

Domain organization of human cleavage factor I_m

Distinct sequence motifs within the 68 kDa subunit of cleavage factor I_m mediate RNA binding, protein-protein interactions and subcellular localization

Sabine Dettwiler¹, Chiara Aringhieri^{2#}, Stefano Cardinale^{2#}, Walter Keller¹ and Silvia M.L. Barabino²

¹Department of Cell Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

²Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, 2, I-20126 Milano, Italy

[#]These authors contributed equally to the work

Running title: Domain organization of human cleavage factor I_m

Corresponding author: Silvia M.L. Barabino, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, 2, I-20126 Milano, Italy; Phone: +39-02-6448 3352; Fax: +39-02-6448 3569; E-mail: silvia.barabino@unimib.it

Summary

Cleavage factor I_m (CFI_m) is required for the first step in pre-mRNA 3' end processing and can be reconstituted *in vitro* from its heterologously expressed 25 kDa and 68 kDa subunits. Binding of CFI_m 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 CFI_m involved in RNA-binding, protein-protein interactions, and subcellular localization. CFI_m68 has a modular domain organization consisting of an N-terminal RNA-recognition motif (RRM) and a C-terminal alternating charge domain. However, the RRM of CFI_m68 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 CFI_m68/25 heterodimer. Whereas multiple protein interactions with other 3' end processing factors are detected with CFI_m25, CFI_m68 interacts with SRp20, 9G8, and hTra2 β , members of the SR family of splicing factors, via its C-terminal alternating charge domain. This domain is also required for targeting CFI_m68 to the nucleus. However, CFI_m68 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.

Introduction

Eukaryotic messenger RNA precursors (pre-mRNAs) are synthesized and processed in the nucleus prior to their export to the cytoplasm, where they serve as templates for protein synthesis. Transcription is coupled spatially and temporally to 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 (1,2)). 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 CPSF¹ (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 (CFI_m), cleavage factor II_m (CFII_m) 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 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 other proteins involved in either transcription, such as the carboxyl terminal domain (CTD) 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 kDa, 59 kDa and 68 kDa and one minor polypeptide of 72 kDa copurify with CFI_m activity from HeLa cell nuclear extract (10). Reconstitution of CFI_m activity with recombinant proteins suggests that CFI_m is a heterodimer consisting of the 25 kDa subunit and one of the larger polypeptides (11). All the three larger proteins appear to be highly related in their amino acid sequence. Moreover, all CFI_m 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

Domain organization of human cleavage factor I_m

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 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 (14)). Previous experiments have shown that preincubation of the RNA substrate with CF_m reduces the lag-phase of the cleavage reaction (11). This observation suggests that binding of this factor to the pre-mRNA may be an early step in the assembly of the 3' splice site processing complex, such that CF_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'UTR of the pre-mRNA of its 68 kDa subunit to which CF_m preferentially binds (15).

In this report we describe the analysis of the functional domains of the 25 kDa and 68 kDa subunits of CF_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_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_m25 binds to RNA. Surprisingly, despite having an RRM, CF_m68 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_m68 identified members of the SR family of splicing factors, suggesting that CF_m may contribute to the coordination of splicing and 3' splice site formation. Furthermore, we investigated the subcellular localization of CF_m. We found that CF_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.

Domain organization of human cleavage factor I_m

Experimental Procedures

Oligonucleotides

Oligonucleotides were synthesized either by Microsynth, or MWG. The sequences were as follows: GST-25: 1-80 sense: gggattac**catatg**tctgtgtaccgccaatc; 1-80 antisense: gc**gggatcctc**atacagtctctctcattc; 81-160 sense: gggattac**catatg**gaaggggtctgattg; 81-160 antisense: gc**gggatcctc**aatatggatactgaggag; 161-227 sense: gggattac**catatg**attcctgcacatattac; 161-227 antisense: gc**gggatcctc**agttgtaaataaaattg; 5'U2AF65: acgcggatccatgtcggacttcgacgag; 3'U2AF65: atccgctcgagctaccagaagtcccggcgg; 5'U2AF35: acgcggatccatggcggagtatctggc; 3'U2AF35: atccgctcgagtcagaatcgccagatc; GST-68: 5'RS: gaacgcggatccctgcaagaacgccattgag; 3'RS: atccgctcgagcttctaacgatgacgatattc; 5'RRM: gatccccgggaatggcggacggcgtgga; 3'RRM: ctcgggatcccaggaaatctgtcccacc; 3'Pro: ctcgggatcccactcaatggcgttcttgc; BclI (5'end): tgggcatgatcagataga; Bsu36I (3'end): gacctaagggtggacg; RNP2 sense: attgcattatatatt**gtagt**tctaactggtggac; RNP2 antisense: gtccaccatgttaga**act**acaatatataatgcaa; \square RS sense: aacagt**ggcatg**cctacatc; \square RS antisense: cacc**aaagctt**ctactcaattccatgaaggca; ythSRs: cacc**ggaattc**gagtccaagtcttatgg; ythSRas: taacc**gctcgag**ctaacgatgacgatattc; 5'-TraX: agt**ctaga**agcgacagcggcgagca; 3'-TraB: ac**ggatc**cttaatagcgacgaggtgagt; 5'-9G8X: agt**ctagat**cgcgttacgggcggga; 3'-9G8B: ac**ggatc**ctcagtcattctttcaggac;

Yeast two-hybrid screen

Yeast two-hybrid library screening and analysis was 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 \square _m68 (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 \square -galactosidase (*lacZ*) and histidine (*HIS3*). The ability to express the reporter *HIS3* gene was tested on SC plates lacking leucine, tryptophane, and histine but containing 25 mM

Domain organization of human cleavage factor I_m

3-aminotriazole. β -galactosidase activity was detected by filter lift assays (18). Library plasmids from positive colonies were recovered in *E. 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_m25 was inserted into the NcoI and XhoI sites. To express fragments of the 25 kDa subunit of CF_m, the ORF 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 ORFs 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_m68 mutants RS, RRM, RRM/Pro, Pro/RS were generated by PCR with the respective oligonucleotides, cloned into pGemTeasy (Promega) and sequenced by C.R.B.I., (Università di Padova, Italy). 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 IPTG. 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 (Pharmacia) and purified according to the manufacturer's instructions.

Domain organization of human cleavage factor I_m

Expression of wild type CF_m68 in baculovirus was achieved with the BAC-To-BAC baculovirus expression system (Life Technologies). The entire coding region together with 162 nt of the 3'UTR 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 formation of the recombinant bacmids. High molecular weight DNA was then prepared from selected *E. coli* clones and 1-2 μ g were used to transfect 1×10^6 *Spodoptera frugiperda* (Sf9) insect cells. After 2-3 days the supernatant was recovered and approximately 1 ml of the virus was used to infect 2×10^6 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 1×10^6 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 and 0.02 % NP-40, 0.5 mM PMSF, 0.7 μ g/ml Pepstatin, 0.4 μ g/ml Leupeptin). The lysate was centrifuged after incubation on ice for 30 minutes and proteins were purified by Ni²⁺-NTA affinity chromatography.

