding appears to be less than that observed with intact CPSF (e.g., Bardwell et al. 1991; Keller et al. 1991). One explanation is that the recombinant protein was not folded properly or was disrupted in some way during purification. But the fact that 160K purified either from *E. coli* by virtue of a His tag or from baculovirus-infected insect cells by immunoaffinity behaved indistinguishably tends to argue against this. We propose instead that 160K is responsible for directly contacting the RNA and recognizing AAUAAA but that the specificity of the interaction is enhanced by other CPSF subunits. This seems unlikely to involve direct contact of the RNA by another CPSF subunit. Although it has been shown that a 30-kD polypeptide associated with CPSF can be cross-linked to RNA in an AAUAAA-dependent manner (e.g., Jenny et al. 1994), the polypeptide appears not to be present in other fully active CPSF preparations (Manley and Murthy 1992; Gilmartin et al. 1995) and therefore could not be necessary for specificity. Instead, we suggest that one or both of the other two subunits (100 or 73K) influences the conformation of 160K so as to increase sequence-specific binding. For example, the observation that 160K can contact both AAUAAA and an up-

stream element in HIV pre-mRNA (Gilmartin et al. 1995) makes it very likely that 160K contains (at least) two distinct RNA-binding domains. Perhaps in the context of the free 160K subunit, this can result in an apparent relaxation of binding specificity, allowing RNA-protein contacts that might not occur in the context of intact CPSF. With respect to RNA-binding domains, it was somewhat unexpected that no clear matches to established RNA-binding consensus sequences (e.g., see Burd and Dreyfuss 1994) were detected. The weak match to the RNP-type RBD, if significant, would indicate a highly diverged RBD, and if 160K contains more than one RNA-binding domain, at least one would need to represent a novel motif.

The fact that 160K interacts with the 77K subunit of CstF offers an explanation for genetic properties of the *Drosophila* homolog of 77K, the suppressor of forked [su(f)] protein (Takagaki and Manley 1994), which in turn offers insights into the basis of poly(A) site strength more generally. Certain mutations in *su(f)* are known to modify the phenotypes of transposable element insertions in the introns of unlinked genes (Mitchelson et al. 1993 and references therein). Given the role of 77K/su(f) in polyadenylation, it is now apparent that this results from changes in relative efficiencies of 3'-end formation. However, because the function of 77K/su(f) does not involve RNA binding, it has not been clear how mutations could influence the relative utilization of poly(A) sites. Based on our data, we propose that at least some of the modifier mutations in *su(f)* affect the 77K–160K interaction, which would influence cooperative RNA binding by CPSF–CstF, and therefore affect the efficiency of poly(A) site recognition. It is likely that these cooperative interactions may be more important for some sites than they are for others. For example, poly(A) sites with weak signal sequences (i.e., suboptimal AAUAAA and G/U-rich elements) would likely be more dependent on cooperativity than would sites with strong signals. This offers an explanation for how mutations in *su(f)* could affect some sites more than others. An extension of this idea is that cooperative interactions between CPSF and CstF (i.e., the 77K–160K interaction) is an important factor in determining poly(A) site strength.

As discussed above, a good deal of previous data suggested an interaction between CPSF and PAP. Thus, our finding of an interaction between PAP and 160K is unexpected only because it suggests that 160K is responsible for, or at least plays a major role in, all of the known functions of CPSF (i.e., AAUAAA recognition and CstF and PAP binding; it is not known how the cleavage factors are recruited to the complex, but this may also require interactions with CPSF). What was more surprising was the dominant-negative effect of free 160K on both specific and nonspecific poly(A) synthesis. The strong inhibition would appear greater than might be expected from the relatively weak protein–protein interaction, although this likely reflects the different conditions of the two assays. It is possible that this inhibition is attributable to an aberrant property of free 160K that is not physiologically relevant. However, another possibility is that

it may be related to the mechanism by which synthesis of the poly(A) tail is terminated. Three factors, CPSF, PAP, and poly(A)-binding protein II, are required for maximal, processive synthesis of ~200-nucleotide poly(A) tails, and in the absence of CPSF poly(A) tail synthesis can continue beyond this, generating chains up to 400 nucleotides (Wahle 1991). This is consistent with the notion that CPSF may not only play a role in facilitating poly(A) synthesis but also in terminating it. Perhaps dissociation of CPSF from AAUAAA during elongation (Sheets and Wickens 1989) leads to a conformational change so that rather than enhancing the activity of PAP, CPSF now inhibits it, in a manner reflecting the ability of free 160K to inhibit poly(A) synthesis.

