

## Distinct Sequence Motifs within the 68-kDa Subunit of Cleavage Factor I<sub>m</sub> Mediate RNA Binding, Protein-Protein Interactions, and Subcellular Localization\*

Received for publication, April 8, 2004, and in revised form, May 28, 2004  
Published, JBC Papers in Press, May 28, 2004, DOI 10.1074/jbc.M403927200

Sabine Dettwiler<sup>‡</sup>, Chiara Aringhieri<sup>§</sup>¶, Stefano Cardinale<sup>§</sup>¶, Walter Keller<sup>‡</sup>,  
and Silvia M. L. Barabino<sup>§</sup>¶

From the <sup>‡</sup>Department of Cell Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland and <sup>§</sup>Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, 2, I-20126 Milano, Italy

**Cleavage factor I<sub>m</sub> (CF I<sub>m</sub>) is required for the first step in pre-mRNA 3'-end processing and can be reconstituted *in vitro* from its heterologously expressed 25- and 68-kDa subunits. The binding of CF I<sub>m</sub> to the pre-mRNA is one of the earliest steps in the assembly of the cleavage and polyadenylation machinery and facilitates the recruitment of other processing factors. We identified regions in the subunits of CF I<sub>m</sub> involved in RNA binding, protein-protein interactions, and subcellular localization. CF I<sub>m68</sub> has a modular domain organization consisting of an N-terminal RNA recognition motif and a C-terminal alternating charge domain. However, the RNA recognition motif of CF I<sub>m68</sub> on its own is not sufficient to bind RNA but is necessary for association with the 25-kDa subunit. RNA binding appears to require a CF I<sub>m68</sub>/25 heterodimer. Whereas multiple protein interactions with other 3'-end-processing factors are detected with CF I<sub>m25</sub>, CF I<sub>m68</sub> interacts with SRp20, 9G8, and hTra2β, members of the SR family of splicing factors, via its C-terminal alternating charge domain. This domain is also required for targeting CF I<sub>m68</sub> to the nucleus. However, CF I<sub>m68</sub> does not concentrate in splicing speckles but in foci that partially colocalize with paraspeckles, a subnuclear component in which other proteins involved in transcriptional control and RNA processing have been found.**

Eukaryotic mRNA precursors (pre-mRNAs)<sup>1</sup> are synthesized and processed in the nucleus prior to their export to the cyto-

plasm where they serve as templates for protein synthesis. Transcription is coupled spatially and temporally to the capping of the pre-mRNA at the 5'-end, splicing, and 3'-end formation. The mature 3'-ends of most eukaryotic mRNAs are generated by endonucleolytic cleavage of the primary transcript followed by the addition of a poly(A) tail to the upstream cleavage product (for reviews see Refs. 1 and 2). In mammals, these reactions are catalyzed by a large multicomponent complex that is assembled in a cooperative manner on specific *cis*-acting sequence elements in the pre-mRNA. The cleavage and polyadenylation specificity factor (CPSF) (3) recognizes the highly conserved hexanucleotide AAUAAA, whereas the cleavage stimulation factor (CstF) (4) binds a more degenerate GU- or U-rich element downstream of the poly(A) site. It has been suggested that *in vivo* CPSF and CstF may become associated with each other prior to pre-mRNA binding, recognizing the two elements in a concerted manner (5). In addition, the cleavage reaction requires mammalian cleavage factor I (CF I<sub>m</sub>), cleavage factor II<sub>m</sub> (CF II<sub>m</sub>), and poly(A) polymerase (PAP). After the first step of 3'-end processing, CPSF remains bound to the upstream cleavage fragment and tethers PAP to the 3'-end of the pre-mRNA (6). In the presence of the nuclear poly(A)-binding protein 1 (PABPN1), PAP elongates the poly(A) tail in a processive manner (6). These factors are both necessary and sufficient to reconstitute cleavage and polyadenylation *in vitro*. However, the other proteins involved in either transcription, such as the C-terminal domain of RNA polymerase II, or capping (nuclear cap-binding complex) and splicing (U2AF65) have been shown to greatly enhance the efficiency of the first step of the reaction (7–9).

Three major polypeptides of 25, 59, and 68 kDa and one minor polypeptide of 72 kDa copurify with CF I<sub>m</sub> activity from HeLa cell nuclear extract (10). Reconstitution of CF I<sub>m</sub> activity with recombinant proteins suggests that CF I<sub>m</sub> is a heterodimer consisting of the 25-kDa subunit and one of the larger polypeptides (11). All of the three larger proteins appear to be highly related in their amino acid sequence. Moreover, all of the CF I<sub>m</sub> subunits are only present in metazoan organisms. The primary sequence of the 25-kDa polypeptide contains a NUDIX motif (12). The amino acid composition of the 68-kDa protein has a domain organization that is reminiscent of spliceosomal SR proteins. Members of the SR family of splicing factors contain one or more N-terminal RNA recognition motifs (RRMs) that function in sequence-specific RNA binding and a C-terminal domain rich in alternating arginine and serine residues, referred to as RS domain, which is required for protein-protein interactions with other RS domains (13). In the 68-kDa protein, the RRM and the RS-like domain are separated by a

\* This work was supported in part by the Telethon Onlus Grant 0010Y02 and a grant from Associazione Italiana per la Ricerca sul Cancro (to S. M. L. B.) and by the University of Basel, the Swiss National Science Fund, and the Louis-Jeantet Foundation for Medicine. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Both authors contributed equally to the work.

¶ To whom correspondence should be addressed: Dept. of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, 2, I-20126 Milano, Italy. Tel.: 39-02-6448-3352; Fax: 39-02-6448-3569; E-mail: silvia.barabino@unimib.it.

<sup>1</sup> The abbreviations used are: pre-mRNA, mRNA precursor; CPSF, cleavage and polyadenylation specificity factor; CstF, cleavage stimulation factor; CF I<sub>m</sub>, cleavage factor I; CF II<sub>m</sub>, cleavage factor II<sub>m</sub>; PAP, poly(A) polymerase; PABPN1, nuclear poly(A)-binding protein; RRM, RNA recognition motif; GST, glutathione S-transferase; GFP, green fluorescent protein; PSP1, paraspeckle protein 1; NRS, nuclear retention signal; PSF, polypyrimidine tract-binding protein-associated splicing factor; PBS, phosphate-buffered saline; aa, amino acid; RNP, ribonuclear particle.

region with high proline content (47%). SR proteins bound to specific RNA sequence elements are thought to recruit key splicing factors, thus enhancing the recognition of splice sites and controlling splice site selection in a concentration-dependent manner (for review see Ref. 14). Previous experiments have shown that preincubation of the RNA substrate with CF  $I_m$  reduces the lag-phase of the cleavage reaction (11). This observation suggests that the binding of this factor to the pre-mRNA may be an early step in the assembly of the 3'-end-processing complex, such that CF  $I_m$  could have a role similar to that of SR proteins in spliceosome assembly. Recent SELEX experiments identified a related set of sequences within the 3'-untranslated region of the pre-mRNA of its 68-kDa subunit to which CF  $I_m$  preferentially binds (15).

In this report, we describe the analysis of the functional domains of the 25- and 68-kDa subunits of CF  $I_m$ . To this end, we generated several deletion and point mutants of the two subunits. We expressed wild type and mutant proteins in heterologous systems and analyzed the purified proteins for protein-protein interactions and for RNA binding. We found that the 25-kDa subunit of CF  $I_m$  interacts not only with PAP (16) but also with PABPN1. Although the amino acid composition of this subunit does not display a known RNA recognition motif, CF  $I_m$ 25 binds to RNA. Surprisingly, despite having an RRM, CF  $I_m$ 68 does not bind very strongly to RNA but requires its RRM for interaction with the 25-kDa protein instead. A yeast two-hybrid screen with the RS-like domain of CF  $I_m$ 68 identified members of the SR family of splicing factors, suggesting that CF  $I_m$  may contribute to the coordination of splicing and 3'-end formation. Furthermore, we investigated the subcellular localization of CF  $I_m$ . We found that CF  $I_m$  subunits are localized to the nucleus and accumulate within a few bright foci that partially overlap with paraspeckles (17). Finally, we report that the RS-like domain is sufficient for targeting the protein to the nucleus.

#### EXPERIMENTAL PROCEDURES

**Oligonucleotides**—Oligonucleotides were synthesized either by Microsynth or MWG. The sequences were as follows: GST-25-(1–80 sense), 5'-GGGATTACATATGTCTGTGGTACCGCCCAATC-3'; GST-25-(1–80 antisense), 5'-GCGGGATCCTCATACAGTCTCCTCATTC-3'; GST-25-(81–160 sense), 5'-GGGATTACATATGGAAGGGGTTCTGATTG-3'; GST-25-(81–160 antisense), 5'-GCGGGATCCTCAATATGGATACTGAGGAG-3'; GST-25-(161–227 sense), 5'-GGGATTACATATGATTCTGTG-CACATATTAC-3'; GST-25-(161–227 antisense), 5'-GCGGGATCCTCA-GTTGTAAATAAAATTG-3'; 5'-U2AF65, 5'-ACGCGGATCCATGTCCG-ACTTCGACGAG-3'; 3'-U2AF65, 5'-ATCCGCTCGAGCTCCAGAAAGT-CCCAGCGG-3'; 5'-U2AF35, 5'-ACGCGGATCCATGCGCGGAGTATCT-GGC-3'; 3'-U2AF35, 5'-ATCCGCTCGAGTCCAGATGCCCCAGATC-3'; GST-68, 5'-RS, 5'-GAACGCGGATCCCTGCAAGAACGCCATTGAG-3'; 3'-RS, 5'-ATCCGCTCGAGTCTTAACGATGACGATATTC-3'; 5'-RRM, 5'-GATCCCCCGGAATGGCGGACGCGGTGGA-3'; 3'-RRM, 5'-CTCCGGATCCCAGAAATCTGTCCACC-3'; 3'-Pro, 5'-CTCCGGATCCCCTCAATGGCGTCTTTCG-3'; Bell (5'-end), 5'-TGGGCATG-ATCAGATAGA-3'; Bsu36I (3'-end), 5'-GACCTAAGGGTGGACG-3'; RNP2 sense, 5'-ATTGCATTATATATTGTAGTTCTAACATGGTGGAC-3'; RNP2 antisense, 5'-TCCACATGTTGAACTCAACAATATGAAATGCAA-3';  $\Delta$ RS sense, 5'-AACAGTGGCATGCCTACATC-3';  $\Delta$ RS anti-sense, 5'-CACCCAAGCTTCTACTCAATTCCATGAAGGCA-3'; ythSRs, 5'-CACCGGAATTCGAGTCCAAGTCTTATGG-3'; ythSRas, 5'-TAACCGCTCGACTAACGATGACGATATTC-3'; 5'-TraX, 5'-AGTCTAGAAG-CGACAGCGGCGAGCA-3'; 3'-TraB, 5'-ACGGATCCTTAATAGCGAC-GAGGTGAGT-3'; 5'-9G8X, 5'-AGTCTAGATCGCGTACGGGCGGTA-3'; and 3'-9G8B, 5'-ACGGATCCTCAGTCCATTCTTTCAGGA-C-3'.