GST pull-down experiments

0.5-1 μ g of purified GST-25 protein was incubated with approximately 100 μ g of either HeLa cell nuclear extract or whole cell extracts prepared from transfected HeLa or HEK293 cells for 1 hour at 4°C. 50 μ l glutathione-Sepharose beads were then added and incubated for additional 2 hours. The resin was washed three times with 1 ml IPP150. Proteins were eluted with 2 \times Laemmli buffer and separated by 12% SDS-PAGE. Proteins were then blotted onto nitrocellulose and the membrane probed with anti-CF_m68 polyclonal antibodies or anti-T7 monoclonal antibody (Novagen).

³⁵S-methionine (Amersham) was used to label proteins in a transcription/translation system (Promega). Purified GST or GST fusion proteins (ca. 500 ng) were incubated with *in vitro* translated proteins in 30 μ l total reaction volume. Volumes were adjusted with GST-buffer (75 μ M KCl, 50 μ M Tris-HCl, pH 7.9, 10% glycerol, 10 μ M reduced Glutathione, 0.01 % NP-40, 1 mM DTT and 50 μ g/ml BSA). After incubation for 1 hr at

Domain organization of human cleavage factor I_m

4°C, 25 μ l of glutathione-Sepharose beads were added and the volume was adjusted to 500 μ l with 1 \times PBS/0.01% NP40. Incubation was continued for 1 hour, 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.

CFI_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 ORF of the 25 kDa subunit was cloned into BamHI and EcoRI, respectively (pBS25). The ORFs of CFI_m59 and CFI_m68 were inserted into the EcoRI site of bluescript (Stratagene). The CFI_m68 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 Δ SR mutant was generated by PCR with pBS68 and primers that contained the SphI (Δ RS sense) and HindIII (Δ 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 (19). Approximately 500 ng GST-fusion protein were incubated with 30-100 fmol labeled RNA under cleavage conditions (20) for 30 minutes at 30°C. GST-Sepharose was added and incubated for one hour at 4°C. The beads were washed extensively with 1 \times PBS/0.01% NP-40 and resuspended in Proteinase K mix. After 30 minutes, 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 μ l reactions were set up on ice as follows: To 8 μ l premix (2 \times mM DTT, 0.01% NP-40, 20 \times mM Creatine Phosphate, 2% PEG, 0.4 \times mM Cordycepin, 5 \times μ l RNA guard, 0.025 μ g/ μ l creatine kinase, 1.5 \times mM MgCl₂) either 1 μ mol of CFI_m25/68, CFI_m68, 5 μ mol of CFI_m25 or 4 μ mol of purified CFI_m were incubated with 200 fmol labeled L3 pre-mRNA. These proteins had been expressed in *E. coli* and assembled *in vitro* as described in (11). 75 mM potassium acetate and

Domain organization of human cleavage factor I_m

200 μmol 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 × 250 mJ, UV Stratalinker 1800, Stratagene) in a microtiter plate at room temperature. Samples were treated with 5 μg/μl RNase A (Sigma) for 1 hour at 37°C and proteins were separated by SDS-PAGE and exposed on X-ray film.

Cell culture and transfections

HeLa and HEK 293 cells were grown in DMEM supplemented with 10% FCS and transfected with 1 μg of plasmid DNA per 60 mm dish containing approximately 60% confluent cells with the calcium phosphate method. The epitope-tagged 9G8 and hTra2 expression plasmids were constructed by amplifying the cDNAs isolated in the yeast two-hybrid screen with specific primers and subcloning of the resulting PCR products as a XbaI-BamHI fragment into pCGTHCF_{FL}T7 expression vector (21). Amplification was performed with Amplitaq Gold (Roche) 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 cover slips. 24 hours after transfection cells were washed with PBS, fixed with 3% paraformaldehyde and permeabilized with 0.2% Triton for 5 min on ice. Cells were incubated 1 hour at 37°C with the primary antibody, washed three times with PBS containing 2% BSA and again incubated for 1 hour at 37°C with the appropriate secondary antibody.

For localization of endogenous CF_m polyclonal antibodies raised against CF_m68 and CF_m25 (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); endogenous CPSF was detected with a mouse monoclonal antibody (3). Secondary

Domain organization of human cleavage factor I_m

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_m68 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 60X, 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 63X, 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_m68 interacts with the 25 kDa protein via its RRM

Four polypeptides with apparent molecular weights of 25 kDa, 59 kDa, 68 kDa and 72 kDa copurify with CF_m activity. Previous studies have shown that the 25 kDa and 68 kDa proteins are sufficient to reconstitute CF_m activity *in vitro*. It has been suggested that CF_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). In order to verify that the larger CF_m subunits associate with the 25 kDa protein, we expressed the 25 kDa protein as a glutathione S-transferase (GST) fusion (GST-25) in *E. coli* and incubated it with HeLa cell nuclear extract. Western blot analysis with anti-CF_m68 antibody detected all three larger CF_m subunits (Fig. 1A, lane 3). This result was confirmed by incubating GST-25 with CF_m68 and CF_m59 that had been transcribed and translated *in vitro* in the presence of ³⁵S-methionine (Fig. 1B, lanes 7 and 8). In addition, CF_m25 appeared to interact weakly with itself (Fig. 1B, lane 9), which might indicate that heterodimers bind to each other to form larger CF_m-complexes. The experiments were carried out in the presence of Rnase A and the interactions were thus 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_m68. For this purpose, 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_m68 only detected association with the full-length GST-25 (Fig. 1C, lane 3). None of the deletion constructs was able to interact with CF_m68, suggesting that intact 25 kDa protein is required for stable binding of the two proteins.

CF_m25 interacts with PAP and PABPN1

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_m with the substrate leads to a faster assembly of the 3' end processing complex (11). CF_m may

Domain organization of human cleavage factor I_m

help to recruit CPSF and CstF to the pre-mRNA and we tested in GST pull-down experiments for the direct interaction of CF_m with CPSF and CstF. Our results with GST-25, however, 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_m, this does not appear to be mediated by a direct interaction with the smallest subunit of CF_m.

An interaction between CF_m25 and PAP has been described before (16). The authors isolated the 25 kDa subunit of CF_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 deletion of the C-terminal 226 residues of PAP did not affect its binding to GST-25 (PAP Δ C, Fig. 2A, 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_m25 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_m25 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, lane 4 and 5, respectively) emphasizing that mainly the region between amino acids 81-160 of CF_m25 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_m25 or that their interaction domains are very close.

