Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1403–1407, February 1992
Biochemistry

The human 64-kDa polyadenylylation factor contains a ribonucleoprotein-type RNA binding domain and unusual auxiliary motifs

YOSHIO TAKAGAKI*, CLINTON C. MACDONALD†, THOMAS SHENK†, AND JAMES L. MANLEY*

*Department of Biological Sciences, Columbia University, New York, NY 10027; and †Howard Hughes Medical Institute, Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Communicated by Mary Edmonds, November 4, 1991 (received for review August 29, 1991)

ABSTRACT **Cleavage stimulation factor is one of the multiple factors required for 3′-end cleavage of mammalian pre-mRNAs. We have shown previously that this factor is composed of three subunits with estimated molecular masses of 77, 64, and 50 kDa and that the 64-kDa subunit can be UV-crosslinked to RNA in a polyadenylylation signal (AAUAAA)-dependent manner. We have now isolated cDNAs encoding the 64-kDa subunit of human cleavage stimulation factor. The 64-kDa subunit contains a ribonucleoprotein-type RNA binding domain in the N-terminal region and a repeat structure in the C-terminal region in which a pentapeptide sequence (consensus MEARA/G) is repeated 12 times and the formation of a long α-helix stabilized by salt bridges is predicted. An ≈270-amino acid segment surrounding this repeat structure is highly enriched in proline and glycine residues (≈20% for each). When cloned 64-kDa subunit was expressed in *Escherichia coli*, an N-terminal fragment containing the RNA binding domain bound to RNAs in a polyadenylylation-signal-independent manner, suggesting that the RNA binding domain is directly involved in the binding of the 64-kDa subunit to pre-mRNAs.**

Nearly all mammalian mRNAs are polyadenylylated at their 3′ ends. Polyadenylylation of an RNA polymerase II transcript occurs in a two-step reaction (refs. 1–4; for reviews, see refs. 5 and 6). A pre-mRNA is first endonucleolytically cleaved at its polyadenylylation site, which is located 10–30 nucleotides (nt) downstream of the polyadenylylation signal sequence, AAUAAA (refs. 7–9; for review, see ref. 10), and a poly(A) stretch of 200–300 nt is then added to the 3′ end of the upstream cleavage product. Although these reactions appear tightly coupled *in vivo*, they can be uncoupled and studied separately *in vitro*. Biochemical fractionation of HeLa cell nuclear extracts has revealed that multiple factors are required for both cleavage and polyadenylylation reactions (11–17). It has been shown that four factors are necessary for cleavage of a simian virus 40 (SV40) late pre-mRNA (13). Only one of these, cleavage-polyadenylylation specificity factor [(CPSF) previously designated cleavage-polyadenylylation factor (12), specificity factor (13), or polyadenylylation factor 2 (14)], is also required for the polyadenylylation reaction. Cleavage factors I and II and cleavage stimulation factor (CstF) are necessary only for cleavage (13). For cleavage of several other pre-mRNAs, poly(A) polymerase, which with CPSF also functions to add poly(A) stretches to the 3′ ends of the upstream cleavage products, is also required (11–17).

CstF has been purified to homogeneity from HeLa cell nuclear extracts (18, 19). CstF is composed of three subunits with estimated molecular masses of 77, 64, and 50 kDa. By immunoprecipitation with a monoclonal antibody (mAb) against the 64-kDa subunit, this polypeptide was shown (18) to be identical to a protein of 64–68 kDa that had been detected (20, 21) in crude nuclear extracts by UV-crosslinking to pre-mRNAs in an AAUAAA sequence-dependent manner. Since both CPSF and CstF are required for this specific UV-crosslinking (19, 22) and for the formation of a stable complex on pre-mRNAs (14, 23), it has been suggested that CstF interacts with both the pre-mRNA and CPSF to stabilize the interaction between the pre-mRNA and CPSF. To understand the molecular mechanisms of CstF function, we cloned cDNAs encoding the 64-kDa subunit and determined its primary structure.‡ We also expressed the cloned cDNA in *Escherichia coli* and examined the RNA binding of the purified protein.

MATERIALS AND METHODS

Screening of cDNA Libraries. To obtain cDNA clones for the 64-kDa subunit, 6×10^5 plaques from a HeLa cell cDNA expression library in λgt11 were screened with a mixture of three mAbs as described (24). Four positive clones were plaque-purified, and cDNA inserts were subcloned in M13mp18 (25). The cDNA derived from λ64-1, the longest clone, was then used to isolate full-length cDNAs from another HeLa-cell cDNA library in λZAP II (Stratagene). Positive clones were plaque-purified and cDNA inserts were excised *in vivo* and sequenced by the chain-termination method (26, 27).