**Yeast Two-hybrid Screen**—Yeast two-hybrid library screening and analysis were performed as described in the HybriZAP-2.1 two-hybrid protocol provided by Stratagene with the yeast strain YRG-2. The C-terminal charged domain of CF  $I_m$ 68 (aa 481–551) was amplified by PCR with the oligonucleotides ythSRs and ythSRas and then cloned into the EcoRI and SmaI restriction sites of the yeast vector pBD-Gal4. This construct was used to screen a pre-made HeLa cell cDNA library that was cloned into the yeast vector pAD-GAL4 (Stratagene). Reporter

genes in the HybriZAP-2.1 vector system are  $\beta$ -galactosidase (*lacZ*) and histidine (*HIS3*). The ability to express the reporter *HIS3* gene was tested on SC plates lacking leucine, tryptophan, and histidine but containing 25 mM 3-aminotriazole.  $\beta$ -Galactosidase activity was detected by filter lift assays (18). Library plasmids from positive colonies were recovered in *Escherichia coli* and sequenced.

**Recombinant Protein Expression and Purification**—For expression of recombinant GST fusion proteins in *E. coli*, the sequence of interest was cloned into the polylinker region of the pGDV expression plasmid (19). The coding region of CF  $I_m$ 25 was inserted into the NcoI and XhoI sites. To express fragments of the 25-kDa subunit of CF  $I_m$ , the open reading frame was divided into three regions corresponding to amino acids 1–80, 81–160, and 161–227 that were amplified by PCR with the respective primers. The GST-25 mutant containing an internal deletion corresponding to amino acids 140–168 was obtained by digestion of the cDNA with Tth111I and Bsu36I, filling-in of the extremities, and religation. To obtain recombinant GST-U2AF subunits, the entire U2AF65 and U2AF35 open reading frames were amplified by PCR from a HeLa cDNA library with the oligonucleotides 5'-U2AF65 and 3'-U2AF65 and 5'-U2AF35 and 3'-U2AF35, respectively. The fragments were cloned into the BamHI and XhoI sites of pGEX vector (Amersham Biosciences).

The CF  $I_m$ 68 mutants RS, RRM, RRM/Pro, and Pro/RS were generated by PCR with the respective oligonucleotides, cloned into pGemTeasy (Promega), and sequenced by Centro Recherche Interdipartimentale Biotecnologie Innovative (Università di Padova). The fragments were excised with XmaI and BamHI and cloned into pGDV that had been linearized with the same enzymes.

*E. coli* BL21 (DE3) were transformed with these plasmids, and protein expression was induced with isopropyl-1-thio- $\beta$ -D-galactopyranoside. Whereas GST-U2AF65, GST-25, and its deletion mutants were present in high amounts in the soluble fraction, soluble full-length GST-59 and GST-68 could neither be recovered from *E. coli* BL21 (DE3) pLysS nor from Epicurian *Coli* BL21 codon plus (Stratagene). In contrast, the expression of several mutant derivatives was successful. Although most of the induced proteins were in inclusion bodies, a significant amount was present in the soluble fraction. The proteins in the supernatant were bound to GST-Sepharose 4B (Amersham Biosciences) and purified according to the manufacturer's instructions.

Expression of wild type CF  $I_m$ 68 in baculovirus was achieved with the BAC-To-BAC baculovirus expression system (Invitrogen). The entire coding region together with 162 nucleotides of the 3'-untranslated region and eight additional adenosines was inserted into the EcoRI site of the baculovirus transfer vector pFASTBAC-Hta. Recombinant viruses were prepared by transforming the plasmid into DH10BAC-competent cells and inducing the formation of the recombinant bacmids. High molecular weight DNA was then prepared from selected *E. coli* clones, and 1–2  $\mu$ g were used to transfect  $1 \times 10^6$  *Spodoptera frugiperda* (Sf9) insect cells. After 2–3 days, the supernatant was recovered and ~1 ml of the virus was used to infect  $2 \times 10^6$  cells that had been seeded on a 10-ml plate. The plates were again incubated at 27 °C for 3–4 days. At this stage, the cells were analyzed for protein expression by Western blot analysis. 2 ml of the resulting virus was added to a 30-ml suspension culture of Sf9 cells at a concentration of  $1 \times 10^6$  cells/ml. After 4–5 days, cells were harvested, washed in PBS, and resuspended in approximately two packed volumes of lysis buffer (200 mM NaCl, 50 mM Tris, 10% glycerol, 0.02% Nonidet P-40, 0.5 mM phenylmethylsulfonyl fluoride, 0.7  $\mu$ g/ml pepstatin, 0.4  $\mu$ g/ml leupeptin). The lysate was centrifuged after incubation on ice for 30 min, and proteins were purified by Ni<sup>2+</sup>-nitrilotriacetic acid affinity chromatography.

**GST Pull-down Experiments**—0.5–1  $\mu$ g of purified GST-25 protein was incubated with ~100  $\mu$ g of either HeLa cell nuclear extract or whole cell extracts prepared from transfected HeLa or human embryonic kidney 293 cells for 1 h at 4 °C. 50  $\mu$ l of glutathione-Sepharose beads then were added and incubated for an additional 2 h. The resin was washed three times with 1 ml of IPP150. Proteins were eluted with 2 $\times$  Laemmli buffer and separated by 12% SDS-PAGE. Proteins then were blotted onto nitrocellulose, and the membrane was probed with anti-CF  $I_m$ 68 polyclonal antibodies or anti-T7 monoclonal antibody (Novagen).

[<sup>35</sup>S]Methionine (Amersham Biosciences) was used to label proteins in a transcription/translation system (Promega). Purified GST or GST fusion proteins (~500 ng) were incubated with *in vitro* translated proteins in 30  $\mu$ l of total reaction volume. Volumes were adjusted with GST buffer (75 mM KCl, 50 mM Tris-HCl, pH 7.9, 10% glycerol, 10 mM reduced glutathione, 0.01% Nonidet P-40, 1 mM dithiothreitol, 50  $\mu$ g/ml bovine serum albumin). After incubation for 1 h at 4 °C, 25  $\mu$ l of glutathione-Sepharose beads were added and the volume was adjusted to 500  $\mu$ l with 1 $\times$  PBS, 0.01% Nonidet P-40. Incubation was continued for 1 h,

and the resin was washed three times with IPP150. Proteins were eluted from the beads by the addition of  $2\times$  Laemmli buffer, separated by 10% SDS-PAGE, and visualized by autoradiography.

CF  $I_m$  subunits were cloned into the BlueScript vector (KS, Stratagene), which allows *in vitro* transcription/translation with the TNT-coupled reticulocyte lysate system (Promega). The open reading frame of the 25-kDa subunit was cloned into BamHI and EcoRI, respectively (pBS25). The open reading frames of CF  $I_m$ 59 and CF  $I_m$ 68 were inserted into the EcoRI site of BlueScript (Stratagene). The CF  $I_m$ 68 mutant that carries two point mutations in the RNP2 motif (G86V,N87V) was generated with two consecutive PCR reactions with primers BclI (5'-end), Bsu36I (3'-end), RNP2 sense, and RNP2 antisense. Two internal BclI and Bsu36I restriction sites were used to substitute the wild type sequence with the PCR fragment carrying the mutations.

The  $\Delta$ SR mutant was generated by PCR with pBS68 and primers that contained the SphI ( $\Delta$ RS sense) and HindIII ( $\Delta$ RS antisense) restriction sites. The amplified fragment was inserted into pBS68 that had been digested with SphI and HindIII and resulted in a C-terminally truncated version lacking aa 482–551 of the full-length protein.

**RNA Binding Assays**—GST pull-down experiments were adapted from Dichtl and Keller (19). Approximately, 500 ng of GST fusion protein were incubated with 30–100 fmol of labeled RNA under cleavage conditions (20) for 30 min at 30 °C. GST-Sepharose was added and incubated for 1 h at 4 °C. The beads were washed extensively with  $1\times$  PBS, 0.01% Nonidet P-40 and resuspended in the proteinase K mixture. After 30 min, the supernatant was transferred to a fresh tube and the RNA was precipitated with ethanol and resolved on 6% denaturing polyacrylamide gels.

For UV cross-linking experiments, 20- $\mu$ l reactions were set up on ice as follows. To 8  $\mu$ l of premixture (2 mM dithiothreitol, 0.01% Nonidet P-40, 20 mM creatine phosphate, 2% polyethylene glycol, 0.4 mM cordycepin, 5 units/ $\mu$ l RNA Guard, 0.025  $\mu$ g/ $\mu$ l creatine kinase, 1.5 mM  $MgCl_2$ ), either 1 pmol of CF  $I_m$ 25/68 and CF  $I_m$ 68, 5 pmol of CF  $I_m$ 25, or 4 pmol of purified CF  $I_m$  was incubated with 200 fmol of labeled L3 pre-mRNA. These proteins had been expressed in *E. coli* and assembled *in vitro* as described previously (11). 75 mM potassium acetate and 200 fmol of RNA substrate (resuspended in water) were added, and volumes were adjusted with buffer E. Reactions were incubated for 10 min at room temperature and exposed to UV light ( $2\times 250$  mJ, UV Stratalinker 1800, Stratagene) in a microtiter plate at room temperature. Samples were treated with 5 ng/ $\mu$ l RNase A (Sigma) for 1 h at 37 °C, and proteins were separated by SDS-PAGE and exposed on x-ray film.