The N-terminal RRM of CF_m68 is sufficient for the interaction with CF_m25

To identify functional domains of the 68 kDa subunit of CF_m we took a similar approach to that used for the characterization of CF_m25. Several CF_m68 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

Domain organization of human cleavage factor I_m

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 Π_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 Π_m 68. As shown in Fig. 3C, both wild type CF Π_m 68 or 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 Π_m 68 (G87V, N88V) abolished interaction with the 25 kDa subunit (Fig. 3C, lane 9). These results suggest that the interaction between CF Π_m 68 and CF Π_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 Π_m

Whereas the 68 kDa subunit of CF Π_m contains an RRM, the 25 kDa polypeptide does not display homology to any known RNA binding domain. Nevertheless, it was shown previously that CF Π_m subunits purified from HeLa cell nuclear extract can be UV cross-linked to an RNA substrate (10). We analyzed the interaction of CF Π_m subunits with RNA in more detail. In a first attempt we performed UV cross-linking experiments with recombinant CF Π_m 68 purified from baculovirus-infected Sf9 cells and recombinant CF Π_m 25 produced in *E. coli*. Similar amounts of proteins were exposed to UV light in the presence of ³²P-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 kDa and 68 kDa subunits of CF Π_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 Π_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

Domain organization of human cleavage factor I_m

the 25 kDa and the 68 kDa proteins binds RNA more efficiently than each subunit separately.

To determine which region of the 25 kDa subunit of CF_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. Briefly, the different GST-fragments were incubated under cleavage conditions with the labeled substrate and RNA-protein complexes were subsequently immobilized on GST-Sepharose. Figure 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 likely reflects an ionic interaction between the positively charged RS-like domain and the RNA.

Synergistic RNA binding of CF_m25 and CF_m68 was confirmed by RNA pull-down assays with the same GST-68 proteins used before and in the presence of increasing amounts of recombinant histidine-tagged 25 kDa subunit (His-CF_m25, Fig. 4D). 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. Addition of CF_m25 to the RS-like domain of CF_m68 did not increase the amount of precipitated RNA (Fig. 4D, lanes 8-11), most likely 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 CF_m-binding sites on the pre-mRNA.

The RS-like domain of CF I_m68 associates with SR proteins

The C-terminal portion of CF I_m68 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_m68, a yeast two hybrid screen was performed with a fragment encoding amino acids 481 to 551 (called 68RS) as bait. A HeLa cell cDNA library was screened and analysis of positive colonies revealed among the strongest interactors some known members of the SR protein family, namely hTra-2 β (23,24), SRp20 (also named X16 (25), and 9G8 (26).

In order 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 β , SRp20, and 9G8 (Fig. 5A). All constructs encoded proteins with a bacteriophage T7 gene 10 (T7) epitope tag at their N-termini allowing detection of the exogenous proteins with a monoclonal antibody that recognizes this epitope (see Experimental Procedures, (27)). Western blot analysis revealed that a significant fraction of hTra-2 β , SRp20 and 9G8 was found to associate with the RS-like domain of CF I_m68 (Fig. 5, lane 3 of each panel). However, none of those SR proteins was detected if incubated with GST (lane 1), GST-25 (lane2), GST-RRM68 (lane4) or GST-U2AF65 (lane 5) indicating that the interaction is specific for the RS-like domain of CF I_m68. The interaction with hTra-2 β , 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_m68 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 punctuate 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

Domain organization of human cleavage factor I_m

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 (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 (32,33).

Because of the similarity of CF_m large subunits and SR proteins, we wanted to determine their subcellular localization. Immunofluorescence microscopy with polyclonal antisera raised against the 25 kDa and the 68 kDa CF_m subunits (11) revealed that endogenous CF_m is localized in the nucleoplasm with exclusion of the nucleoli, with a granular distribution and concentration in few discrete foci (Fig. 6c, and 6c'). To confirm the staining pattern observed with the antibodies, we transiently expressed both CF_m subunits fused to the green fluorescent protein (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 hr 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 (Figures 6b, and 6b'). In order 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_m68 do not correspond to cleavage bodies visualized by anti-CPSF100. This unexpected result prompted us to assay markers for known subnuclear compartments in order to clarify the nature of the CF_m foci. We used antibodies specific for p80 coilin to detect Cajal bodies (34) and for SC-35 to detect splicing speckles (35,36). GFP-68 was however neither localized to Cajal bodies (Fig. 7B, panels a-a'') nor to speckles (Fig. 7B, panels b-b''), but always localized in close proximity to both of them. The observation that GFP-68 foci were adjacent to speckles prompted us to co-localize 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 were therefore termed paraspeckles. Figure 7B (panels c-c'') shows that GFP-68 and PSP1 at least partially co-localize in paraspeckles.

Role of individual domains of CF_m68 in cellular distribution and subnuclear localization

To determine the role of individual domains of CF_m68 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 the proteins by indirect immunofluorescence microscopy (Fig. 8). We verified the expression of all transfected cDNAs by Western blot analysis of whole-cell lysates (results not shown).

While 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 (68_N, Fig. 8, 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 deletion mutants lacking the C-terminal RS-like domain localized to the cytoplasm (68_{RS46}, Fig. 8 panel c'; RRM and RRM/Pro, not shown). Interestingly, a domain search with PSORT (37) reveals the presence of a putative NLS (RRHR) located 46 amino acids upstream of the stop codon. Consistent with this prediction, deletion of this portion of the charged domain was sufficient to prevent nuclear import of the truncated protein (68_{RS46}, Fig. 8 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 (RRM/RS, Fig. 8 panel d'). This protein was found both in the cytoplasm and in the nucleus. Furthermore, fusion of the RS domain to GFP is sufficient to direct nuclear import of the chimeric protein (RS, Fig. 8 panel e'). These results indicate that the C-terminal charged domain may contain one or more nuclear localization signals.

DISCUSSION

Mammalian cleavage factor I (CFI_m) is a component of the basic pre-mRNA 3' end processing complex. Reconstitution experiments suggest that CFI_m occurs as several isomers consisting of a heterodimer of a 25 kDa polypeptide and one of the three larger subunits of 59 kDa, 68 kDa and 72 kDa (11). In this work we have characterized the functional domains of the 25 kDa 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 CFI_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 CFI_m68/25 heterodimer that identified specific binding sites on the RNA (15). RNA-binding of CFI_m25 per se does not require the C-terminus of the protein.

The primary sequence of CFI_m68 reveals the presence of two known sequence motifs: an N-terminal RNA-recognition motif of the RNP-type (RRM) 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 CFI_m68 is primarily engaged in protein-protein interaction with the 25 kDa subunit of CFI_m. Although RRMs are generally considered nucleic acid binding domains, several reports have implicated them in protein-protein interactions. For example, the U2 snRNP-specific protein U2B^{''} was found to interact with U2A' via its RRM (38). Recently, the crystal structure of the complex between the *D. melanogaster* Y14 and Mago proteins was determined (39). Y14 and Mago associate with spliced mRNAs and are components of the exon junction complex. Like CFI_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 U2AF65 (40). In addition, U2AF65 was found to bind to the human splicing

Domain organization of human cleavage factor I_m

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 snRNP-binding, we suspect that the 25 kDa protein and one of the larger subunits of CF_m cooperate to recognize CF_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_m68 provides new insight into the coordination of splicing and 3' end formation. Recognition of 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 (51), it has been suggested that CF_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 51). Two recent reports identified CF_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 α 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 *survival of motor neuron 2* mRNA *in vivo* (57). Therefore, our results suggest that CF_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_m68 and CF_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_m59 with SR proteins, albeit this CF_m subunit also possesses an RS-like C-terminal domain (our unpublished observations).