RNA and Protein Analysis. Hybrid selection of the 64-kDa-subunit-specific mRNA was performed as described (28), using 100 μg of HeLa-cell poly(A)$^+$ RNA. For *in vitro* transcription (29), pBSSK plasmid containing the longest cDNA insert (pZ64-18) was digested with *Eag* I. Hybrid-selected or *in vitro*-transcribed capped mRNA were translated *in vitro* using rabbit reticulocyte lysate (Promega). *In vitro* translation products were subjected to immunoprecipitation with anti-64-kDa-subunit mAb (18) or with anti-polyoma large tumor antigen mAb as described (18). For Northern blot analysis, 10 μg of HeLa-cell poly(A)$^+$ RNA was treated with glyoxal and fractionated on a 1% agarose gel in 10 mM sodium phosphate (pH 7.0). Hybridization was performed essentially as described (25).

Partial Protease Analysis of 64-kDa Proteins. HeLa cells were labeled with Tran^{35}S-Label (ICN) for 1 hr, harvested, and lysed, and the 64-kDa protein was purified by immunoprecipitation (18). The RNA transcript from pZ64-18 was translated *in vitro* as described above and the 64-kDa protein was purified by SDS/PAGE followed by immunoprecipitation. *In vivo*- and *in vitro*-synthesized 64-kDa proteins were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: CPSF, cleavage-polyadenylylation specificity factor; CstF, cleavage stimulation factor; mAb, monoclonal antibody; RBD, RNA binding domain; SV40, simian virus 40; nt, nucleotide(s).

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M85085).

treated with *Staphylococcus aureus* V8 protease at 37°C for 30 min (30) and fractionated on an SDS/15% polyacrylamide gel (31).

Expression of the Cloned 64-kDa Subunit in *E. coli*. To construct an *E. coli* expression plasmid, a *Sal* I site was first created immediately upstream of the translation initiation codon ATG (residues 1–3) by the PCR using pZ64-18 DNA as a template, and the *Sal* I–*Hin*dIII fragment was cloned into pDS56-6xHis vector (32). *E. coli* RR-1 cells were transformed with this plasmid and protein expression was induced with isopropyl β-D-thiogalactopyranoside. The fusion protein (rHis64Δ247), which contains six consecutive histidine residues encoded by the vector sequence followed by a 247-residue N-terminal fragment of the 64-kDa subunit, was purified as described (32).

UV-Crosslinking. rHis64Δ247 protein (5 μg) was incubated with ^{32}P-labeled SV40 late pre-mRNAs (1×10^4 cpm) in buffer D (33) at 30°C for 30 min. The mixtures were UV-irradiated and then treated with RNase A (20). Proteins were precipitated with 10% trichloroacetic acid and fractionated on an SDS/10% polyacrylamide gel.

RESULTS

Cloning of cDNAs for the 64-kDa Subunit of CstF. We have purified CstF to near homogeneity and prepared mAbs against the 64-kDa subunit of CstF (18). Using a mixture of these mAbs, a HeLa-cell cDNA expression library in λgt11 was immunoscreened. The insert from λ64-1, the longest positive clone, was used to hybrid-select complementary RNA from HeLa-cell poly(A)$^+$ RNA (28). When this mRNA was translated *in vitro*, only a single major protein with estimated molecular mass of ≈64 kDa was detected, which was immunoprecipitated with a specific anti-64-kDa subunit mAb but not with a nonspecific anti-polyoma large tumor antigen mAb (Fig. 1*A*).

Since the length of the cDNA insert of the clone λ64-1 (≈1.2 kilobase pairs) was much shorter than that of the specific mRNA estimated by Northern blot analysis (≈2.2 kilobases, Fig. 1*B*), we screened a second HeLa-cell cDNA library with the cDNA insert of λ64-1 and obtained 22 positive clones. One of these (pZ64-18), containing the longest cDNA insert, was used as a template for *in vitro* transcription. When the RNA transcript was translated *in vitro*, again a single major product (≈64 kDa) was detected and was immunoprecipitated only with a specific anti-64-kDa-subunit mAb (Fig. 2*A*). To provide additional evidence that the cDNA clone pZ64-18 encodes the 64-kDa subunit, the V8 protease partial digestion patterns of the 64-kDa subunit protein purified from [^{35}S]methionine-labeled HeLa cells and the [^{35}S]methionine-labeled *in vitro* translation product were compared (Fig. 2*B*). The *in vivo*- and *in vitro*-synthesized proteins showed exactly the same polypeptide profiles. These results indicate that pZ64-18 encodes the 64-kDa subunit and that it contains the entire protein coding region of the 64-kDa-subunit mRNA.