An intriguing aspect of polyadenylation has been the question of how conserved the process is between lower and higher eukaryotes. Unlike most other basic cellular processes, where similarities outweigh the differences, the situation with polyadenylation has been less clear. Although it is known that the reaction in yeast also involves endonucleolytic cleavage and poly(A) addition (Butler and Platt 1988) and requires multiple separable factors (Chen and Moore 1992), some confusion has existed because of the lack of any detectable consensus sequence (e.g., AAUAAA) in the RNA. This has led to speculation that the factors involved, at least those that specify the poly(A) site, may be distinct. However, recent experiments have provided evidence that some of these factors are related and, together with the data presented here, allow us to propose a model that provides evolutionary and mechanistic links between poly(A) site specification in yeast and metazoans.

Our cloning of the 77-kD subunit of CstF (Takagaki et al. 1994) not only led to the conclusion that the protein is the human homolog of su(f) in *Drosophila* but also that it is very likely the homolog of the yeast protein RNA14. [The homology between su(f) and RNA14 had been noted previously; Mitchelson et al. 1993]. Furthermore, a significant similarity exists between the 64-kD subunit of CstF and RNA15, which had been linked to RNA14 by genetic studies (Minvielle-Sebastia et al. 1991). Based on this, we proposed that RNA14/15 is the yeast counterpart of CstF. RNA 14 and 15 have now been shown to be required for yeast polyadenylation in vitro and to be subunits of the factor CF1 (Minvielle-Sebastia et al. 1994). However, CF1 is necessary for both cleavage and polyadenylation (Chen and Moore 1992) and thus would appear to be functionally more related to CPSF than to CstF. One possibility is that further purification of CF1 will yield two factors, one equivalent to CPSF and the other to CstF. However, there is no AAUAAA in yeast and at least the recognition function of CPSF may therefore not be required. Thus another possibility is that there is no CPSF homolog in yeast and the CstF equivalent (RNA 14/15 or CF1) plays the central role in poly(A) site specification. Perhaps the shorter and simpler yeast primary transcripts need only one RNA-binding factor, and AAUAAA and CPSF evolved together with the longer, more complex transcription units found in metazoans.

But how does CF1 communicate with PAP, given that CstF does not? The answer is provided by a recent study characterizing a yeast protein, FIP1, which binds to both RNA14 and PAP, thereby providing the necessary bridging function (Preker et al. 1995). It is striking that these are the yeast equivalents of the two mammalian proteins we have shown here to interact with 160K. This suggests that there may be some similarity between 160K and FIP1 that escaped detection by standard data base searches. We therefore compared the two protein sequences directly, using the BESTFIT algorithm and found a weak but extensive similarity (Fig. 7). By introducing multiple gaps (10) in the FIP1 sequence (such that its 327 residues are spread over 470 160K residues), a match with 19% identity, 33% similarity allowing only biochemically conservative changes, and 48% similarity with evolutionarily conserved changes was apparent. The longest uninterrupted similarity encompasses the 95 carboxy-terminal FIP1 residues, which are 19% identical and up to 70% similar. Although further work is required to evaluate the significance of this similarity, we suggest that 160K has evolved to incorporate the functions of FIP1 through the use of structurally related domains.