The large subunit of CF_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

Domain organization of human cleavage factor I_m

the protein. In particular, the RS-like domain is a major determinant for CF_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 distribution observed with this mutant protein represents incomplete nuclear import and/or incomplete retention of the protein in the nucleus. The cytoplasmic and nuclear distribution of the RRM/RS mutant could be explained by the absence of the proline-rich region that may contain a weak NLS 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_m68 could contain a nuclear retention signal (NRS). The first NRS has been identified in hnRNP C1, a non-shuttling hnRNP 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 primary sequence between the NRS regions of SC35 and hnRNP C1 is within the proline-rich region. It remains to be established whether CF_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 (32,59,60). The presence of an RS-like domain in CF_m68 prompted us to test whether this factor localizes to splicing speckles. Localization of CF_m subunits, both with specific antibodies and as GFP-fusions, showed that CF_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_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

Domain organization of human cleavage factor I_m

behavior of the PSP1, a recently described protein of unknown function (17). This protein was first identified by mass spectrometry as a nucleolar component. Localization studies, however, 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: PSP2, p54nrb, and the polypyrimidine tract binding protein- (PTB) associated splicing factor PSF. PSF was first identified as a factor that associates with PTB and was shown to be required at an early step in spliceosome assembly (61). PSF was also identified as a component of a snRNP-free complex (SF-A) that contains U1 snRNP-specific protein A and is implicated in 3' end processing (62). PSF and p54nrb share considerable sequence homology and have been shown to bind the CTD of the largest subunit of RNA polymerase II (63). Therefore, PSF and p54nrb might participate in coupling transcription to pre-mRNA splicing. The identification of CF_I_m68 as an additional component of paraspeckles suggests a possible function for these structures in 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 manuscript. We also thank M. Urbano, A. Stotz, V. Widmer and D. Blank for excellent technical support. This work was supported by the Telethon Onlus grant nr. 0010Y02 and by a grant from A.I.R.C. to S.B. and by the University of Basel, the Swiss National Science Fund and the Louis-Jeantet Foundation for Medicine.

¹The abbreviations used are: CPSF, cleavage and polyadenylation specificity factor; CstF, cleavage stimulation factor; CF_I_m, cleavage factor I; CF_I_m, cleavage factor I_m; PAP, poly(A) polymerase; PABPN1, nuclear poly(A) binding protein; CTD, carboxyl terminal domain of RNA polymerase_{II}; RRM, RNA recognition motifs; 3'UTR, 3' untranslated region; GST, glutathione S-transferase; GFP, green fluorescent protein; PSP1,

Domain organization of human cleavage factor I_m

paraspeckle protein 1; NLS, nuclear localization signal; NRS, nuclear retention signal; PTB, polypyrimidine tract binding protein; PBS, phosphate-buffered saline; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; BSA, bovine serum albumine; ORF, open reading frame; PEG, polyethylene glycol; DMEM, Dulbecco's modified Eagle's medium; FCS, fetal calf serum.

Figure Legends

Figure 1. **The 25 kDa protein interacts with the larger CF_m subunits.**

- A.** The 25 kDa protein can co-precipitate the large subunits from nuclear extract. GST protein (lane 2) or GST-25 (lane 3) was preincubated with HeLa cell nuclear extract as described in 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_m subunits (11). Lane 1: 2 μg of HeLa cell nuclear extract.
- B.** GST-25 interacts directly with CF_m59, CF_m68 and with itself. *In vitro* translated [³⁵S]methionine-labelled 25 kDa, 59 kDa, and 68 kDa polypeptides were incubated with either GST or GST-25 protein as described in 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_m68 requires full-length 25 kDa protein. *In vitro* translated [³⁵S]methionine-labelled CF_m68 subunit was incubated with the indicated purified GST fusion proteins (GST-25, 25/1-80, 25/81-160, 25/161-227, 25/1-160, 25/81-227, or 25/140-168) as described in 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.

Figure 2. **The central region of CF I_m25 is required for the interaction with PAP and PABPN1.**

- A.** GST-25 directly interacts with PAP and PABPN1. GST (lanes 3, 6, 9) or GST-25 (lanes 2, 5, 8) was preincubated with *in vitro* translated [³⁵S]methionine-labelled PAP, PAP_C or PABPN1 proteins as described in 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_m25 are sufficient for interaction with PABPN1. *In vitro* translated [³⁵S]methionine-labelled PABPN1 was incubated with

Domain organization of human cleavage factor I_m

the indicated purified GST fusion proteins (GST-25, 1-160, 81-227, 1-80, 81-160, or 161-227) as described in Experimental Procedures. Bound proteins were analyzed by SDS-PAGE and autoradiography. 10% of the input is shown in lane 1.

Figure 3. **The RRM domain of CF Π_m 68 is required for the interaction with the 25 kDa protein.**

- A. Schematic representation of the CF Π_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_m68 interacts directly with the 25 kDa subunit via the RRM domain. *In vitro* translated [³⁵S]methionine-labelled CF Π_m 25 subunit was incubated with the purified GST fusion proteins indicated (68RRM, 68RS, 68RRM/RS, or 68Pro/RS) as described in 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 Π_m 68 are essential for the interaction with CF Π_m 25. GST (lanes 2, 5, 8) or GST-25 (lanes 3, 6, 9) was preincubated with the indicated *in vitro* translated [³⁵S] methionine-labeled proteins (CF Π_m 68, 68 \square RS, 68RNP2) as described in 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.

Figure 4. **The small and the large subunit of CF Π_m cooperate in RNA binding.**

Domain organization of human cleavage factor I_m

- A.** CF_m subunits can be UV cross-linked to RNA. ³²P-labelled Adenovirus L3 pre-mRNA was incubated for 10 min with the indicated proteins (pCF_m: purified CF_m and rCF_m: recombinant CF_m) and exposed to UV light as described in Experimental Procedures. After digestion with RNaseA, the products were analyzed by 12% SDS-PAGE and autoradiography.
- B.** The N-terminal region of CF_m25 is required for RNA-binding. RNA pull-down experiment with GST fusions of CF_m25 deletion mutants. ³²P-labelled SV40 pre-mRNA was incubated with the indicated purified GST fusion proteins as described in Experimental Procedures. Lane 1: GST protein.
- C.** The CF_m68RRM domain does not bind to RNA. RNA pull-down experiment with GST fusions of CF_m68 domain deletion mutants. ³²P-labelled Adenovirus L3 substrate was incubated with the indicated purified GST fusion proteins as described in Experimental Procedures. Lane 1: 10% of the input L3 pre-mRNA; lane 2: GST protein.
- D.** CF_m25 interacts with the RRM of CF_m68 and enhances binding to RNA. RNA pull-down experiment with GST fusions of CF_m68 domain deletion mutants either in the absence (lanes 2, 7, and 12,) or in the presence of increasing amounts of recombinant histidine-tagged 25 kDa subunit (lanes 3-6, 8-11 and 13-16). ³²P-labelled Adenovirus L3 substrate was incubated with the indicated purified GST fusion proteins as described in 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.