Structure of the cDNA Encoding the 64-kDa Subunit. The complete nucleotide sequence of pZ64-18 was then determined (Fig. 3). Since this cDNA did not appear to extend entirely to the 3′ end of the mRNA, part of another clone (pZ64-17) was also sequenced. Altogether the 64-kDa-subunit cDNA is composed of 1978 base pairs [excluding poly(A)] and only a single long open reading frame starting with the putative translation initiation codon ATG (nt 1–3) was found. The 5′ untranslated region predicted from the cDNA is quite short (22 nt). Although the actual untranslated region may be somewhat longer, the correspondence between the sizes of the cDNA and mRNA suggest that the difference, if any, is small. The nucleotide sequence surrounding the ATG codon (underlined) is an excellent match to the consensus sequence A/GCCATGG (34). The translation termination codon (TGA) and the polyadenylylation signal (AATAAA) are found at nt 1732–1734 and at nt 1925–1930, respectively. The protein is composed of 577 amino acids and the molecular weight was calculated to be 60,920.

Structural Features of the 64-kDa Protein. The 64-kDa subunit contains three characteristic structural features. First, a search for homologous proteins in the GenBank data base (February 1991) using the FASTA program (35) revealed that the N-terminal region of the 64-kDa subunit (residues 17–96) shares homology with the ribonucleoprotein-consensus-type RBD (36) or RNA recognition motif (37).

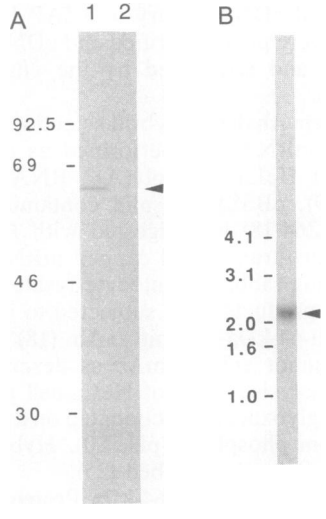

FIG. 1. Characterization of the 64-kDa-subunit mRNA. (*A*) mRNA hybrid-selected with the cDNA insert of λ64-1 was translated *in vitro* and the reaction product was immunoprecipitated with anti-64-kDa-subunit mAb (lane 1) or with anti-polyoma large tumor antigen mAb (lane 2). Immunoprecipitated protein was fractionated on an SDS/10% polyacrylamide gel. Positions of the molecular mass markers are indicated in kDa on the left. Arrowhead indicates the position of the 64-kDa protein. (*B*) HeLa-cell poly(A)$^+$ RNA fractionated and transferred to nitrocellulose was probed with the cDNA insert of λ64-1. Positions of the DNA size markers are indicated in kilobase(s) on the left. Arrowhead indicates the position of the 64-kDa-subunit mRNA.

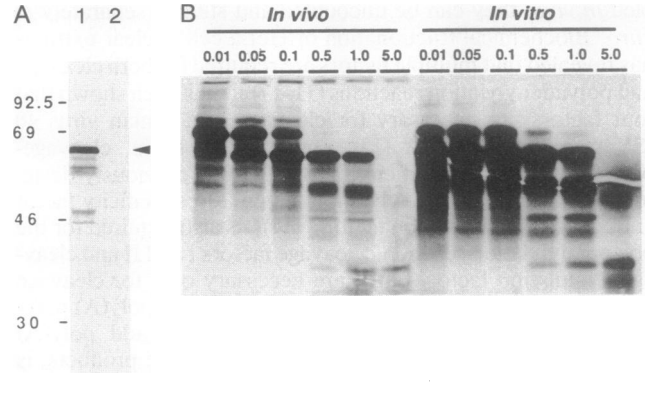

FIG. 2. Characterization of a protein encoded by a 64-kDa-subunit cDNA (pZ64-18). (*A*) *In vitro* translation of an RNA transcript from pZ64-18. The translation product was immunoprecipitated and analyzed as described in Fig. 1*A*. (*B*) Partial protease analysis of *in vivo*- and *in vitro*-synthesized 64-kDa proteins. The 64-kDa proteins synthesized in HeLa cells or *in vitro* were purified. Both proteins were subjected to digestion with the indicated amounts (in μg) of *S. aureus* V8 protease and fractionated on an SDS/15% polyacrylamide gel.