In summary, we have shown that the 160-kD subunit of CPSF plays a key role in coordinating pre-mRNA poly-

```

FIP1  MSSSED.....EDDKFLYGS...DSELAL..... 21
      :.:.:.....:..:..:..:..:..:..:..:..:
160K  LGSRQSRPYLLVHVDQELLIYEAFPHDSQLGQGNLKVRFKKVPHNINPRE 903

FIP1  ....PSSKRSRDDADAGASSNPDIK..... 44
      ||.:.:..:..:..:..:..:..:..:..:
160K  KKPSPKSKKKAEGGGAEGAGARGVARFRYPEDIYGYSGVFICGSPSHWL 953

FIP1  .....RQKFDSPVEETPATARDRSDEDIYSDSSDDSDSDLEV 83
      :.:.:..:..:..:..:..:..:..:..:
160K  LVTGRGALRLHPMAIDGPDVDFAPFHNVCPRGFLYFNQGEIRISVLPA 1003

FIP1  IISLGPD.....PTRLDKLLDSYSTAATSSSKDVISVAT 118
      :.:.:..:..:..:..:..:..:..:..:
160K  YLSYDAPWPVRKIPLRCTAHYVAYHVESKVVAVATSTNTPCAR.IPRMTG 1052

FIP1  DVSNTITKTSDERLI.....TEGEANQGVTA 144
      :.:.:..:..:..:..:..:..:..:..:
160K  EEKEFETIERDERYIHPQAEAFSIQLISPVSWEAIPNARIELQWEHVTC 1102

FIP1  TTVKATESDGNVP.....KAMTGSIDLKKEGIFDSVGITTIDPE.. 183
      :.:.:..:..:..:..:..:..:..:..:
160K  MKTVSLRSEEVTSGSLKGYVAAGTCLMQGEVTCRGRILIMDVIEVVPPEG 1152

FIP1  .....VLKEKPWRQPGANLSY.....FNYGFNEFTW 210
      ||.:.:..:..:..:..:..:..:..:..:
160K  QPLTKNKFVLYEKEQKGPVTALCHCNHLSVAIGQKIFLWSLRASELTG 1202

FIP1  MEYLHRQEKQQDYNPRILMG.....LLSLQQQKLNANDTDSNL 252
      |.:.:..:..:..:..:..:..:..:..:
160K  MAFIDTQLYIHQMISVKNFILAADVMSKISLLRYQEESKTLSLVSRDAKP 1252

FIP1  GNIIDNNNNVNNANMSNLNSNMGMSGTNPAPPMPHSPFPLPMFGSF 302
      :.:.:..:..:..:..:..:..:..:..:
160K  LEVYSVDFMVDNAQLGFLVSDRDLNLMVYMLFEA...KESFGGMRLRR. 1299

FIP1  PPFPMFGMPFPMNQPNQPNQNSK 327
      :.:.:..:..:..:..:..:..:..:..:
160K  ADFHVGAVHNTFWRTFCRATEGLSK 1324

```

Figure 7. Similarity between human 160K and yeast FIP1. The amino acid sequences (single-letter code) of 160K and FIP1 are compared. The amino acids of each protein are numbered at right. Dots in the sequence indicate gaps. (Vertical lines) Identical residues; (double dots) biochemically related residues; (single plus double dots) evolutionarily related residues.

adenylation. It contacts the RNA to facilitate recognition of the AAUAAA signal sequence, and it participates in protein-protein interactions with other factors to bring about assembly of the polyadenylation complex and synthesis of the poly(A) tail. Given these multiple and fundamental functions, it seems probable that 160K will be a target of factors that regulate polyadenylation. We suspect that future studies on 160K will provide additional insights into the mechanism and regulation of pre-mRNA polyadenylation.

Materials and methods

Purification of CPSF and peptide microsequencing

CPSF was purified from calf thymus as described previously (Murthy and Manley 1992). Approximately 100 μ g of the phenyl Superose fraction was electrophoresed on a preparative 8% SDS-polyacrylamide gel and the resolved proteins were stained with Coomassie blue. The 160-kD band was excised from the gel, washed extensively with water, and partially digested *in situ* with *Staphylococcus aureus* V8 protease (Sigma) as described (Kennedy et al. 1988). The cleaved peptides were separated on an 15% SDS-polyacrylamide gel, transferred to an Immobilon-P membrane (Millipore), and stained with Coomassie blue (Matsudaria 1987). Three well-separated bands were cut from the filter and subjected to peptide microsequencing. Sequencing was performed by M. Gavinowicz of the Columbia HHMI Protein Analysis Facility.