Figure 5. **The C-terminal RS-like domain of CF I_m68 interacts with SR proteins.**

HeLa or HEK 293 cells were transfected with pCG constructs expressing hTra2, SRp20, or 9G8. Whole cell extracts were incubated with either GST or the indicated GST fusion proteins (GST-25, GST-68RS, GST-68RRM, GST-U2AF65) as described in Experimental Procedures. Bound proteins were separated by SDS-PAGE and analyzed by Western blotting with an anti-T7 tag monoclonal antibody.

Figure 6. **Nuclear localization of CF_m subunits.**

Domain organization of human cleavage factor I_m

HeLa cells were either immuno-labeled with anti-CF_m68 (panel c) or with anti- CF_m25 (panel c') polyclonal antibodies (11) or transfected with GFP-fusion constructs (panel b: GFP68; panel b': GFP25) as described in Experimental Procedures. Panels a and a': DAPI-staining. The white arrowheads indicate punctuate CF_m foci. Scale bars, 5 μ m.

Figure 7. CF_m68 localizes to paraspeckles.

- A. Double-labeling of HeLa cells with anti-CPSF100 monoclonal antibody (panel a, (3), and with anti-CF_m68 polyclonal antiserum (panel b, (11) as described in Experimental Procedures. The merge of the two channels is shown in panel c (green, anti-CPSF100, red, anti-CF_m68). Small arrows indicate the cleavage bodies that can be detected with the anti-CPSF100 antibody (3). Arrowheads indicate discrete CF_m68 foci that do not co-localize with CPSF cleavage bodies.
- B. CF_m68 partially co-localizes 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 in 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). Splicing speckles are detected with an anti-SC35 monoclonal antibody (b – b", SIGMA). Arrowheads indicate adjacent GFP-68 "foci" (green) and speckles (red). Paraspeckles are detected with an anti-PSP1 polyclonal antiserum (c – c", 16) Arrowheads indicate co-localization of GFP-68 and PSP1 (yellow).

Figure 8. Role of CF_m68 domains in cellular localization and subnuclear distribution.

The CF_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'** GFP; **b'** CF_m68 Δ N; **c'** CF_m68 Δ RS₄₆; **d'** CF_m68RRM/RS; **e'** CF_m68RS. Bar, 5 μ m.

REFERENCES

1. Wahle, E., and Rügsegger, U. (1999) *FEMS Microbiol. Rev.* **23**, 277-295
2. Colgan, D. F., and Manley, J. L. (1997) *Genes Dev.* **11**, 2755-2766
3. Jenny, A., Hauri, H.-P., and Keller, W. (1994) *Mol. Cell. Biol.* **14**, 8183-8190
4. MacDonald, C. C., Wilusz, J., and Shenk, T. (1994) *Mol. Cell. Biol.* **14**, 6647-6654
5. Takagaki, Y., and Manley, J. L. (2000) *Mol. Cell. Biol.* **20**, 1515-1525
6. Bienroth, S., Keller, W., and Wahle, E. (1993) *EMBO J.* **12**, 585-594
7. Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I., and Gilmartin, G. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11893-11898
8. Hirose, Y., and Manley, J. L. (1998) *Nature* **395**, 93-96
9. Millevoi, S., Geraghty, F., Idowu, B., Tam, J. L., Antoniou, M., and Vagner, S. (2002) *EMBO Rep* **3**, 869-874
10. Rügsegger, U., Beyer, K., and Keller, W. (1996) *J. Biol. Chem.* **271**, 6107-6113
11. Rügsegger, U., Blank, D., and Keller, W. (1998) *Mol. Cell* **1**, 243-253
12. Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996) *J. Biol. Chem.* **271**, 25059-25062
13. Smith, C. W., and Valcarcel, J. (2000) *Trends Biochem. Sci.* **25**, 381-388.
14. Manley, J. L., and Tacke, R. (1996) *Genes Dev.* **10**, 1569-1579
15. Brown, K. M., and Gilmartin, G. M. (2003) *Mol. Cell* **12**, 1467-1476
16. Kim, H., and Lee, Y. (2001) *Biochem. Biophys. Res. Commun.* **289**, 513-518
17. 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
18. Chevray, P. M., and Nathans, D. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5789-5793
19. Dichtl, B., and Keller, W. (2001) *EMBO J.* **20**, 3197-3209
20. Barabino, S. M., Hübner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997) *Genes Dev.* **11**, 1703-1716
21. Wilson, A., Peterson, M., and Herr, W. (1995) *Genes Dev.* **9**, 2445-2458
22. Caceres, J. F., Misteli, T., Screatton, G. R., Spector, D. L., and Krainer, A. R. (1997) *J. Cell Biol.* **138**, 225-238
23. Segade, F., Hurle, B., Claudio, E., Ramos, S., and Lazo, P. S. (1996) *FEBS Lett.* **387**, 152-156
24. Beil, B., Screatton, G., and Stamm, S. (1997) *DNA Cell Biol* **16**, 679-690
25. Zahler, A. M., Lane, W. S., Stolk, J. A., and Roth, M. B. (1992) *Genes Dev.* **6**, 837-847
26. Cavaloc, Y., Popielarz, M., Fuchs, J. P., Gattoni, R., and Stevenin, J. (1994) *EMBO J.* **13**, 2639-2649
27. Wilson, A. C., Peterson, M. G., and Herr, W. (1995) *Genes Dev.* **9**, 2445-2458
28. 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
29. Schul, W., van Driel, R., and de Jong, L. (1998) *Exp. Cell Res.* **238**, 1-12
30. Krause, S., Fakan, S., Weis, K., and Wahle, E. (1994) *Exp. Cell Res.* **214**, 75-82
31. Lamond, A. I., and Spector, D. L. (2003) *Nat. Rev. Mol. Cell. Biol.* **4**, 605-612
32. Caceres, J. F., Misteli, T., Screatton, G. R., Spector, D. L., and Krainer, A. R. (1997) *J. Cell. Biol.* **138**, 225-238