**Cell Culture and Transfections**—HeLa and human embryonic kidney 293 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum and transfected with 1  $\mu$ g of plasmid DNA/60-mm dish containing ~60% confluent cells with the calcium phosphate method. The epitope-tagged 9G8 and hTra2 $\beta$  expression plasmids were constructed by amplifying the cDNAs isolated in the yeast two-hybrid screen with specific primers and subcloning the resulting PCR products as a XbaI-BamHI fragment into pCGTHCF<sub>1</sub>T7 expression vector (21). Amplification was performed with Amplitaq Gold (Roche Applied Science) and oligonucleotides (5'-9G8X, 3'-9G8B, 5'-TraX, and 3'-TraB) purchased from MWG. Construction of pCG-SRp20 was described previously (22).

**Indirect Immunofluorescence**—Cells were either plated on glass coverslips and transfected or distributed after transfection on polylysine-coated coverslips. 24 h after transfection, cells were washed with PBS, fixed with 3% paraformaldehyde, and permeabilized with 0.2% Triton X-100 for 5 min on ice. Cells were incubated 1 h at 37 °C with the primary antibody, washed three times with PBS containing 2% bovine serum albumin, and again incubated for 1 h at 37 °C with the appropriate secondary antibody.

For localization of endogenous CF  $I_m$ , polyclonal antibodies raised against CF  $I_m$ 68 and CF  $I_m$ 25 (11) were used at a 1:500 and 1:200 dilution, respectively. Colocalization studies of GFP fusion proteins and endogenous SRp20 protein were performed with a mouse monoclonal antibody (Zymed Laboratories Inc.) used as a 1:100 dilution and detected with a Texas Red-conjugated anti-mouse IgG (Molecular Probes, 1:1000) secondary antibody. SC35 was detected with a mouse monoclonal antibody (Sigma, 1:2000), and endogenous CPSF was detected with a mouse monoclonal antibody (3). Secondary antibodies were Alexa 488-conjugated anti-mouse IgG or Alexa 594-conjugated anti-rabbit IgG (Molecular Probes) used as a 1:1000 dilution.

GFP expression plasmids were constructed by cloning DNA fragments encoding full-length and mutant CF  $I_m$ 68 into pEGFP-C1 plasmid (Clontech). GFP-68 was obtained by digesting pBS68 with BamHI and SalI and inserting the two fragments that are generated (BamHI/

SalI and SalI/SalI) into pEGFP-C1 linearized with BglII and SalI. GFP-68 deletion mutants were excised from the corresponding GST constructs with XmaI and BamHI (filled) and cloned into pEGFP-C1 linearized with XmaI and XbaI (filled).

**Confocal Fluorescence Microscopy**—Fluorescence microscopy of fixed cells was carried out with a Nikon YFL microscope equipped with a  $\times 60$ , 1.4 oil Plan-Apochromat objective or a Leica DM IRE2 confocal microscope equipped with an argon/krypton laser (488 nm) to excite GFP fluorescence and a helium/neon laser (543 nm) to excite the Alexa 594 fluorescence and a  $\times 63$ , 1.4 oil HCX Plan-Apochromat objective. For double-labeling experiments, images from the same focal plane were sequentially recorded in different channels and superimposed.

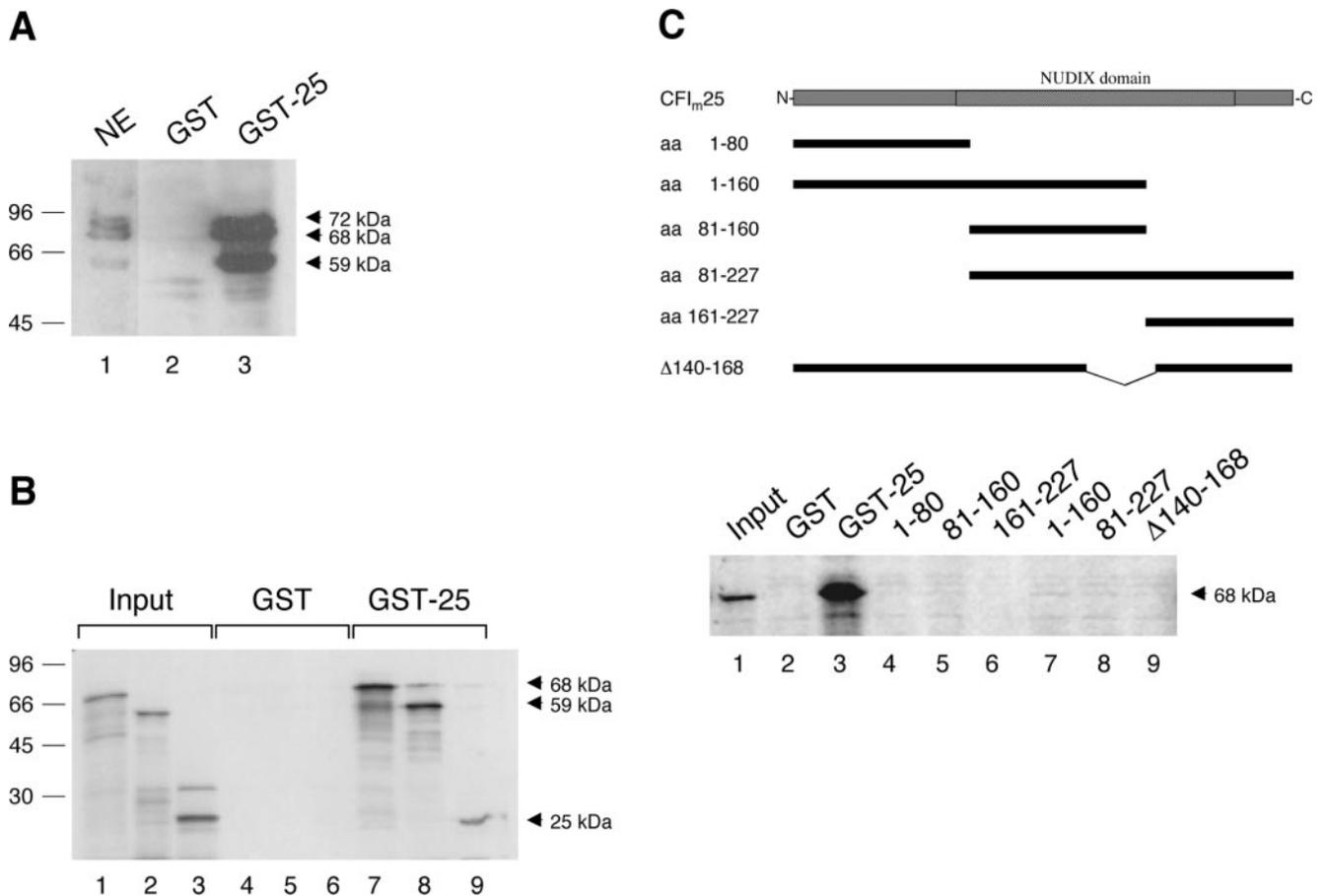
## RESULTS

**CF  $I_m$ 68 Interacts with the 25-kDa Protein via Its RRM**—Four polypeptides with apparent molecular masses of 25, 59, 68, and 72 kDa copurified with CF  $I_m$  activity. Previous studies have shown that the 25- and 68-kDa proteins are sufficient to reconstitute CF  $I_m$  activity *in vitro*. It has been suggested that CF  $I_m$  forms a heterodimer composed of the 25-kDa subunit and one of the larger polypeptides, which seem to be related as they are recognized by the same antibody that was raised against the 68-kDa protein (11). To verify that the larger CF  $I_m$  subunits associate with the 25-kDa protein, we expressed the 25-kDa protein as a GST fusion (GST-25) in *E. coli* and incubated it with HeLa cell nuclear extract. Western blot analysis with anti-CF  $I_m$ 68 antibody detected all three larger CF  $I_m$  subunits (Fig. 1A, lane 3). This result was confirmed by incubating GST-25 with CF  $I_m$ 68 and CF  $I_m$ 59 that had been transcribed and translated *in vitro* in the presence of [<sup>35</sup>S]methionine (Fig. 1B, lanes 7 and 8). In addition, CF  $I_m$ 25 appeared to interact weakly with itself (Fig. 1B, lane 9), which might indicate that heterodimers bind to each other to form larger CF  $I_m$  complexes. The experiments were carried out in the presence of RNase A, and the interactions thus were not mediated through RNA.

Since no known protein-protein interaction motif has so far been identified in the sequence of the 25-kDa polypeptide, we wished to map the region responsible for its binding to CF  $I_m$ 68. For this purpose, the fragments of the 25-kDa protein were expressed as GST fusions in *E. coli* (a schematic representation of the different mutants is shown in Fig. 1C). Pull-down experiments performed with the GST-25 fragments and *in vitro* transcribed/translated CF  $I_m$ 68 only detected association with the full-length GST-25 (Fig. 1C, lane 3). None of the deletion constructs was able to interact with CF  $I_m$ 68, suggesting that intact 25-kDa protein is required for the stable binding of the two proteins.

**CF  $I_m$ 25 Interacts with PAP and PABPN1**—The binding of CPSF and CstF to conserved sequence elements in the pre-mRNA located upstream and downstream of the poly(A) site mediates specific recognition of the polyadenylation site. As has been shown previously, the association of CF  $I_m$  with the substrate leads to a faster assembly of the 3'-end-processing complex (11). Therefore, CF  $I_m$  may help to recruit CPSF and CstF to the pre-mRNA. To test this hypothesis, we tested in GST pull-down experiments for the direct interaction of CF  $I_m$  with CPSF and CstF. However, our results with GST-25 suggest only weak association with *in vitro* transcribed/translated CstF50, CstF77, and CPSF160 and no detectable interaction with the other CstF and CPSF subunits (results not shown). Therefore, if the recruitment of CstF and CPSF is facilitated by CF  $I_m$ , this does not appear to be mediated by a direct interaction with the smallest subunit of CF  $I_m$ .