cDNA cloning

From the amino acid sequence (LDASGFATQGPTVFAGNIG) of the longest of the three peptides, a 44-mer inosine-containing degenerate oligonucleotide (5'-GGITTYGCIACICARGGICCIACIGTITTYGCIIGGIAAYATIGG-3') was synthesized and end-labeled by phosphorylation. The labeled 44-mer was used to screen 1×10^6 plaques of a bovine heart cDNA library (Clontech) in duplicate. Hybridization was performed at 50°C in 6 $\times$ SSC, 5 $\times$ Denhardt's solution, 0.5% SDS, and 100 μ g/ml of salmon sperm DNA. Four positive clones were identified and subcloned into pBluescript (Stratagene) for sequencing. The DNA fragment (3.3 kb) from the longest insert-containing clone, which encoded the 19-amino-acid peptide, was gel purified, labeled by random priming, and used to screen a HeLa cDNA library in the λ EXIox(+) vector (a kind gift from Dr. Jian Wu, Rockefeller University, NY) to isolate full-length cDNAs. Twenty positives from 1.0×10^6 plaques were purified and plasmids were excised *in vivo*. Purified plasmids were digested with *Eco*RI and subjected to Southern blot analysis. One of the positive clones, containing a 4.5-kb insert (clone 27), was subcloned into pBluescript and sequenced in both directions using nested deletions. The nucleotide sequence of the insert was determined by dideoxy sequencing of single-stranded DNA with Sequenase 2.0 (U.S. Biochemical).

Generation of antibodies

A *Sa*II fragment of the 160K cDNA (coding for the amino-terminal 927 residues) was ligated into the polylinker region of the expression vector pDS56-6H (Gentz et al. 1989) and then transformed into strain JM101. Following induction with IPTG for 2 hr, bacterial extracts were prepared by resuspending the cells in lysis buffer (20 mM Tris-HCl at pH 8.0; 100 mM NaCl; 0.5 mM DTT, and 100 μ g/ml of lysozyme), followed by sonication and centrifugation. Pellets were resuspended in 6 M guanidine hy-

drochloride and incubated at room temperature for 1 hr with gentle shaking. After brief centrifugation, the supernatant was applied to a Ni²⁺ nitriloacetic acid affinity column. Proteins were eluted with a step gradient of imidazole (5, 20, 40, and 100 mM) and analyzed by SDS-polyacrylamide gel electrophoresis. Fractions from the 100 mM imidazole eluate, which contained most of the recombinant protein, were dialyzed against PBS. Antisera were prepared in rabbits by the Pocono Rabbit Farm.

Expression of 160K in *E. coli* and insect cells

To construct an *E. coli* expression plasmid, an *Eco*RV site (in the polylinker region of pBluescript) was converted to a *Bam*HI site. The *Bam*HI fragment, coding for the full-length 160-kD protein, was cloned into the *E. coli* expression vector pRSETC (Invitrogen). A *Nhe*I-*Nar*I fragment (from the linker region to immediately upstream of the translation initiation codon ATG) was deleted to remove the 5'-untranslated region. The recombinant protein contained 6 histidines encoded by the vector at its amino terminus. *E. coli* BL21 (DE3) cells were transformed with this plasmid and protein expression was induced with IPTG. Although most of the induced protein was in inclusion bodies, a significant amount was present in the soluble fraction. This was purified by Ni²⁺ affinity chromatography (40 mM imidazole wash, 100 mM elution) and dialyzed against buffer D (20 mM HEPES at pH 7.9, 20% glycerol, 50 mM (NH₄)₂SO₄, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF). The final concentration was ~150 μ g/ml.