Domain organization of human cleavage factor I_m

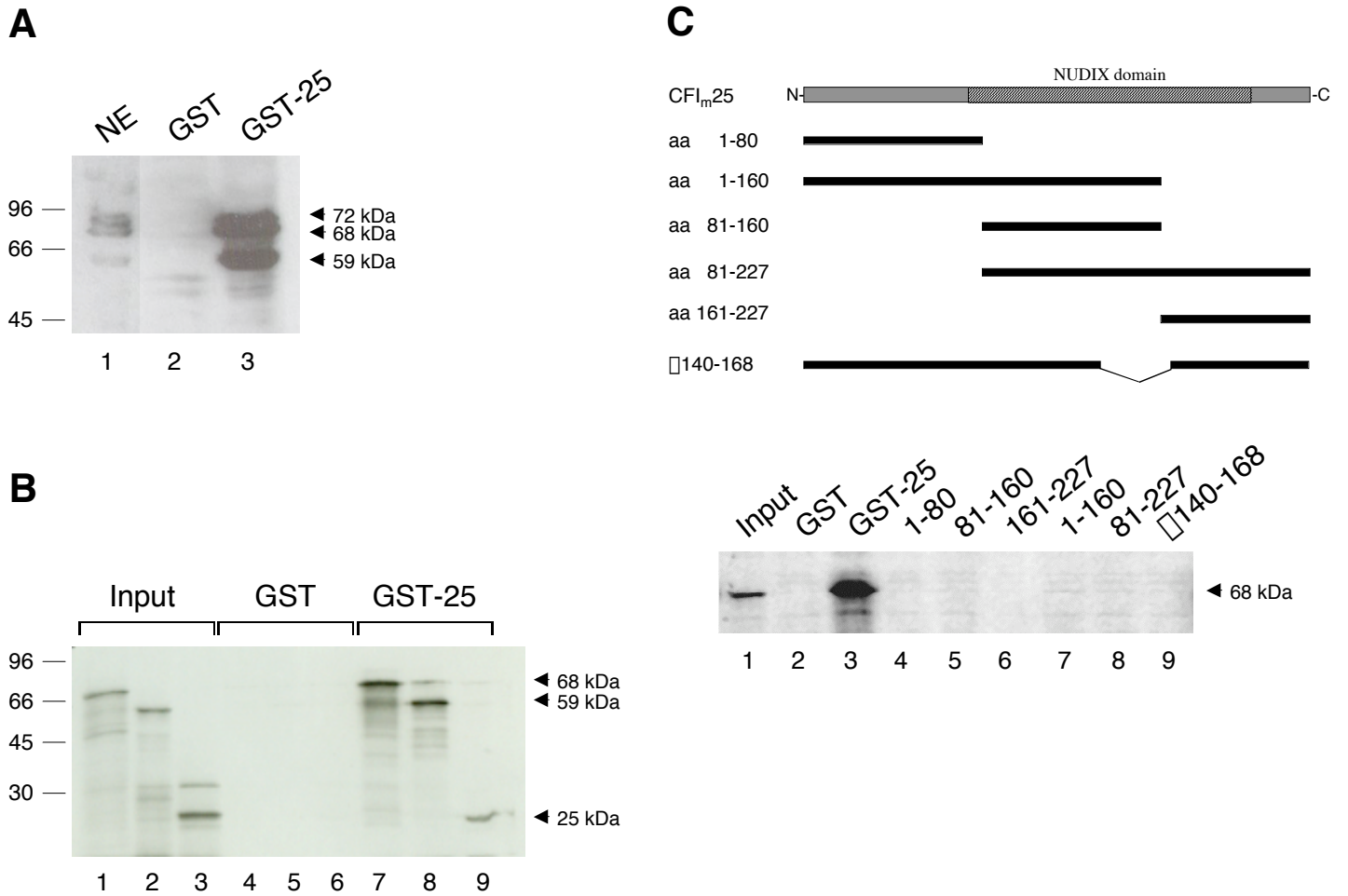
33. Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A. R., and Caceres, J. F. (2002) *Mol. Cell Biol.* **22**, 6871-6882
34. Carmo-Fonseca, M., Ferreira, J., and Lamond, A. I. (1993) *J. Cell Biol.* **120**, 841-852
35. Carmo-Fonseca, M., Tollervey, D., Pepperkok, R., Barabino, S. M., Merdes, A., Brunner, C., Zamore, P. D., Green, M. R., Hurt, E., and Lamond, A. I. (1991) *EMBO J.* **10**, 195-206
36. Spector, D. L., Fu, X. D., and Maniatis, T. (1991) *EMBO J.* **10**, 3467-3481
37. Nakai, K., and Horton, P. (1999) *Trends Biochem. Sci.* **24**, 34-36
38. Price, S. R., Evans, P. R., and Nagai, K. (1998) *Nature* **394**, 645-650
39. Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003) *Nat. Struct. Biol.* **10**, 433-439
40. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) *Cell* **106**, 595-605
41. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003) *Mol. Cell* **11**, 965-976
42. Niwa, M., Rose, S. D., and Berget, S. M. (1990) *Genes Dev.* **4**, 1552-1559
43. Niwa, M., and Berget, S. M. (1991) *Genes Dev.* **5**, 2086-2095
44. Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., and Mattaj, I. W. (1994) *Cell* **76**, 531-541
45. 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
46. Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I. W. (1997) *Genes Dev.* **11**, 761-773
47. Vagner, S., Vagner, C., and Mattaj, I. W. (2000) *Genes Dev.* **14**, 403-413
48. Awasthi, S., and Alwine, J. C. (2003) *RNA* **9**, 1400-1409
49. McCracken, S., Lambermon, M., and Blencowe, B. J. (2002) *Mol Cell Biol* **22**, 148-160
50. Rosonina, E., Bakowski, M. A., McCracken, S., and Blencowe, B. J. (2003) *J Biol Chem* **278**, 43034-43040
51. Caceres, J. F., and Kornblihtt, A. R. (2002) *Trends Genet.* **18**, 186-193
52. Zhou, Z., Sim, J., Griffith, J., and Reed, R. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12203-12207
53. Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002) *Genome Res.* **12**, 1231-1245
54. Zahler, A. M., Neugebauer, K. M., Lane, W. S., and Roth, M. B. (1993) *Science* **260**, 219-222
55. Lou, H., Neugebauer, K. M., Gagel, R. F., and Berget, S. M. (1998) *Mol. Cell Biol.* **18**, 4977-4985
56. Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998) *Cell* **93**, 139-148
57. Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J., and Wirth, B. (2000) *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9618-9623
58. Nakielny, S., and Dreyfuss, G. (1996) *J. Cell Biol.* **134**, 1365-1373
59. Hedley, M. L., Amrein, H., and Maniatis, T. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11524-11528
60. Li, H., and Bingham, P. M. (1991) *Cell* **67**, 335-342

Domain organization of human cleavage factor I_m

61. Patton, J. G., Porro, E. B., Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993) *Genes Dev.* **7**, 393-406
62. Lutz, C. S., Cooke, C., O'Connor, J. P., Kobayashi, R., and Alwine, J. C. (1998) *RNA* **4**, 1493-1499
63. 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