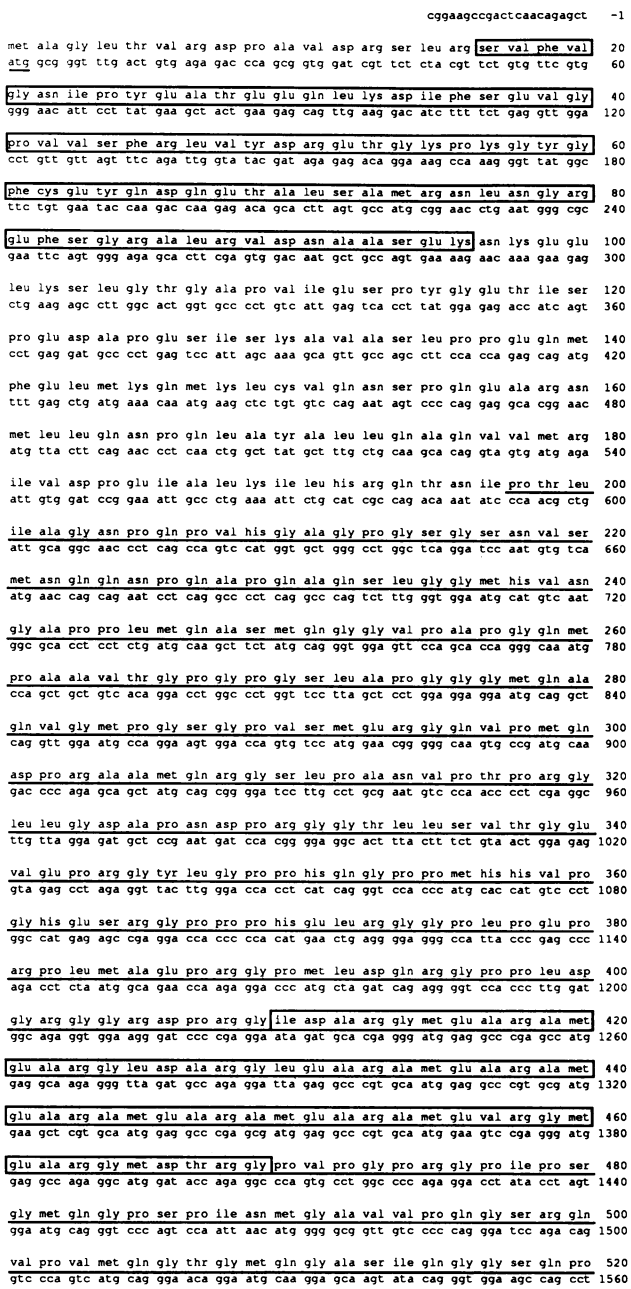

FIG. 3. Sequences of the 64-kDa-subunit cDNA and the predicted protein. Nucleotides (lower lines) and amino acids (upper lines) are numbered on the right, starting with the translation initiation site. The translation initiation codon (ATG), the translation termination codon (TGA), and the polyadenylylation signal (AATAAA) are underlined. The RNA binding domain (RBD) and the repeat structure are boxed and the proline/glycine-rich regions are underlined.

This ≈80-residue domain (Fig. 4) has been found in a variety of RNA binding proteins, and, where studied, appears directly responsible for the interaction of the protein with RNA.

A second feature is a repeat structure found in the C-terminal region of the 64-kDa subunit (residues 410–469). Each repeat is composed of 5 amino acids (consensus MEARA/G) and is repeated 12 times (Fig. 5*A*). No protein containing this kind of pentapeptide repeat was found in GenBank. Prediction of secondary structure according to Chou and Fasman (38) strongly suggests that this repeat structure forms an exceptionally long α-helix, consisting of 16 turns. When this α-helical structure was depicted by flattened helical display (Fig. 5*B*), it became evident that all basic amino acid residues (arginine) can electrostatically interact with closely located acidic residues (glutamic or aspartic acid) in the next turn to form as many as 11 salt bridges (39, 40). These salt bridges very likely provide considerable stabilization for the α-helical structure. In addition, helical wheel analysis (Fig. 5*C*) shows that salt bridges can be formed on all faces of the helix.