The entire coding region of 160K, with the amino-terminal His tag (*Nde*I-*Bam*HI fragment from pRSETC), was inserted at the *Eco*RI site of the baculovirus transfer vector pVL1392 (PharMingen). Recombinant viruses were prepared by cotransfecting sf21 insect cells with a mixture of 0.1 μ g of BaculoGold linearized virus DNA (PharMingen) and 1 μ g of recombinant transfer vector. After transfection, recombinant baculoviruses were isolated, amplified, and titered as described (Miller et al. 1986). To express 160K, sf21 cells were infected with recombinant viruses at a m.o.i. of 5. Forty hours after infection, cells were washed three times with ice-cold PBS and suspended in lysis buffer (50 mM Tris-HCl at pH 8.0, 150 mM NaCl, 1% NP40, 0.1% aprotinin, 1 mM DTT, 0.35 mM PMSF). The lysate was centrifuged after incubation on ice for 30 min. Most of the fusion protein was in the nuclear pellet, which was resuspended in buffer D and sonicated. The supernatant was dialyzed against buffer D lacking EDTA after removing the insoluble material by centrifugation. Recombinant 160K was purified further by a Ni²⁺ affinity column, dialyzed against buffer D, and stored at a final concentration of 150 μ g/ml. All experiments presented here that utilized recombinant 160K were done with both bacterial and insect cell preparations, which in all cases gave nearly indistinguishable results.

Immunodepletion

Immunodepletion experiments were carried out as described previously (Takagaki et al. 1990). Briefly, 100 μ l of protein A-agarose beads was incubated with 500 μ l of either preimmune or anti-160K antisera for 8 hr at 4°C. Beads were washed extensively with IPP buffer (10 mM Tris-HCl at pH 7.9, 500 mM NaCl, 0.1% NP-40) and then equilibrated with buffer D. Five hundred-microliter aliquots of 1:5 diluted HeLa nuclear extract or CPSF Mono Q fraction (Takagaki et al. 1989) were added to the preimmune or immune beads, or left untreated, and the samples were gently mixed overnight at 4°C. After a brief centrifugation, supernatants were recovered and 5- μ l samples were assayed for CPSF activity in 25- μ l specific polyadenylation assays as described (Murthy and Manley 1992; see below).

RNA binding

³²P-labeled RNA substrates were synthesized by in vitro transcription. Antibodies were conjugated to 20 µl of protein A beads as described for immunodepletion experiments. After equilibrating the beads with NET buffer [50 mM Tris-HCl at pH 7.9, 150 mM NaCl, 0.05% NP-40], 5 µg of purified, recombinant 160K was added and incubated at 4°C overnight. Beads were washed five times with the same buffer and once with binding buffer [13 mM HEPES at pH 7.9, 28 mM (NH₄)₂SO₄, 33 mM KCl, 1 mM MgCl₂, 0.08 mM EDTA, 0.2 mM DTT, 0.2 mM PMSF]. Ten nanograms of labeled RNA substrate in 300 µl of binding buffer containing 0.4 µg of tRNA was added to the beads and incubated at 4°C for 1 hr. After extensive washing with binding buffer, RNA was recovered by protease K digestion and phenol/chloroform extraction, and fractionated on a 5% denaturing polyacrylamide gel.

Polyadenylation reactions and UV cross-linking analysis

Plasmids pG3SVL-A and pG3L3-A, which contain the SV40 late and adenovirus-2 L3 polyadenylation sites, respectively, were digested with *HpaI* and *BamHI*, respectively, and used as templates to synthesize ³²P-labeled RNA substrates (Ryner et al. 1989b). Specific and nonspecific polyadenylation reaction conditions were essentially as described previously (Murthy and Manley 1992). CPSF was purified from calf thymus through the poly(U) step (Murthy and Manley 1992) and further purified by Superose 6 chromatography. PAP was purified through the Mono S step of Ryner et al. (1989a), or in some cases homogeneous recombinant bovine PAP was purified from baculovirus-infected sf21 cells (gift of D. Foukal, Columbia University, New York). A standard 12.5-µl reaction mixture contained ~50 ng of CPSF, the equivalent of 5–10 ng of purified PAP, and 1–2 ng of labeled RNA. Reaction mixtures were incubated at 30°C for either 30 min (specific) or 20 min (nonspecific), and the products were purified and analyzed on 5% denaturing gels.