An interaction between CF  $I_m$ 25 and PAP has been described previously (16). The authors isolated the 25-kDa subunit of CF  $I_m$  in a yeast two-hybrid screen with the C-terminal domain of mouse PAP (aa 472–739) as a bait. In addition to PAP, we found that GST-25 interacts with PABPN1 (Fig. 2A, lane 8). A



**FIG. 1. The 25-kDa protein interacts with the larger CF  $I_m$  subunits.** *A*, the 25-kDa protein can coprecipitate the large subunits from nuclear extract. GST protein (lane 2) or GST-25 (lane 3) was preincubated with HeLa cell nuclear extract as described under “Experimental Procedures.” GST-Sepharose-bound proteins were resolved by SDS-PAGE and analyzed by Western blotting with a polyclonal antibody specific for the larger CF  $I_m$  subunits (11). Lane 1, 2  $\mu$ g of HeLa cell nuclear extract. *B*, GST-25 interacts directly with CF  $I_m$ 59, CF  $I_m$ 68, and with itself. *In vitro* translated [ $^{35}$ S]methionine-labeled 25-, 59-, and 68-kDa polypeptides were incubated with either GST or GST-25 protein as described under “Experimental Procedures.” Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input was loaded in lanes 1, 2, and 3. *C*, interaction with CF  $I_m$ 68 requires a full-length 25-kDa protein. *In vitro* translated [ $^{35}$ S]methionine-labeled CF  $I_m$ 68 subunit was incubated with the indicated purified GST fusion proteins (GST-25, GST-25/1–80, GST-25/81–160, GST-25/161–227, GST-25/1–160, GST-25/81–227, or GST-25/ $\Delta$ 140–168) as described under “Experimental Procedures.” Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input is shown in lane 1. A schematic representation of the recombinant 25-kDa proteins used in this study is shown in the upper panel. The position of the NUDIX domain is indicated by the shaded box. NE, nuclear extract.

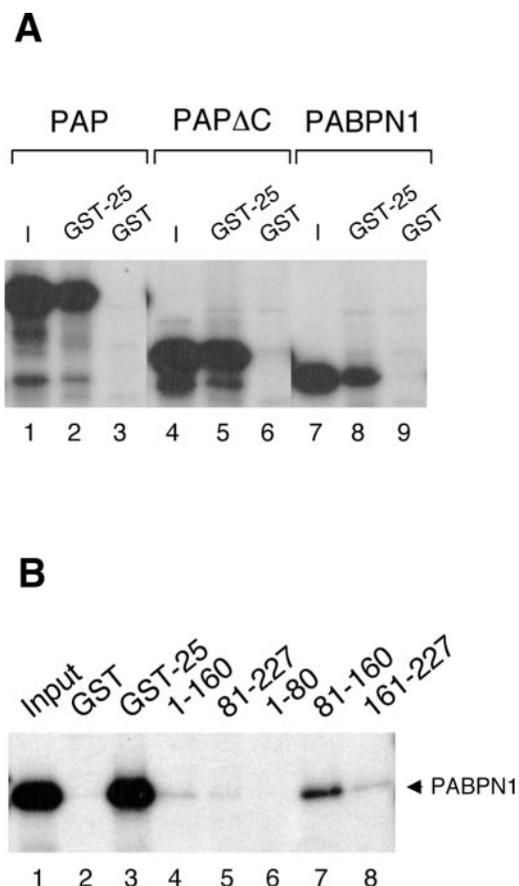
deletion of the C-terminal 226 residues of PAP did not affect its binding to GST-25 (Fig. 2A, PAP $\Delta$ C, lane 5). Incubation of GST-25 with HeLa cell nuclear extract and subsequent Western blot analysis confirmed the interaction of GST-25 with PAP and PABPN1 (results not shown).

The region within CF  $I_m$ 25 responsible for its interaction with PAP and PABPN1 was mapped and is shown in Fig. 2B. GST fragments comprising amino acids 1–80, 81–160, 161–227, 1–160, and 81–227 of CF  $I_m$ 25 were incubated with *in vitro* transcribed/translated PABPN1. The full-length 25-kDa protein was able to efficiently interact with PABPN1 (Fig. 2B, lane 3), and a weaker interaction was found with fragment 81–160 (Fig. 2B, lane 7). Several experiments confirmed that weak interactions could take place with the constructs comprising amino acids 1–160 and 81–227 (Fig. 2B, lanes 4 and 5, respectively), emphasizing that mainly the region between amino acids 81 and 160 of CF  $I_m$ 25 mediates association with PABPN1. Similar results were obtained with PAP (results not shown), suggesting that PAP and PABPN1 may share the same binding site on CF  $I_m$ 25 or that their interaction domains are very close.

*The N-terminal RRM of CF  $I_m$ 68 Is Sufficient for the Interaction with CF  $I_m$ 25*—To identify functional domains of the 68-kDa subunit of CF  $I_m$ , we took a similar approach to that

used for the characterization of CF  $I_m$ 25. Several CF  $I_m$ 68 deletion and point mutants were generated and cloned into a bacterial expression vector (Fig. 3A). Unfortunately, neither the full-length 68-kDa protein nor any of the mutants containing the proline-rich domain within the central region, with the exception of the Pro/RS fragment, could be efficiently expressed as GST fusion proteins in *E. coli*. *In vitro* translated CF  $I_m$ 25 was found to associate with both GST-68RRM and GST-68RRM/RS (Fig. 3B, lanes 3 and 5) but not with the GST-68RS or GST-68Pro/RS fragments (Fig. 3B, lanes 4 and 6). These results were confirmed by pull-down assays in which recombinant GST-25 was incubated with *in vitro* translated wild type or mutant CF  $I_m$ 68. As shown in Fig. 3C, both wild type CF  $I_m$ 68 and a C-terminal truncation lacking the RS-like domain are efficiently precipitated by GST-25 (Fig. 3C, lanes 3 and 6 respectively). Instead, the substitution of two amino acids within the RNP2 motif of the RRM of CF  $I_m$ 68 (G87V,N88V) abolished interaction with the 25-kDa subunit (Fig. 3C, lane 9). These results suggest that the interaction between CF  $I_m$ 68 and CF  $I_m$ 25 is mediated by the RRM and, in particular, by the RNP2 motif within the 68-kDa protein.

*The 25-kDa Subunit Is the Major RNA-binding Component of CF  $I_m$* —Whereas the 68-kDa subunit of CF  $I_m$  contains an RRM, the 25-kDa polypeptide does not display homology to any



**FIG. 2. The central region of CF  $I_m$ 25 is required for the interaction with PAP and PABPN1.** *A*, GST-25 directly interacts with PAP and PABPN1. GST (lanes 3, 6, and 9) or GST-25 (lanes 2, 5, and 8) was preincubated with *in vitro* translated [ $^{35}$ S]methionine-labeled PAP, PAP $\Delta$ C, or PABPN1 proteins as described under "Experimental Procedures." Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input is shown in lanes 1, 4, and 7. *B*, residues 81–160 of CF  $I_m$ 25 are sufficient for interaction with PABPN1. *In vitro* translated [ $^{35}$ S]methionine-labeled PABPN1 was incubated with the indicated purified GST fusion proteins (GST-25, 1–160, 81–227, 1–80, 81–160, or 161–227) as described under "Experimental Procedures." Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input is shown in lane 1.

known RNA binding domain. Nevertheless, it was shown previously that the CF  $I_m$  subunits purified from HeLa cell nuclear extract can be UV-cross-linked to an RNA substrate (10). We analyzed the interaction of CF  $I_m$  subunits with RNA in more detail. In a first attempt, we performed UV-cross-linking experiments with recombinant CF  $I_m$ 68 purified from baculovirus-infected Sf9 cells and recombinant CF  $I_m$ 25 produced in *E. coli*. Similar amounts of proteins were exposed to UV light in the presence of  $^{32}$ P-labeled pre-mRNA substrate (for details see "Experimental Procedures"). The first lane of Fig. 4A shows that in addition to the 59-kDa protein, the 25- and 68-kDa subunits of CF  $I_m$ , which had been purified from HeLa cells, could be UV-cross-linked to RNA. However, when the single recombinant subunits were assayed separately for RNA binding, only the 25-kDa protein could efficiently be cross-linked to RNA (Fig. 4A, lane 2), whereas the signal detected with the 68-kDa polypeptide was very weak (Fig. 4A, lane 3). In contrast, the cross-linking efficiency of CF  $I_m$  assembled *in vitro* from recombinant 25- and 68-kDa polypeptides was comparable to that obtained with the purified factor (Fig. 4A, lane 4). These results suggest that a heterodimer consisting of the 25- and the 68-kDa proteins binds RNA more efficiently than each subunit separately.

To determine which region of the 25-kDa subunit of CF  $I_m$  was responsible for the interaction with the pre-mRNA, the same GST fragments that were previously used for the protein-protein interaction experiments (Fig. 1C) were assayed in an RNA pull-down experiment. The different GST fragments were incubated under cleavage conditions with the labeled substrate, and RNA-protein complexes were subsequently immobilized on GST-Sepharose. Fig. 4B shows that the fragment consisting of amino acids 1–160 binds to RNA as efficiently as the full-length GST-25 (Fig. 4B, lanes 2 and 3). Smaller fragments comprising amino acids 1–80 and 81–160, respectively, also interact with RNA albeit weaker (Fig. 4B, lanes 5 and 6), whereas the C-terminal fragment (aa 161–227) and the fragment consisting of amino acids 81–227 fail to do so. We conclude from our data that the C-terminally 80 amino acids are not required for RNA binding.

In a similar way, we also mapped the RNA binding region of the 68-kDa polypeptide. As shown in Fig. 4C, the N-terminal fragment containing the RRM did not significantly bind the RNA (lanes 3–6). In contrast, both the C-terminal RS-like domain and the RRM/RS fusion polypeptides were able to efficiently pull-down the substrate (lanes 7–14). This probably reflects an ionic interaction between the positively charged RS-like domain and the RNA.