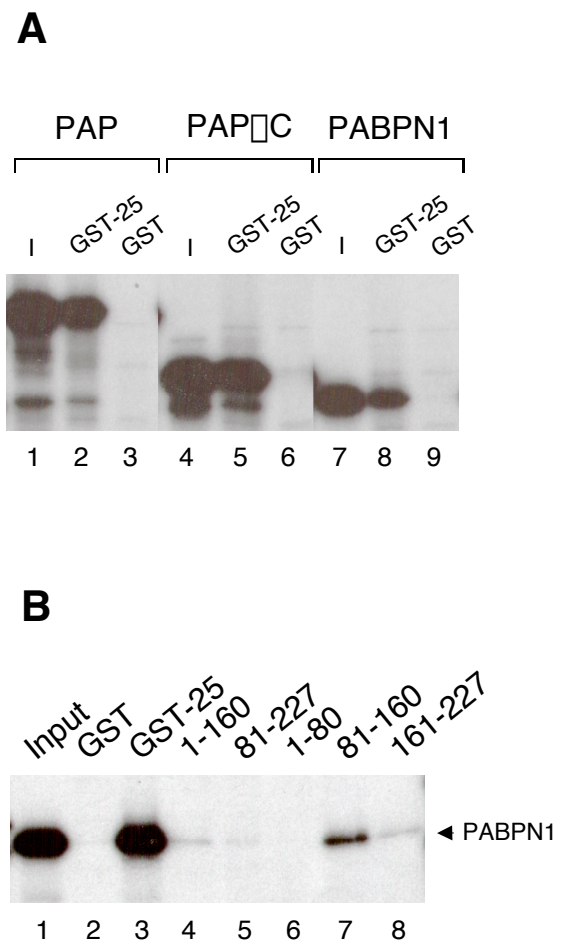
Domain organization of human cleavage factor I_m

Figure 1.



Domain organization of human cleavage factor I_m

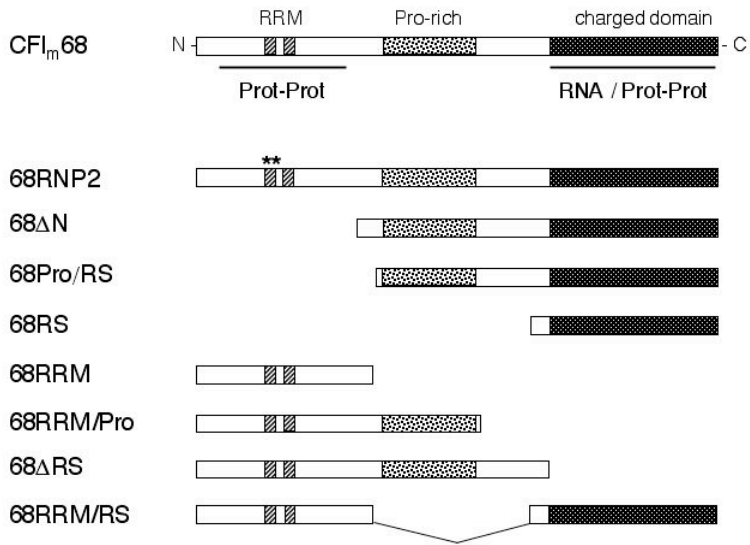
Figure 2.



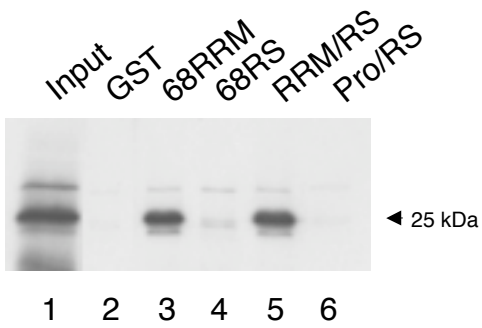
Domain organization of human cleavage factor I_m

Figure 3.

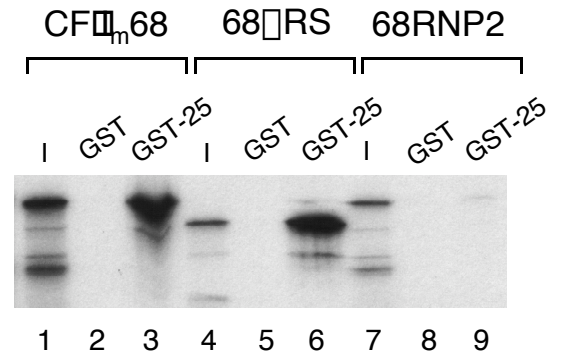
A



B

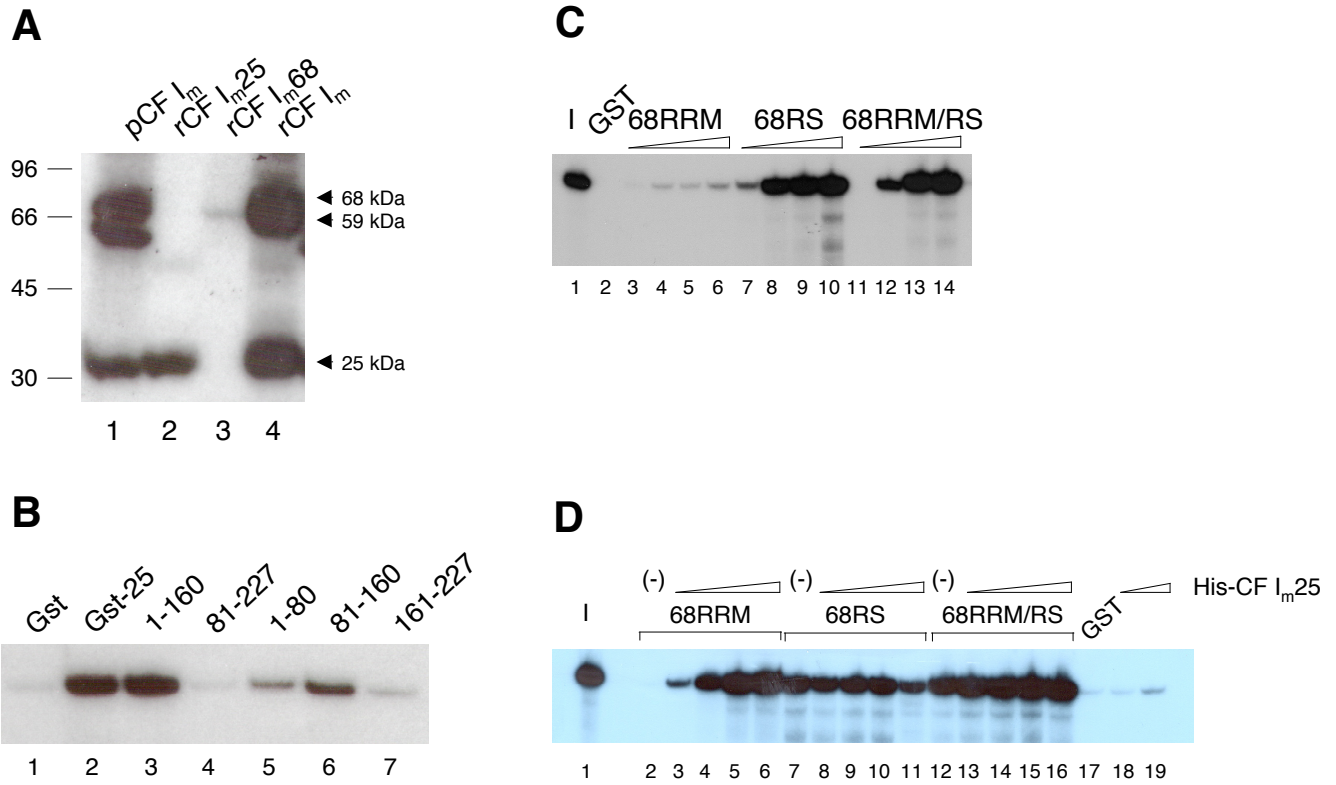


C



Domain organization of human cleavage factor I_m

Figure 4.