UV cross-linking experiments were performed using polyadenylation reaction conditions as described above and in Murthy and Manley (1992). CstF (Mono S fraction; Takagaki et al. 1990), CPSF [poly(U) fraction further purified by Superose 6 chromatography], and/or recombinant 160K protein were added in the indicated amounts and combinations to reaction mixtures containing ³²P-labeled RNA substrates. After incubation at 30°C for 10 min, samples were irradiated with a handheld shortwave UV light for 10 min at 4°C, digested with RNase A, and separated on 10% SDS-polyacrylamide gels.

In vitro transcription/translation and protein-binding assays

Plasmids were linearized at appropriate restriction sites 3' of the stop codon to generate templates for in vitro RNA synthesis with SP6 (for PAP and PAP3) or T7 (for CstF 77K) polymerase and the resulting RNAs were translated in rabbit reticulocyte lysate (Promega) in the presence of [³⁵S] methionine according to the manufacturer's suggested conditions. A coupled transcription/translation system (TNT; Promega) was used to synthesize labeled 160K and the U1 snRNP 70K protein. Purified baculovirus-expressed 160K or PAP (5 µg) was conjugated to antibody-protein A beads as described above. The final wash was with IPBB [Kohtz et al. 1994, 13 mM HEPES at pH 7.9, 28 mM (NH₄)₂SO₄, 100 mM KCl, 0.5 mM DTT, 0.2 mM PMSF, 0.1% Tween 20]. In vitro-translated proteins (5–10 µl) were incubated with resin-bound proteins in 300 µl of IPBB containing 2% BSA for 4 hr at 4°C with rocking. Beads were then washed three times with IPBB + 100 mM KCl and once with IPBB + 200 mM KCl. Proteins were eluted by boiling and analyzed by SDS-gel electrophoresis.

Acknowledgments

We are grateful to Yoshio Takagaki, Tobias Raabe, and Wenqing Zhao for plasmids, Jian Wu for the HeLa cDNA library and Diana Foukal for purified recombinant PAP and anti-PAP antibodies. We thank Y. Takagaki, J. Kohtz, W. Zhao, and R. Tacke for helpful discussions and M. Wang and L. Zhong for technical assistance. This work was supported by National Institutes of Health grant GM 28983.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Note added in proof

The putative amino acid sequence of bovine 160K was recently published (Jenny, A. and W. Keller. 1995. *Nucleic Acids Res.* **23**: 2629–2635.) and is 96% identical to the human sequence reported here.