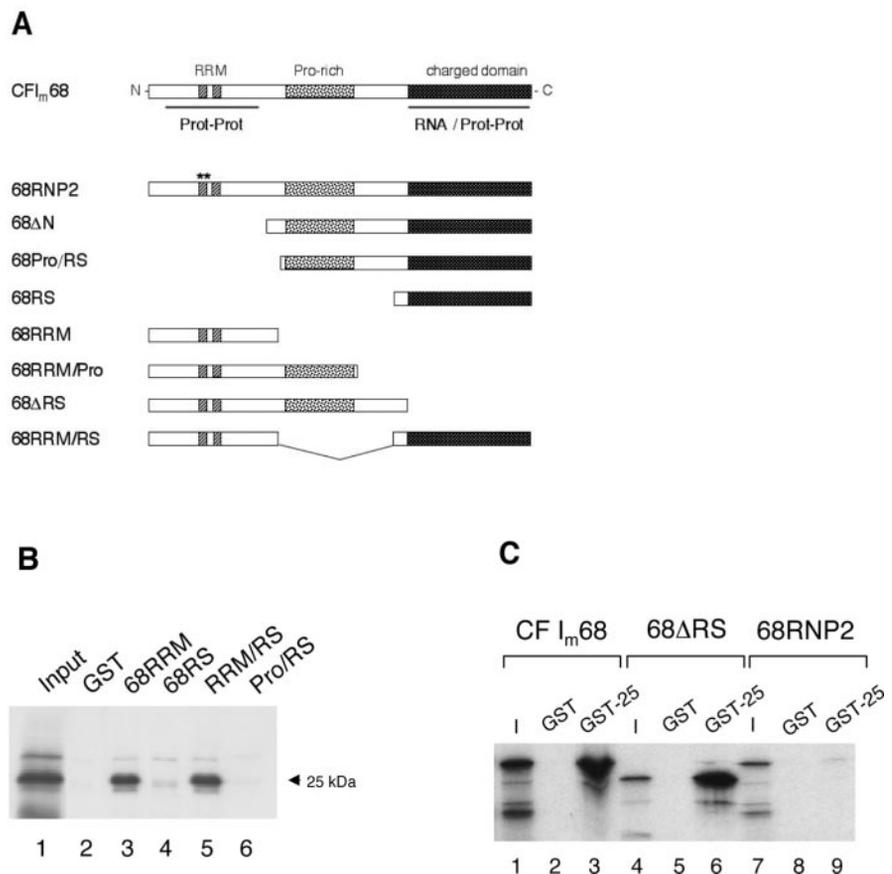
Synergistic RNA binding of CF  $I_m$ 25 and CF  $I_m$ 68 was confirmed by RNA pull-down assays with the same GST-68 proteins used before and in the presence of increasing amounts of the recombinant histidine-tagged 25-kDa subunit (Fig. 4D, His-CF  $I_m$ 25). The addition of the 25-kDa subunit increased the amount of RNA precipitated by the GST-RRM/RS protein (Fig. 4D, lanes 13–16) and enabled the GST-RRM fragment to pull-down RNA (Fig. 4D, lanes 3–6). This could be explained by the concomitant interaction of the 25-kDa polypeptide with the RNA and the GST-tagged RRM fragment of the 68-kDa protein. The addition of CF  $I_m$ 25 to the RS-like domain of CF  $I_m$ 68 did not increase the amount of precipitated RNA (Fig. 4D, lanes 8–11), most probably because the 25-kDa protein cannot interact with GST-68RS. Taken together, these results suggest that the 25-kDa protein and one of the larger subunits cooperate to recognize the CF  $I_m$  binding sites on the pre-mRNA.

**The RS-like Domain of CF  $I_m$ 68 Associates with SR Proteins**—The C-terminal portion of CF  $I_m$ 68 consists of RD-, RE-, and RS-dipeptide repeats up to a total length of 60 amino acids. Similar alternating charge domains consisting of RS repeats have been described to occur in a family of splicing factors termed SR proteins. The RS domain has been shown to promote protein-protein interactions, direct subcellular localization, and, in certain situations, nucleocytoplasmic shuttling. To isolate proteins that interact with the C terminus of CF  $I_m$ 68, a yeast two-hybrid screen was performed with a fragment encoding amino acids 481–551 (called 68RS) as bait. A HeLa cell cDNA library was screened, and an analysis of positive colonies revealed among the strongest interactors some known members of the SR protein family, namely hTra-2 $\beta$  (23, 24), SRp20 (also named X16) (25), and 9G8 (26).

To verify these interactions *in vivo*, we used GST-68RS in pull-down assays with total extracts of human cells transfected with plasmids encoding cDNAs for hTra-2 $\beta$ , SRp20, and 9G8 (Fig. 5A). All of the constructs encoded proteins with a bacteriophage T7 gene 10 (T7) epitope tag at their N termini, allowing the detection of the exogenous proteins with a monoclonal antibody that recognizes this epitope (see "Experimental Procedures") (21). Western blot analysis revealed that a significant fraction of hTra-2 $\beta$ , SRp20, and 9G8 was found to associate with the RS-like domain of CF  $I_m$ 68 (Fig. 5, lane 3 of each panel). However, none of those SR proteins was detected if

FIG. 3. The RRM domain of CF  $I_m$ 68 is required for the interaction with the 25-kDa protein.

**A**, schematic representation of the CF  $I_m$ 68 domain mutants. The *open* and *solid* boxes indicate the regions of the proteins present in each mutant relative to the 552 amino acid wild type protein shown at the *top*. Asterisks represent substituted amino acids within the RNP2. The *line* indicates that residues 222–419 are missing in the RRM/RS protein. The regions required for protein-protein interaction and RNA-binding are indicated. **B**, CF  $I_m$ 68 interacts directly with the 25-kDa subunit via the RRM domain. *In vitro* translated [ $^{35}$ S]methionine-labeled CF  $I_m$ 25 subunit was incubated with the purified GST fusion proteins indicated (68RRM, 68RS, 68RRM/RS, or 68Pro/RS) as described under "Experimental Procedures." Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input is shown in *lane 1*. **C**, residues within the RRM and in particular the RNP2 motif of CF  $I_m$ 68 are essential for the interaction with CF  $I_m$ 25. GST (*lanes 2, 5, and 8*) or GST-25 (*lanes 3, 6, and 9*) was preincubated with the indicated *in vitro* translated [ $^{35}$ S]methionine-labeled proteins (CF  $I_m$ 68, 68 $\Delta$ RS, and 68RNP2) as described under "Experimental Procedures." In the 68RNP2 mutant protein, glycine 86 and asparagines 87 within the RNP2 were changed to valines. Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input is shown in *lanes 1, 4, and 7*.



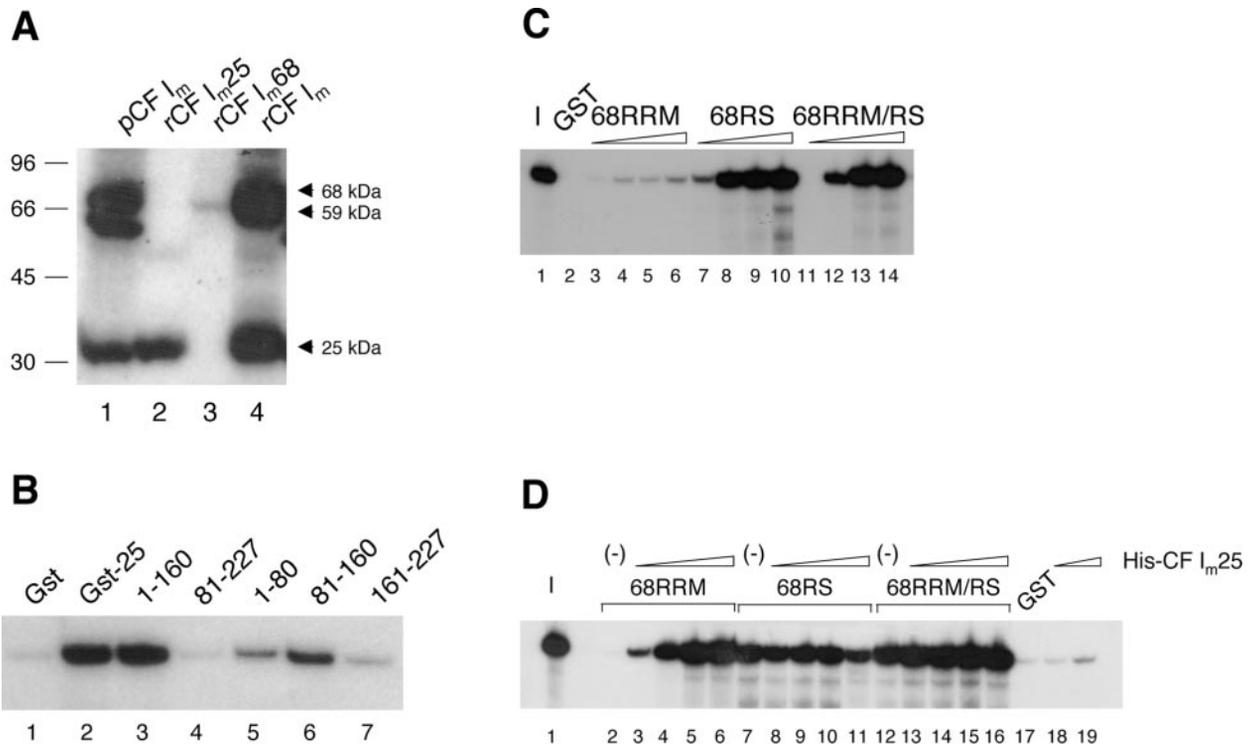
incubated with GST (*lane 1*), GST-25 (*lane 2*), GST-RRM68 (*lane 4*), or GST-U2AF65 (*lane 5*), indicating that the interaction is specific for the RS-like domain of CF  $I_m$ 68. The interaction with hTra- $\beta$ 2, SRp20, and 9G8 was confirmed *in vitro* by GST-pull down experiments with *in vitro* transcribed/translated proteins (results not shown). These results indicate that CF  $I_m$ 68 interacts directly with a specific subset of SR proteins, raising the intriguing possibility that CF  $I_m$  contributes to the definition of the last exon during the processing of intron-containing pre-mRNAs.