An additional feature stems from the fact that the contents of proline (12.0%) and glycine (12.5%) residues in the 64-kDa subunit are very high. In particular, these two residues are highly enriched in regions surrounding the repeat. A long region preceding the repeat structure (residues 198–409) is 19.3% proline and 18.9% glycine; a shorter region of 57 residues immediately following the repeat (residues 470–526) also displays a high content of proline (19.3%) and glycine (24.6%). Secondary structure prediction suggests the existence of as many as 10 β-turns in these regions (data not shown). Although some proteins in connective tissues (e.g., collagens) also show high contents of proline and glycine, their characteristic repeating motif $(\text{Gly-Xaa-Yaa})_n$, where Yaa is often proline (41), is not found in the 64-kDa subunit.

The 64-kDa subunit of purified CstF displays apparent size heterogeneity during SDS/PAGE (18), and recent studies have shown that this is due at least in part to phosphorylation (C.C.M. and T.S., unpublished data). In this regard, we note the presence of a possible protein kinase C consensus site (residues 80–85), which is in the C-terminal region of the RBD, and two possible cGMP-dependent protein kinase sites at residues 364–365 and 498–499 (42). Additionally, despite the fact that the 64-kDa subunit is localized exclusively in the nucleus (18), it lacks sequences resembling a nuclear localization signal (43). It may be that heterotrimer formation occurs in the cytoplasm, and one of the other subunits of CstF contains the requisite nuclear localization signal.

The 64-kDa Protein Synthesized in *E. coli* Binds RNA. To study the RNA binding function of the 64-kDa subunit by itself, we expressed an N-terminal fragment of the protein in *E. coli* using the pDS56-6xHis vector (32). The fusion protein was purified and tested for its ability to be UV-crosslinked to SV40 late pre-mRNAs containing a wild-type (AAUAAA) or a point-mutant (AAGAAA) polyadenylylation signal. The *E. coli*-expressed protein was UV-crosslinked to both pre-mRNAs with equal efficiencies (Fig. 6). These results agree with those obtained with extensively purified CstF (22) and support the view that the 64-kDa subunit does not recognize the AAUAAA polyadenylylation signal.

DISCUSSION

In this study, we have cloned cDNAs encoding the 64-kDa subunit of CstF and predicted elements of the protein's structure. The 64-kDa subunit contains an RBD in the N-terminal region. The highly conserved ribonucleoprotein 1 and 2 motifs (36, 37) are both present, as are other well-conserved residues and possible structural motifs (44). This domain is very likely involved in the interaction of the 64-kDa subunit with pre-mRNAs detected by UV-crosslinking experiments, a view supported by the fact that an N-terminal fragment of the 64-kDa protein expressed in *E. coli* could be efficiently crosslinked to RNA.

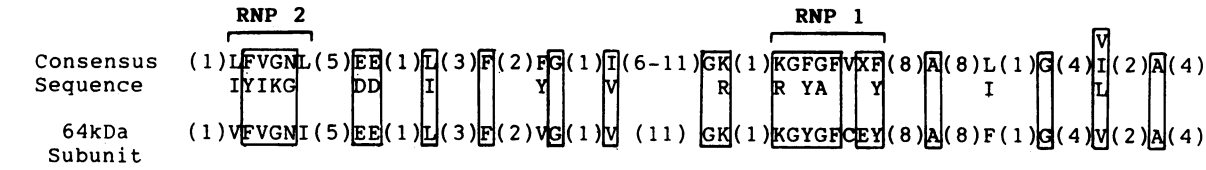

FIG. 4. Comparison of the RBD in the 64-kDa subunit and the RBD consensus. Highly conserved residues, deduced from compilations of RBDs (36, 37), and the spacings between them (in parentheses) are indicated on the top. Corresponding amino acid residues and the spacings in the 64-kDa-subunit RBD are shown on the bottom. Positions at which identical residues are found between the consensus and the 64-kDa subunit are boxed. Ribonucleoprotein 1 (RNP1; octamer) and 2 (RNP2) consensus sequences are indicated.