References

- Bardwell, V.J., M. Wickens, S. Bienroth, W. Keller, B. Sproat, and A.I. Lamond. 1991. Site-directed ribose methylation identifies 2'OH groups in polyadenylation substrates critical for AAUAAA recognition and poly(A) addition. *Cell* **65**: 125–133.
- Bienroth, S., E. Wahle, C. Satler-Crazzolara, and W. Keller. 1991. Purification of the cleavage and polyadenylation factor involved in 3' processing of mRNA precursors. *J. Biol. Chem.* **266**: 19768–19776.
- Bienroth, S., W. Keller, and E. Wahle. 1993. Assembly of a processive mRNA polyadenylation complex. *EMBO J.* **12**: 585–594.
- Bilger, A., C.A. Fox, E. Wahle, and M. Wickens. 1994. Nuclear polyadenylation factors recognize cytoplasmic polyadenylation elements. *Genes & Dev.* **8**: 1106–1115.
- Boelens, W.C., E.J.R. Jansen, W.J. VanVenrooij, R. Striepecke, I.W. Mattaj, and S.I. Gunderson. 1993. The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* **72**: 881–892.
- Burd, C.G. and G. Dreyfuss. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**: 615–621.
- Butler, J.S. and T. Platt. 1988. RNA processing generates the mature 3' end of yeast CYC1 mRNA in vitro. *Science* **242**: 1270–1274.
- Chen, J. and C. Moore. 1992. Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. *Mol. Cell. Biol.* **12**: 3470–3481.
- Christofori, G. and W. Keller. 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.
- . 1989. Poly(A) polymerase purified from HeLa cell nuclear extract is required for both cleavage and polyadenylation of pre-mRNA in vitro. *Mol. Cell. Biol.* **9**: 193–203.
- Dingwall, C., and R.L. Laskey. 1991. Nuclear targeting sequences—A consensus? *Trends Biochem. Sci.* **16**: 478–481.
- Gentz, R., C.-H. Chen, and C.A. Rosen. 1989. Bioassay for transactivation using purified human immunodeficiency virus tat encoded protein: Transactivation requires mRNA synthesis. *Proc. Natl. Acad. Sci.* **86**: 821–824.
- Gilmartin, G.M. and J.R. Nevins. 1989. An ordered pathway of assembly of components required for polyadenylation site recognition and processing. *Genes & Dev.* **3**: 2180–2189.
- . 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., E.S. Fleming, J. Oetjen, and B.R. Graveley. 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.
- Gunderson, S.I., K. Boyer, G. Martin, W. Keller, W. Boelens, and I.W. Mattaj. 1994. The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* **76**: 531–541.
- Jenny, A., H.-P. Hauri, and W. Keller. 1994. Characterization of cleavage and polyadenylation specificity factor and cloning of its 100 kD subunit. *Mol. Cell. Biol.* **14**: 8183–8190.
- Keller, W., S. Bienroth, K.M. Lang, and G. Christofori. 1991. Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J.* **10**: 4241–4249.
- Kenan, D.J., C.C. Query, and J.D. Keene. 1991. RNA recognition: Towards identifying determinants of specificity. *Trends Biochem. Sci.* **16**: 214–220.
- Kennedy, T.E., M.A. Gavinowicz, A. Barzilai, E.R. Kandel, and J.D. Sweatt. 1988. Sequencing of proteins from two-dimensional gels by using *in situ* digestion and transfer of peptides to polyvinylidene difluoride membranes: Application to proteins associated with sensitization in *Aplysia*. *Proc. Natl. Acad. Sci.* **85**: 7008–7012.
- Kohtz, J.D., S.F. Jamison, C.L. Will, P. Zuo, R. Luhrmann, M.A. Garcia-Blanco, and J.L. Manley. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**: 119–124.
- Lutz, C. and J.C. Alwine. 1994. Direct interaction of the U1 snRNP-A protein with the upstream efficiency element of the SV40 late polyadenylation signal. *Genes & Dev.* **8**: 576–586.
- MacDonald, C.C., J. Wilusz, and T. Shenk. 1994. The 64kD 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.
- Manley, J.L. 1988. Polyadenylation of mRNA precursors. *Biochem. Biophys. Acta* **950**: 1–12.
- . 1995. A complex protein assembly catalyzes polyadenylation of mRNA precursors. *Curr. Opin. Genet. Dev.* **5**: 222–228.
- Matsudaira, P. 1987. Sequence from picomole quantities of protein electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**: 10035–10038.
- Miller, D.W., P. Safer, and L.K. Miller. 1986. In *Genetic engineering*, (ed. J.K. Setlow and A. Hollaender), vol. 8. pp. 277–298. Plenum Press, New York.