**CF  $I_m$  Localizes in a Punctate Pattern in the Nucleus**—Several studies have addressed the subcellular distribution of components of the 3'-end-processing machinery. CstF64 and CPSF100 were found concentrated in foci adjacent to Cajal bodies that were termed "cleavage bodies" to reflect their content of pre-mRNA cleavage factors (28). Different 3'-end formation factors can also accumulate at sites of transcription (29). Despite being a polyadenylation factor, PABPN1 appears to localize in nuclear speckles (30). Speckles (also known as "interchromatin granule clusters") are subnuclear structures that are believed to serve as storage sites for splicing factors, including SR proteins (for review see Ref. 31). The RS domain is a major determinant for the subcellular localization of SR proteins to speckles, and it can also function as a nuclear localization signal (22, 33).

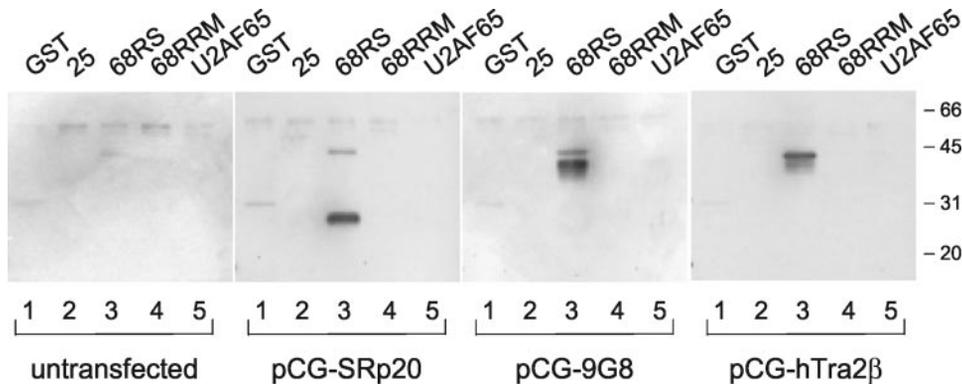
Because of the similarity of CF  $I_m$  large subunits and SR proteins, we wanted to determine their subcellular localization. Immunofluorescence microscopy with polyclonal antisera raised against the 25- and the 68-kDa CF  $I_m$  subunits (11) revealed that endogenous CF  $I_m$  is localized in the nucleoplasm with the exclusion of the nucleoli, with a granular distribution, and with concentration in few discrete foci (Fig. 6, *c* and *c'*). To confirm the staining pattern observed with the antibodies, we transiently expressed both CF  $I_m$  subunits fused to the GFP. A

portion of the transfected cell population expressed the GFP fusion proteins at very high levels and showed aberrant localization patterns. Therefore, we selected cells that expressed low levels of the GFP fusion proteins. In these cells, indirect immunofluorescence showed that, 20–24 h after transfection, the transiently expressed GFP-68 and GFP-25 localized exclusively to the nucleus with a diffuse pattern and, in addition, within a few distinct foci, in agreement with the localization observed with the specific antibodies (Fig. 6, *b* and *b'*). To determine whether these foci corresponded to cleavage bodies, we performed double-labeling experiments with a monoclonal antibody that specifically recognizes the 100-kDa subunit of CPSF (3). However, Fig. 7A shows that the foci detected with anti-CF  $I_m$ 68 do not correspond to cleavage bodies visualized by anti-CPSF100. This unexpected result prompted us to assay markers for known subnuclear compartments to clarify the nature of the CF  $I_m$  foci. We used antibodies specific for p80 coilin to detect Cajal bodies (34) and for SC-35 to detect splicing speckles (35, 36). However, GFP-68 was neither localized to Cajal bodies (Fig. 7B, *panels a–a'*) nor to speckles (Fig. 7B, *panels b–b'*) but was always localized in close proximity to both of them. The observation that GFP-68 foci were adjacent to speckles prompted us to colocalize GFP-68 and the recently identified PSP1 protein (17). This protein accumulates in punctate subnuclear structures that are frequently localized adjacent to SC35 speckles and, therefore, are termed paraspeckles. Fig. 7B (*panels c–c'*) shows that GFP-68 and PSP1, at least partially, colocalize in paraspeckles.

**Role of Individual Domains of CF  $I_m$ 68 in Cellular Distribution and Subnuclear Localization**—To determine the role of individual domains of CF  $I_m$ 68 in nuclear and subnuclear localization, we transiently overexpressed cDNAs encoding several mutant derivatives (listed in Fig. 3A) as GFP fusion proteins in HeLa cells and determined the cellular distribution of



**FIG. 4. The small and the large subunit of CF  $I_m$  cooperate in RNA binding.** *A*, CF  $I_m$  subunits can be UV-cross-linked to RNA.  $^{32}P$ -Labeled adenovirus L3 pre-mRNA was incubated for 10 min with the indicated proteins (pCF  $I_m$ , purified CF  $I_m$ ; rCF  $I_m$ , recombinant CF  $I_m$ ) and exposed to UV light as described under "Experimental Procedures." After digestion with RNaseA, the products were analyzed by 12% SDS-PAGE and autoradiography. *B*, the N-terminal region of CF  $I_m$ 25 is required for RNA binding. RNA pull-down experiment with GST fusions of CF  $I_m$ 25 deletion mutants.  $^{32}P$ -Labeled SV40 pre-mRNA was incubated with the indicated purified GST fusion proteins as described under "Experimental Procedures." Lane 1, GST protein. *C*, the CF  $I_m$ 68RRM domain does not bind to RNA. RNA pull-down experiment with GST fusions of CF  $I_m$ 68 domain deletion mutants is shown.  $^{32}P$ -Labeled adenovirus L3 substrate was incubated with the indicated purified GST fusion proteins as described under "Experimental Procedures." Lane 1, 10% of the input L3 pre-mRNA; lane 2, GST protein. *D*, CF  $I_m$ 25 interacts with the RRM of CF  $I_m$ 68 and enhances binding to RNA. RNA pull-down experiment with GST fusions of CF  $I_m$ 68 domain deletion mutants either in the absence (lanes 2, 7, and 12) or in the presence of increasing amounts of the recombinant histidine-tagged 25-kDa subunit (lanes 3–6, 8–11, and 13–16) is shown.  $^{32}P$ -Labeled adenovirus L3 substrate was incubated with the indicated purified GST fusion proteins as described under "Experimental Procedures." Lane 1, 10% of input L3 pre-mRNA; lane 17, GST protein; lanes 18–19, increasing amounts of histidine-tagged 25-kDa protein.



**FIG. 5. The C-terminal RS-like domain of CF  $I_m$ 68 interacts with SR proteins.** HeLa or human embryonic kidney 293 cells were transfected with pCG constructs expressing hTra2 $\beta$ , SRp20, or 9G8. Whole cell extracts were incubated with either GST or the indicated GST fusion proteins (GST-25, GST-68RS, GST-68RRM, and GST-U2AF65) as described under "Experimental Procedures." Bound proteins were separated by SDS-PAGE and analyzed by Western blotting with an anti-T7 tag monoclonal antibody.

the proteins by indirect immunofluorescence microscopy (Fig. 8). We verified the expression of all of the transfected cDNAs by Western blot analysis of whole cell lysates (results not shown).

Whereas GFP alone was diffusely distributed throughout the cell (Fig. 8, *panel a'*), mutant proteins in which the N-terminal RNA binding domain was either deleted (Fig. 8,  $68\Delta N$ , *panel b'*) or carried amino acid substitutions in the RNP2 of the RRM (results not shown) were concentrated in the nucleus (excluding the nucleoli), similarly to the wild type protein (Fig. 6). All of the deletion mutants lacking the C-terminal RS-like domain localized to the cytoplasm (Fig. 8,  $68\Delta RS_{46}$ , *panel c'*; RRM and

RRM/Pro, data not shown). Interestingly, a domain search with PSORT (37) reveals the presence of a putative nuclear localization signal (RRHR) located 46 amino acids upstream of the stop codon. Consistent with this prediction, the deletion of this portion of the charged domain was sufficient to prevent nuclear import of the truncated protein (Fig. 8,  $68\Delta RS_{46}$ , *panel c'*). The requirement for the RS domain for nuclear localization was confirmed by the cellular distribution of a mutant protein in which the N-terminal RRM is fused directly to the RS-like domain (Fig. 8, *RRM/RS*, *panel d'*). This protein was found both in the cytoplasm and in the nucleus. Furthermore, the

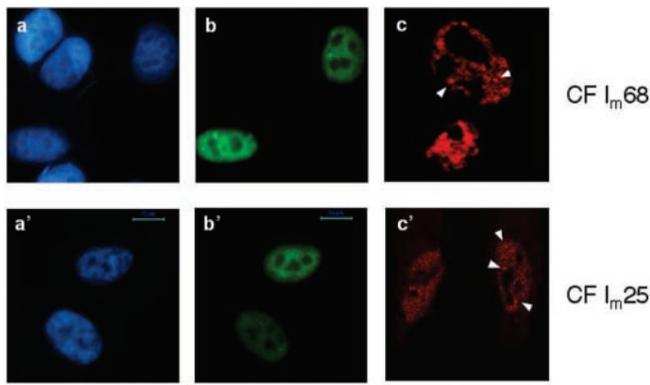


FIG. 6. **Nuclear localization of CF  $I_m$  subunits.** HeLa cells were either immunolabeled with anti-CF  $I_m68$  (panel *c*) or with anti-CF  $I_m25$  (panel *c'*) polyclonal antibodies (11) or transfected with GFP fusion constructs (panel *b*, GFP68; panel *b'*, GFP25) as described under "Experimental Procedures." Panels *a* and *a'*, 4,6-diamidino-2-phenylindole staining. The white arrowheads indicate punctuate CF  $I_m$  foci. Scale bars, 5  $\mu$ m.

fusion of the RS domain to GFP is sufficient to direct nuclear import of the chimeric protein (Fig. 8, *68RS*, panel *e'*). These results indicate that the C-terminal charged domain may contain one or more nuclear localization signals.

#### DISCUSSION

CF  $I_m$  is a component of the basic pre-mRNA 3'-end-processing complex. Reconstitution experiments suggest that CF  $I_m$  occurs as several isomers consisting of a heterodimer of a 25-kDa polypeptide and one of the three larger subunits of 59, 68, and 72 kDa (11). In this work, we have characterized the functional domains of the 25- and 68-kDa subunits. The small subunit lacks any known sequence motif involved in RNA binding or protein-protein interactions. However, our results suggest that the central part of CF  $I_m25$  (aa 81–160) mediates specific interactions with PAP and PABPN1. In contrast, the intact protein is required for association with the 68-kDa protein, indicating that a heterodimeric interaction of the 25-kDa protein with the 68-kDa polypeptide is essential for binding to the pre-mRNA substrate. This view is supported by recent SELEX experiments with a CF  $I_m68/25$  heterodimer that identified specific binding sites on the RNA (15). The RNA binding of CF  $I_m25$  *per se* does not require the C terminus of the protein.