Both the N-terminal fragment and full-length bacterially expressed 64-kDa protein (unpublished data) can be crosslinked with equal efficiencies to RNAs containing or lacking an intact AAUAAA. These findings are consistent with previous results obtained with extensively purified CstF (19, 22), which, coupled with functional studies (13, 14), led to the suggestion that CstF does not directly recognize AAUAAA (e.g., ref. 18). What specific sequences might then be recognized by the 64-kDa subunit? The only sequence element other than AAUAAA implicated in polyadenylylation is the so-called downstream element, which is a U- or G/U-rich sequence found 5–50 nt 3′ to the cleavage site in many pre-mRNAs (for review, see ref. 6). Indeed, recent studies by Nevins and colleagues (14, 23) have suggested that CstF (called CF-1 by those authors) interacts specifically with these downstream sequences. However, several other studies are not entirely consistent with this view. First, the original detection of AAUAAA-dependent UV-crosslinking of the 64-kDa subunit did not show any dependence on downstream sequences (20, 21), and our preliminary experiments with purified CstF or the recombinant 64-kDa subunit are consistent with these earlier findings. Furthermore, we have also shown that CstF is fully functional in bringing about efficient cleavage on substrates containing mutations that destroy the downstream element (K. Murthy, Y.T., and J.L.M., unpublished data). Therefore, we suggest that the 64-kDa protein may have only limited sequence specificity in its interaction with RNA and that its function in polyadenylylation may involve specific protein–protein interactions.

The repeat structure in the C-terminal region is likely to form an extended α-helix. Several different types of α-helical structures have been suggested to make specific protein–protein interactions in a number of transcription factors, for example, in leucine-zipper (45–47) and helix–loop–helix (46–49) proteins. In all of these cases, hydrophobic amino acids are localized on one side of the helix, and it has been suggested that these residues are directly involved in protein–protein interactions. On the other hand, other transcription factors can apparently form amphipathic α-helices in which charged residues are located on the same face of the helix (e.g., refs. 50 and 51). In contrast to all of these examples, the putative α-helix of the 64-kDa subunit protein is considerably longer (≈60 residues), and charged residues (both acidic and basic) are distributed on all faces of the helix (Fig. 5*C*). This raises the possibility that the α-helix of the 64-kDa subunit is completely or largely solvent-exposed, an extremely unusual situation observed previously only in certain calcium-binding contractile proteins (39). In addition, the α-helix is surrounded by proline/glycine-rich regions that can potentially form multiple β-turns, which may provide a flexible framework allowing extensive movement of the α-helix. We speculate that the repeat structure is involved in a protein–protein interaction mediated by electrostatic forces.

As mentioned above, there is now considerable evidence suggesting a functional interaction between CstF and CPSF (14, 19, 22, 23). When CstF and CPSF are mixed, efficient UV-crosslinking of the 64-kDa subunit to RNA can be detected only with substrates containing AAUAAA. In contrast, purified CstF or the *E. coli*-expressed N-terminal fragment of the 64-kDa subunit shows no specificity for AAUAAA-containing RNAs. Thus it appears that protein–protein interactions between CPSF and CstF convert the latter to an AAUAAA-dependent RNA binding protein, perhaps through cooperative interactions of the two factors with the pre-mRNA. This is analogous to the interaction between CPSF and poly(A) polymerase, which converts the enzyme from a sequence-independent polymerase to an AAUAAA-dependent one. We propose that the α-helical