Domain organization of human cleavage factor I_m

Figure 5.

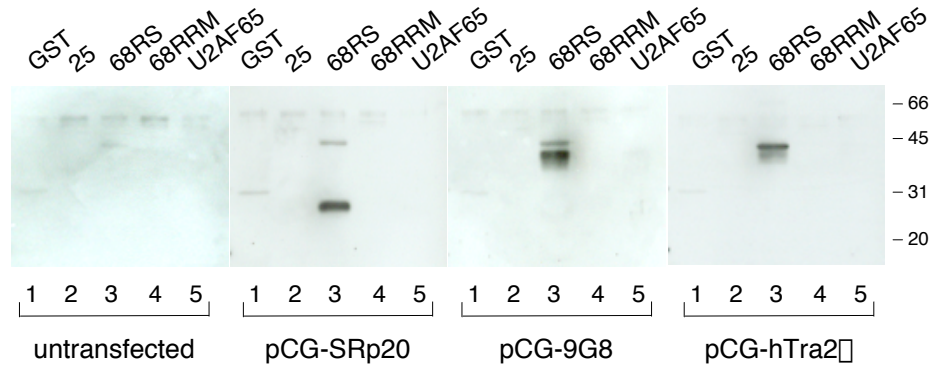
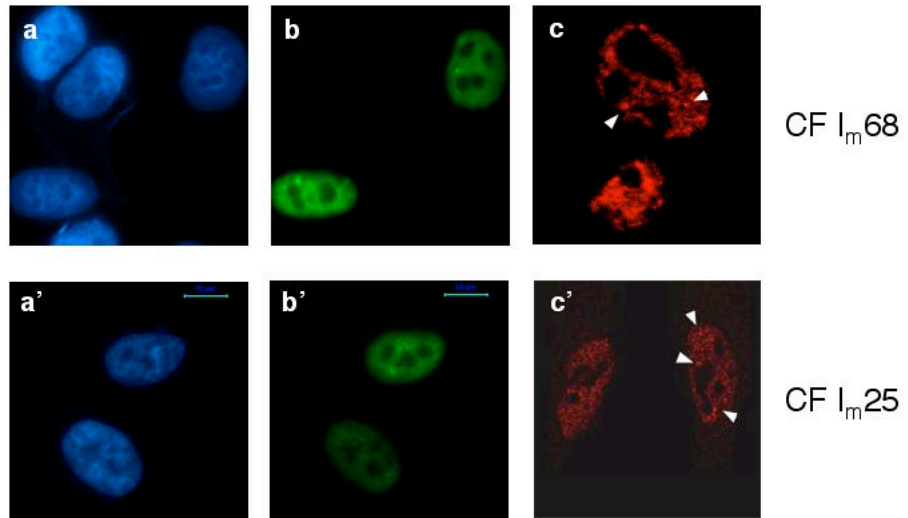
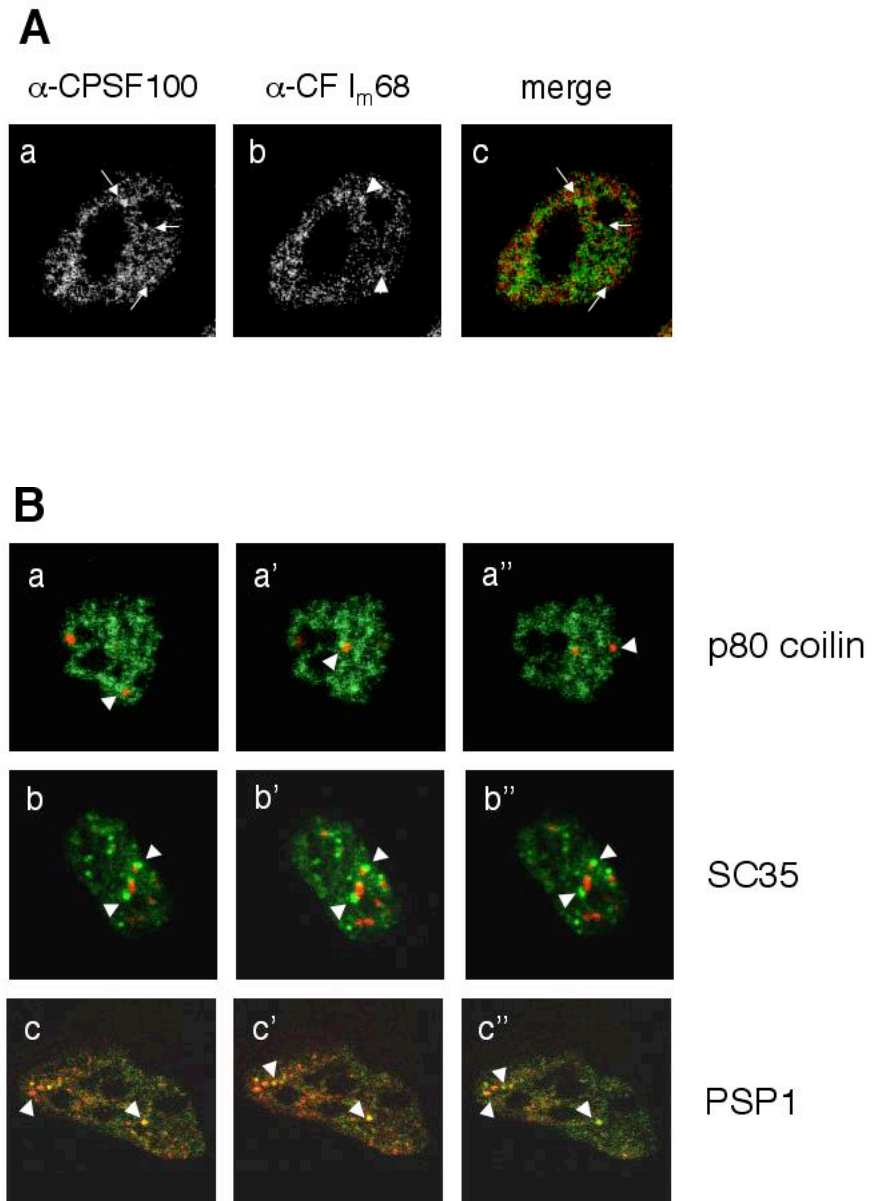


Figure 6.



Domain organization of human cleavage factor I_m

Figure 7.



Domain organization of human cleavage factor I_m

Figure 8.