The primary sequence of CF  $I_m68$  reveals the presence of two known sequence motifs: an N-terminal RRM of the RNP type and a C-terminal alternating charge domain enriched in RS, RD, and RE repeats that resembles the RS domain of spliceosomal SR proteins. Our results indicate that the RRM of CF  $I_m68$  is primarily engaged in protein-protein interaction with the 25-kDa subunit of CF  $I_m$ . Although RRMs are generally considered nucleic acid binding domains, several reports have implicated them in protein-protein interactions. For example, the U2 small nuclear RNP-specific protein U2B' was found to interact with U2A' via its RRM (38). Recently, the crystal structure of the complex between the *Drosophila melanogaster* Y14 and Mago proteins was determined (39). Y14 and Mago associate with spliced mRNAs and are components of the exon junction complex. Similar to CF  $I_m25$ , Mago does not reveal recognizable motifs. Whereas the amino acid sequence of Y14 predicts a canonical RRM, the structure of the complex reveals that Y14 RRM is engaged in protein recognition rather than in RNA binding. A similar interaction occurs between the two subunits of the splicing factor U2AF. Human U2AF is a heterodimer composed of a 65-kDa large subunit (U2AF65) and a 35-kDa small subunit (U2AF35). It was shown that the atypical RRM of U2AF35 interacts with the proline-rich region of

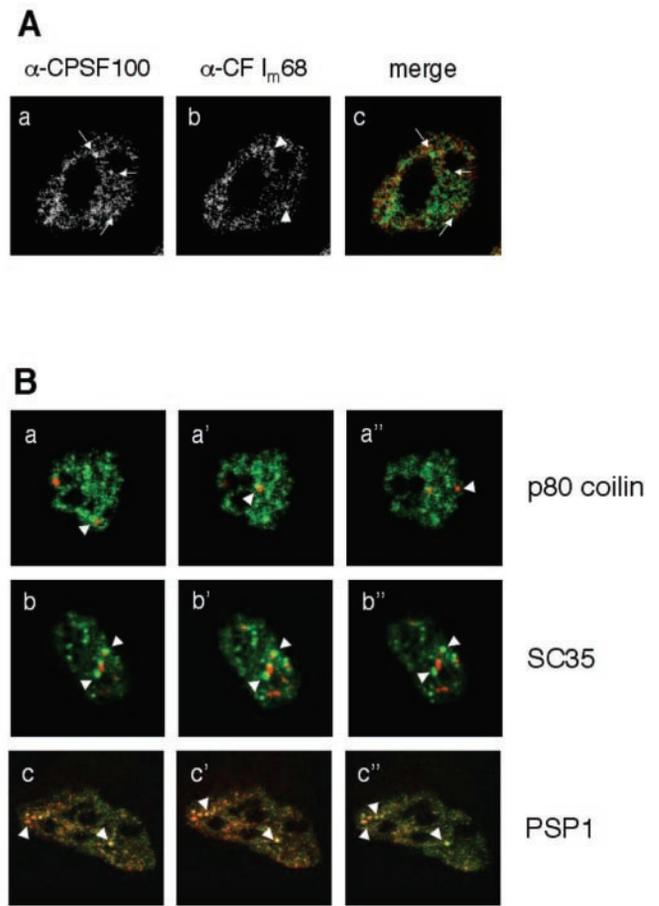
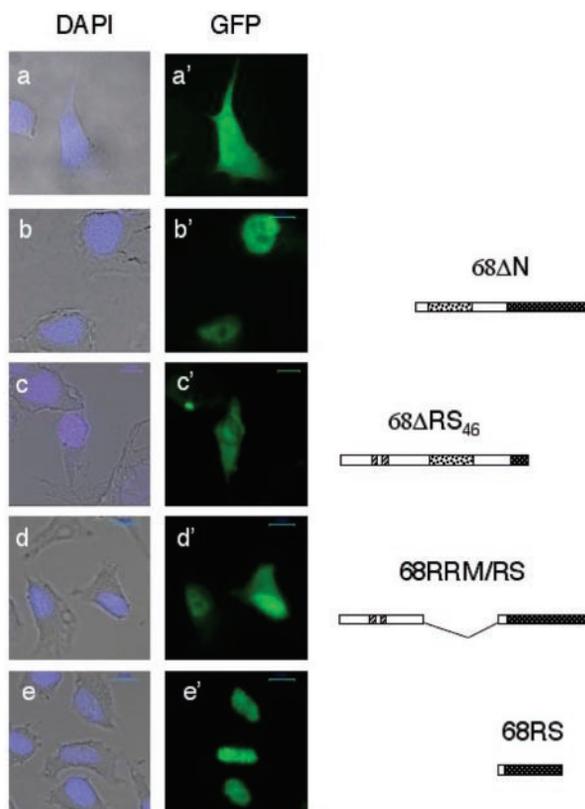


FIG. 7. **CF  $I_m68$  localizes to paraspeckles.** *A*, double labeling of HeLa cells with anti-CPSF100 monoclonal antibody (panel *a*) (3) and with anti-CF  $I_m68$  polyclonal antiserum (panel *b*) (11) as described under "Experimental Procedures." Is shown. The merge of the two channels is shown in panel *c* (green, anti-CPSF100; red, anti-CF  $I_m68$ ). Small arrows indicate the cleavage bodies that can be detected with the anti-CPSF100 antibody (3). Arrowheads indicate discrete CF  $I_m68$  foci that do not colocalize with CPSF cleavage bodies. *B*, CF  $I_m68$  partially colocalizes with PSP1 protein. HeLa cells were transiently transfected with a plasmid expressing GFP-68 protein. The cells were fixed and immunostained with antibodies against nuclear markers (red). Endogenous proteins were detected with specific antibodies followed by Alexa 594-conjugated secondary antibody as described under "Experimental Procedures." Three sequential focal planes are shown. Cajal bodies were detected with a polyclonal anti-p80 coilin antiserum (*a-a'*) (28). Arrowheads indicate adjacent GFP-68 foci (green) and coiled bodies (red). *b-b'*, splicing speckles are detected with an anti-SC35 monoclonal antibody (Sigma). Arrowheads indicate adjacent GFP-68 foci (green) and speckles (red). *c-c'*, paraspeckles are detected with an anti-PSP1 polyclonal antiserum (16). Arrowheads indicate colocalization of GFP-68 and PSP1 (yellow).

U2AF65 (40). In addition, U2AF65 was found to bind to the human splicing factor SF1 via its noncanonical third RRM (41). As in the case of U2AF where both subunits work in concert to recognize weak polypyrimidine tracts and promote U2 small nuclear RNP binding, we suspect that the 25-kDa protein and one of the larger subunits of CF  $I_m$  cooperate to recognize the CF  $I_m$  binding sites on the pre-mRNA and thus promote assembly of the 3'-end-processing complex.

The isolation of three SR proteins in a yeast two-hybrid screen with the RS-like domain of CF  $I_m68$  provides new insight into the coordination of splicing and 3'-end formation. Recognition of the 3'-terminal exons has been postulated to involve an interaction of splicing and polyadenylation factors, and a growing number of reports support this view (42–50). Because the RS domain of SR proteins mediates protein-protein interactions that are important for spliceosome assembly



**FIG. 8. Role of CF  $I_m68$  domains in cellular localization and subnuclear distribution.** The CF  $I_m68$  RS domain is required for nuclear localization. HeLa cells were transfected with plasmids encoding GFP-fusion proteins and fixed 24 h after transfection. The localization of the expressed proteins was determined by indirect immunofluorescence. *a* and *a'*, GFP; *b* and *b'*, CF  $I_m68\Delta N$ ; *c* and *c'*, CF  $I_m68\Delta RS_{46}$ ; *d* and *d'*, CF  $I_m68RRM/RS$ ; *e* and *e'*, CF  $I_m68RS$ . Bar, 5  $\mu m$ . DAPI, 4,6-diamidino-2-phenylindole.

(51), it has been suggested that CF  $I_m68$  could interact via its C-terminal alternating charge domain with one or more SR proteins and/or SR-related splicing factors (for review see Ref. 51). Two recent reports identified CF  $I_m$  as a component of purified spliceosomes (52, 53). SRp20 and 9G8 that were identified in our screen belong to the family of SR proteins that function in the constitutive splicing reaction and also as alternative splicing regulators (25, 54). Moreover, SRp20 has been shown to mediate the recognition of an alternative 3'-terminal exon by affecting the efficiency of CstF binding to the pre-mRNA (55).

hTra2 $\beta$  is one of the two homologues of the alternative splicing regulator of the *D. melanogaster* sex determination cascade. It binds to purine-rich exonic splicing enhancers (56) and was reported to promote exon 7 inclusion of the "survival of motor neuron 2" mRNA *in vivo* (57). Therefore, our results suggest that CF  $I_m68$  through the interaction with a specific subset of SR proteins could participate in the definition of the last exon and in the choice between alternative 3'-terminal exons. It is interesting to note that both CF  $I_m68$  and CF  $I_m25$  but not the 59-kDa polypeptide have been found in purified spliceosomes (52, 53). So far, we have been unable to demonstrate an interaction of CF  $I_m59$  with SR proteins, albeit this CF  $I_m$  subunit also possesses an RS-like C-terminal domain.<sup>2</sup>

The large subunit of CF  $I_m$  has a modular domain organization consisting of an N-terminal RRM, a proline-rich central region, and a C-terminal RS-like domain. Our results suggest

that each of these domains contributes to the correct intracellular distribution of the protein. In particular, the RS-like domain is a major determinant for CF  $I_m68$  nuclear localization since fusion of this domain to GFP results in exclusive nuclear distribution. In contrast, the RRM does not appear to contain any nuclear targeting signal. Fusion of this region to GFP results in a diffuse distribution throughout the cell, similar to that of GFP alone. However, when the RRM is fused to the RS-like domain, the resulting polypeptide is predominantly nuclear but cytoplasmic as well. We do not know at present whether the nuclear and cytoplasmic distributions observed with this mutant protein represent incomplete nuclear import and/or incomplete retention of the protein in the nucleus. The cytoplasmic and nuclear distributions of the RRM/RS mutant could be explained by the absence of the proline-rich region that may contain a weak nuclear localization signal that is required for complete nuclear targeting. This explanation is consistent with the observation that a mutant lacking the entire RS-like domain is still partially nuclear (results not shown). Alternatively, the central proline-rich part of CF  $I_m68$  could contain a nuclear retention signal (NRS). The first NRS has been identified in human RNP C1, a non-shuttling human RNP that contains a proline-rich region, clusters of basic residues, potential phosphorylation sites for casein kinase II and protein kinase C, and a potential glycosylation site (58). Recently, NRSs have been mapped in two non-shuttling SR proteins, SC35 and SRp40 (33). The only apparent similarity in the primary sequence between the NRS regions of SC35 and human RNP C1 is within the proline-rich region. It remains to be established whether CF  $I_m68$  contains an NRS and if it is a shuttling protein.