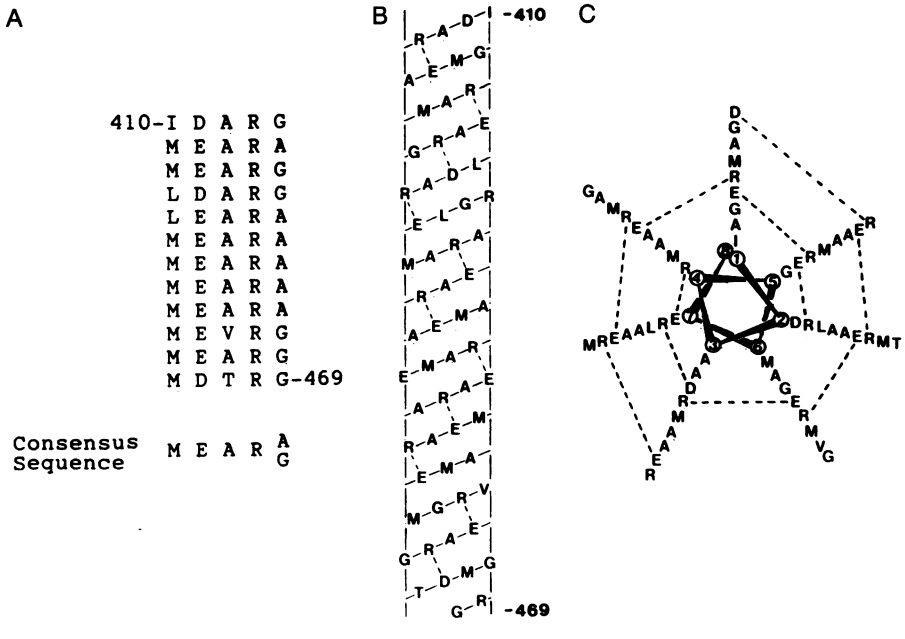

FIG. 5. Structure of the pentapeptide repeating motif in the 64-kDa subunit. (*A*) Amino acid sequences of the pentapeptide repeating units are shown. A consensus sequence is shown at the bottom. The first and the last residues of the repeat structure are indicated by number. (*B*) Flattened helical display of the repeat structure. The helical region containing the repeat structure was flattened into two dimensions by splitting the helix lengthwise. Note that Arg-428, Glu-446, and Gly-464, which are bisected by the split along the helix, are displayed in duplicate. Salt bridges are indicated by dotted lines. (*C*) Helical wheel analysis of the repeat structure. The amino acid sequence of the repeat structure is displayed down the axis of a schematic α-helix. Ile-410 is placed at position 1 of the helix, followed by Asp-411 at position 2, and so on. Gly-469, the most C-terminal residue, is placed at position 4.

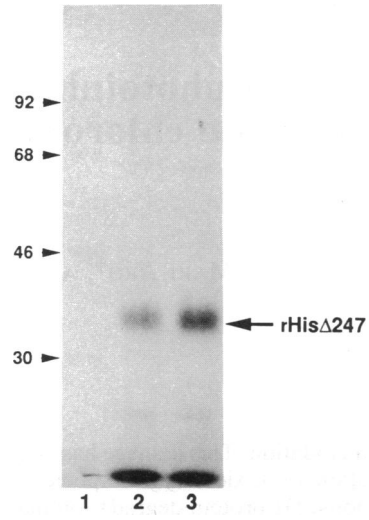

FIG. 6. UV-crosslinking of the bacterially expressed 64-kDa subunit to SV40 late pre-mRNAs. The *E. coli*-expressed N-terminal fragment of the 64-kDa subunit (rHis64Δ247) was incubated with ^{32}P-labeled SV40 late pre-mRNAs containing a wild-type (AAUAAA, lane 2) or a mutated (AAGAAA, lane 3) polyadenylylation signal. The reaction mixtures were UV-irradiated, treated with RNase A, and fractionated on an SDS/10% polyacrylamide gel. Positions of the molecular mass markers (lane 1) are indicated on the left in kDa.

structure in the 64-kDa subunit directly interacts with CPSF. It may be that this interaction alters the structure of the 64-kDa subunit, increasing the accessibility and/or affinity of the RBD for the pre-mRNA. An intriguing possibility is that some of the intramolecular salt bridges in the repeat structure of the 64-kDa subunit are converted to intermolecular ones when CstF interacts with CPSF, leading to a stable interaction between CstF and CPSF.

We are grateful to L. Zhong for expert technical assistance. We thank T. Kadesch and C. Prives for supplying a HeLa cell cDNA library and mAb, respectively. We are also grateful to A. Lupas for pointing out the potential for salt bridge formation and to W. Weast for preparing the manuscript. This work was supported by National Institutes of Health Grants GM-28983 to J.L.M. and CA-38965 to T.S. C.C.M. is an American Cancer Society postdoctoral fellow. T.S. is an American Cancer Society professor.