- Minvielle-Sebastia, L. B. Winsor, N. Bonneand, and F. Lacroute. 1991. Yeast *RNA14* and *RNA15* genes result in an abnormal mRNA decay rate; sequence analysis reveals an RNA binding domain in the RNA 15 protein. *Mol. Cell. Biol.* **11**: 3075–3087.
- Minvielle-Sebastia, L., P.J. Preker, and W. Keller. 1994. RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science* **266**: 1702–1705.
- Mitchelson, A., M. Simonelig, C. Williams, and K. O'Hare. 1993. Homology with *S. cerevisiae* RNA 14 suggests that phenotypic suppression in *Drosophila* by *suppressor of forked* occurs at the level of RNA stability. *Genes & Dev.* **7**: 241–249.
- Moore, C.L., J. Chen, and J. Whoriskey. 1988. Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. *EMBO J.* **7**: 3159–3169.
- Murthy, K.G. and J.L. Manley. 1992. Characterization of the multisubunit cleavage polyadenylation specificity factor from calf thymus. *J. Biol. Chem.* **267**: 14804–14811.
- Preker, P.J., J. Lingner, L. Minvielle-Sebastia, and W. Keller. 1995. The *FIP1* gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. *Cell* **81**: 379–389.
- Proudfoot, N. 1991. Poly(A) signals. *Cell* **64**: 671–674.
- Raabe, T., F. Bollum, and J.L. Manley. 1991. Primary structure and expression of bovine poly(A) polymerase. *Nature* **353**: 229–234.
- Raabe, T., K.G. Murthy, and J.L. Manley. 1994. Poly(A) polymerase contains multiple functional domains. *Mol. Cell. Biol.* **14**: 2946–2957.
- Ryner, L.C., Y. Takagaki, and J.L. Manley. 1989a. Multiple forms of poly(A) polymerase purified from HeLa cells function in specific mRNA 3' end formation. *Mol. Cell. Biol.* **9**: 4229–4238.
- . 1989b. Sequences downstream of AAUAAA signals affect pre-RNA cleavage and polyadenylation *in vitro* both directly and indirectly. *Mol. Cell. Biol.* **9**: 1759–1771.
- Sheets, M.D. and M. Wickens. 1989. Two phases in the addition of a poly(A) tail. *Genes & Dev.* **3**: 1401–1412.
- . 1994. A polyadenylation factor subunit is the human homologue of the *Drosophila* suppressor of forked protein. *Nature* **372**: 471–474.
- Takagaki, Y., L.C. Ryner, and J.L. Manley. 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *Cell* **52**: 731–742.
- . 1989. Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes & Dev.* **3**: 1711–1724.
- Takagaki, Y., J.L. Manley, C. MacDonald, J. Wilusz, and T. Shenk. 1990. A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes & Dev.* **4**: 2112–2120.
- Takagaki, Y., C.C. MacDonald, T. Shenk, and J.L. Manley. 1991. The 64kD polyadenylation factor contains an RNP-type RNA binding domain and unusual auxiliary motifs. *Proc. Natl. Acad. Sci.* **89**: 1403–1407.
- Takagaki, Y. and J.L. Manley. 1992. A human polyadenylation factor is a G-protein β subunit homologue. *J. Biol. Chem.* **267**: 23471–23475.
- Wahle, E. 1991. A novel poly(A) binding protein acts as a specificity factor in the second phase of mRNA polyadenylation. *Cell* **66**: 759–768.
- Wahle, E. and W. Keller. 1992. The biochemistry of 3' end cleavage and polyadenylation of mRNA precursors. *Annu. Rev. Biochem.* **61**: 419–440.
- Wahle, E., G. Martin, E. Schlitz, and W. Keller. 1991. Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. *EMBO J.* **10**: 4251–4257.
- Weiss, E.A., G.M. Gilmartin, and J.R. Nevins. 1991. Poly(A) site efficiency reflects the stability of complex formation involving the downstream termination elements. *EMBO J.* **10**: 215–219.
- Wickens, M. 1992. Forward, backward, how much, when: Mechanisms of poly(A) addition and removal and their role in early development. *Semin. Dev. Biol.* **3**: 399–412.
- Wilusz, J. and T. Shenk. 1988. A 64kD nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. *Cell* **52**: 221–228.
- Wilusz, J., T. Shenk, Y. Takagaki, and J.L. Manley. 1990. A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64kD protein to polydenylation substrates. *Mol. Cell. Biol.* **10**: 1244–1248.

The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation.

K G Murthy and J L Manley

Genes Dev. 1995, **9**:

Access the most recent version at doi:[10.1101/gad.9.21.2672](https://doi.org/10.1101/gad.9.21.2672)

References

This article cites 51 articles, 27 of which can be accessed free at:
<http://genesdev.cshlp.org/content/9/21/2672.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).