Immunofluorescence localization studies showed that SR proteins are organized in the interphase nucleus in a characteristic speckled pattern and that their RS domains are required for targeting proteins to these structures (22, 59, 60). The presence of an RS-like domain in CF  $I_m68$  prompted us to test whether this factor localizes to splicing speckles. Localization of CF  $I_m$  subunits, both with specific antibodies and as GFP fusions, showed that CF  $I_m$  is distributed throughout the nucleoplasm excluding the nucleoli and, in addition, concentrates in a few discrete foci. Double-labeling experiments with specific antibodies directed against marker proteins of known subnuclear compartments revealed that CF  $I_m$  foci are often located in close proximity to both Cajal bodies and splicing speckles that do not correspond to cleavage bodies. This distribution is very similar to the behavior of the PSP1, a recently described protein of unknown function (17). This protein was first identified by mass spectrometry as a nucleolar component. However, localization studies detected it in a novel type of subnuclear bodies termed paraspeckles because they are often found close to splicing speckles (17).

Only three other proteins have been reported to localize in paraspeckles in addition to PSP1. They are PSP2, p54nrb, and the polypyrimidine tract-binding protein-associated splicing factor (PSF). PSF was first identified as a factor that associates with polypyrimidine tract-binding protein and was shown to be required at an early step in spliceosome assembly (61). PSF was also identified as a component of a small nuclear RNP-free complex (SF-A) that contains U1 small nuclear RNP-specific protein A and is implicated in 3'-end processing (27). PSF and p54nrb share considerable sequence homology and have been shown to bind the C-terminal domain of the largest subunit of RNA polymerase II (32). Therefore, PSF and p54nrb might participate in coupling transcription to pre-mRNA splicing. The identification of CF  $I_m68$  as an additional component of

<sup>2</sup> S. Dettwiler, C. Aringhieri, S. Cardinale, W. Keller, and S. M. L. Barabino, unpublished observations.

paraspeckles suggests a possible function for these structures in the transcription and processing of pre-mRNAs.

**Acknowledgments**—We thank J. F. Caceres, G. Martin, U. Kühn, and A. Fox for the generous gift of plasmids and antisera. We are grateful to M. Carmo-Fonseca, M. O'Connell, and G. Martin for critical comments on the paper. We also thank M. Urbano, A. Stotz, V. Widmer, and D. Blank for excellent technical support.

## REFERENCES

- Wahle, E., and Rügsegger, U. (1999) *FEMS Microbiol. Rev.* **23**, 277–295
- Colgan, D. F., and Manley, J. L. (1997) *Genes Dev.* **11**, 2755–2766
- Jenny, A., Hauri, H.-P., and Keller, W. (1994) *Mol. Cell. Biol.* **14**, 8183–8190
- MacDonald, C. C., Wilusz, J., and Shenk, T. (1994) *Mol. Cell. Biol.* **14**, 6647–6654
- Takagaki, Y., and Manley, J. L. (2000) *Mol. Cell. Biol.* **20**, 1515–1525
- Bienroth, S., Keller, W., and Wahle, E. (1993) *EMBO J.* **12**, 585–594
- Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I., and Gilmartin, G. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11893–11898
- Hirose, Y., and Manley, J. L. (1998) *Nature* **395**, 93–96
- Millevoi, S., Geraghty, F., Idowu, B., Tam, J. L., Antoniou, M., and Vagner, S. (2002) *EMBO Rep.* **3**, 869–874
- Rügsegger, U., Beyer, K., and Keller, W. (1996) *J. Biol. Chem.* **271**, 6107–6113
- Rügsegger, U., Blank, D., and Keller, W. (1998) *Mol. Cell* **1**, 243–253
- Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996) *J. Biol. Chem.* **271**, 25059–25062
- Smith, C. W., and Valcarcel, J. (2000) *Trends Biochem. Sci.* **25**, 381–388
- Manley, J. L., and Tacke, R. (1996) *Genes Dev.* **10**, 1569–1579
- Brown, K. M., and Gilmartin, G. M. (2003) *Mol. Cell* **12**, 1467–1476
- Kim, H., and Lee, Y. (2001) *Biochem. Biophys. Res. Commun.* **289**, 513–518
- Fox, A. H., Lam, Y. W., Leung, A. K., Lyon, C. E., Andersen, J., Mann, M., and Lamond, A. I. (2002) *Curr. Biol.* **12**, 13–25
- Chevray, P. M., and Nathans, D. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5789–5793
- Dichtl, B., and Keller, W. (2001) *EMBO J.* **20**, 3197–3209
- Barabino, S. M., Hübner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997) *Genes Dev.* **11**, 1703–1716
- Wilson, A., Peterson, M., and Herr, W. (1995) *Genes Dev.* **9**, 2445–2458
- Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., and Krainer, A. R. (1997) *J. Cell Biol.* **138**, 225–238
- Segade, F., Hurlle, B., Claudio, E., Ramos, S., and Lazo, P. S. (1996) *FEBS Lett.* **387**, 152–156
- Beil, B., Screaton, G., and Stamm, S. (1997) *DNA Cell Biol.* **16**, 679–690
- Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992) *Genes Dev.* **6**, 837–847
- Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R., and Stevenin, J. (1994) *EMBO J.* **13**, 2639–2649
- Lutz, C. S., Cooke, C., O'Connor, J. P., Kobayashi, R., and Alwine, J. C. (1998) *RNA* **4**, 1493–1499
- Schul, W., Groenhout, B., Koberna, K., Takagaki, Y., Jenny, A., Manders, E. M., Raska, I., van Driel, R., and de Jong, L. (1996) *EMBO J.* **15**, 2883–2892
- Schul, W., van Driel, R., and de Jong, L. (1998) *Exp. Cell Res.* **238**, 1–12
- Krause, S., Fakan, S., Weis, K., and Wahle, E. (1994) *Exp. Cell Res.* **214**, 75–82
- Lamond, A. I., and Spector, D. L. (2003) *Nat. Rev. Mol. Cell. Biol.* **4**, 605–612
- Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P. W., Kobayashi, R., Blencowe, B. J., and Ingles, C. J. (2002) *RNA* **8**, 1102–1111
- Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A. R., and Caceres, J. F. (2002) *Mol. Cell. Biol.* **22**, 6871–6882
- Carmo-Fonseca, M., Ferreira, J., and Lamond, A. I. (1993) *J. Cell Biol.* **120**, 841–852
- Carmo-Fonseca, M., Tollervey, D., Pepperkok, R., Barabino, S. M., Merdes, A., Brunner, C., Zamore, P. D., Green, M. R., Hurt, E., and Lamond, A. I. (1991) *EMBO J.* **10**, 195–206
- Spector, D. L., Fu, X. D., and Maniatis, T. (1991) *EMBO J.* **10**, 3467–3481
- Nakai, K., and Horton, P. (1999) *Trends Biochem. Sci.* **24**, 34–36
- Price, S. R., Evans, P. R., and Nagai, K. (1998) *Nature* **394**, 645–650
- Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003) *Nat. Struct. Biol.* **10**, 433–439
- Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) *Cell* **106**, 595–605
- Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003) *Mol. Cell* **11**, 965–976
- Niwa, M., Rose, S. D., and Berget, S. M. (1990) *Genes Dev.* **4**, 1552–1559
- Niwa, M., and Berget, S. M. (1991) *Genes Dev.* **5**, 2086–2095
- Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., and Mattaj, I. W. (1994) *Cell* **76**, 531–541
- Lutz, C. S., Murthy, K. G., Schek, N., O'Connor, J. P., Manley, J. L., and Alwine, J. C. (1996) *Genes Dev.* **10**, 325–337
- Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I. W. (1997) *Genes Dev.* **11**, 761–773
- Vagner, S., Vagner, C., and Mattaj, I. W. (2000) *Genes Dev.* **14**, 403–413
- Awasthi, S., and Alwine, J. C. (2003) *RNA* **9**, 1400–1409
- McCracken, S., Lambermon, M., and Blencowe, B. J. (2002) *Mol. Cell. Biol.* **22**, 148–160
- Rosonina, E., Bakowski, M. A., McCracken, S., and Blencowe, B. J. (2003) *J. Biol. Chem.* **278**, 43034–43040
- Caceres, J. F., and Kornblihtt, A. R. (2002) *Trends Genet.* **18**, 186–193
- Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12203–12207
- Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) *Genome Res.* **12**, 1231–1245
- Zahler, A. M., Neugebauer, K. M., Lane, W. S., and Roth, M. B. (1993) *Science* **260**, 219–222
- Lou, H., Neugebauer, K. M., Gagel, R. F., and Berget, S. M. (1998) *Mol. Cell. Biol.* **18**, 4977–4985
- Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998) *Cell* **93**, 139–148
- Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J., and Wirth, B. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 9618–9623
- Nakielnny, S., and Dreyfuss, G. (1996) *J. Cell Biol.* **134**, 1365–1373
- Hedley, M. L., Amrein, H., and Maniatis, T. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 11524–11528
- Li, H., and Bingham, P. M. (1991) *Cell* **67**, 335–342
- Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993) *Genes Dev.* **7**, 393–406