1. Nevins, J. R. & Darnell, J. E., Jr. (1978) *Cell* **15,** 1477–1493.
2. Manley, J. L., Sharp, P. A. & Gefter, M. L. (1982) *J. Mol. Biol.* **159,** 581–599.
3. Moore, C. L., Skolnik-David, H. & Sharp, P. A. (1986) *EMBO J.* **5,** 1929–1938.
4. Sheets, M. D., Stephenson, P. & Wickens, M. (1987) *Mol. Cell. Biol.* **7,** 1518–1529.
5. Humphrey, T. & Proudfoot, N. J. (1988) *Trends Genet.* **4,** 243–245.
6. Manley, J. L. (1988) *Biochim. Biophys. Acta* **950,** 1–12.
7. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
8. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24,** 251–260.
9. Higgs, D. R., Goodbourn, S. E. Y., Lamb, J., Clegg, J. B., Weatherall, D. J. & Proudfoot, N. J. (1983) *Nature (London)* **306,** 398–400.
10. Proudfoot, N. (1991) *Cell* **64,** 671–674.
11. Takagaki, Y., Ryner, L. C. & Manley, J. L. (1988) *Cell* **52,** 731–742.
12. Christofori, G. & Keller, W. (1988) *Cell* **54,** 875–889.
13. Takagaki, Y., Ryner, L. C. & Manley, J. L. (1989) *Genes Dev.* **3,** 1711–1724.
14. Gilmartin, G. M. & Nevins, J. R. (1989) *Genes Dev.* **3,** 2180–2189.
15. Ryner, L. C., Takagaki, Y. & Manley, J. L. (1989) *Mol. Cell. Biol.* **9,** 4229–4238.
16. Terns, M. P. & Jacob, S. T. (1989) *Mol. Cell. Biol.* **9,** 1435–1444.
17. Bardwell, V. J., Zarkower, D., Edmonds, M. & Wickens, M. (1990) *Mol. Cell. Biol.* **10,** 846–849.
18. Takagaki, Y., Manley, J. L., MacDonald, C. C., Wilusz, J. & Shenk, T. (1990) *Genes Dev.* **4,** 2112–2120.
19. Gilmartin, G. M. & Nevins, J. R. (1991) *Mol. Cell. Biol.* **11,** 2432–2438.
20. Wilusz, J. & Shenk, T. (1988) *Cell* **52,** 221–228.
21. Moore, C. L., Chen, J. & Whoriskey, J. (1988) *EMBO J.* **7,** 3159–3169.
22. Wilusz, J., Shenk, T., Takagaki, Y. & Manley, J. L. (1990) *Mol. Cell. Biol.* **10,** 1244–1248.
23. Weiss, E. A., Gilmartin, G. M. & Nevins, J. R. (1991) *EMBO J.* **10,** 215–219.
24. Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
25. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
26. Henikoff, S. (1987) *Methods Enzymol.* **155,** 156–165.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Jagus, R. (1987) *Methods Enzymol.* **152,** 567–572.
29. Konarska, M. M., Padgett, R. A. & Sharp, P. A. (1984) *Cell* **38,** 731–736.
30. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. & Laemmli, E. K. (1977) *J. Biol. Chem.* **252,** 1102–1106.
31. Laemmli, U. K. (1970) *Nature (London)* **227,** 680–685.
32. Gentz, R., Chen, C.-H. & Rosen, C. A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 821–824.
33. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489.
34. Kozak, M. (1989) *J. Cell Biol.* **108,** 229–241.
35. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
36. Bandziulis, R. J., Swanson, M. S. & Dreyfuss, G. (1989) *Genes Dev.* **3,** 431–437.
37. Query, C. C., Bentley, R. C. & Keene, J. D. (1989) *Cell* **57,** 89–101.
38. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148.
39. Sundaralingam, M., Drendel, W. & Greaser, M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 7944–7947.
40. Marqusee, S. & Baldwin, R. L. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 8898–8902.
41. Vuorio, E. & de Crombrugghe, B. (1990) *Annu. Rev. Biochem.* **59,** 837–872.
42. Kemp, B. E. & Pearson, R. B. (1990) *Trends Biochem. Sci.* **15,** 342–346.
43. Silver, P. A. (1991) *Cell* **64,** 489–497.
44. Kenan, D. J., Query, C. C. & Keene, J. D. (1991) *Trends Biochem. Sci.* **16,** 214–221.
45. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240,** 1759–1764.
46. Prendergast, G. C. & Ziff, E. B. (1989) *Nature (London)* **341,** 392.
47. Blackwood, E. M. & Eisenman, R. N. (1991) *Science* **251,** 1211–1217.
48. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56,** 777–783.
49. Williams, T. & Tjian, R. (1991) *Science* **251,** 1067–1071.
50. Horikoshi, M., Wang, C. K., Fujii, H., Cromlish, J. A., Weil, P. A. & Roeder, R. G. (1989) *Nature (London)* **341,** 299–303.
51. Ptashne, M. (1988) *Nature (London)* **335,** 683–689.