

Human pre-mRNA cleavage factor II_m contains homologs of yeast proteins and bridges two other cleavage factors

Henk de Vries¹, Ursula Rügsegger^{1,2},
Wolfgang Hübner¹, Arno Friedlein³,
Hanno Langen³ and Walter Keller^{1,4}

¹Department of Cell Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel and ³Roche Genetics, F.Hoffmann-La Roche Ltd, Grenzacherstrasse 124, CH-4070 Basel, Switzerland

²Present address: Department of Biochemistry and Biophysics, University of California, 513 Parnassus Avenue, San Francisco, CA 94143-0448, USA

⁴Corresponding author
e-mail: walter.keller@unibas.ch

H.de Vries and U.Rügsegger contributed equally to this work

Six different protein factors are required *in vitro* for 3' end formation of mammalian pre-mRNAs by endonucleolytic cleavage and polyadenylation. Five of the factors have been purified and most of their components cloned, but cleavage factor II_m (CF II_m) remained uncharacterized. We have purified CF II_m from HeLa cell nuclear extract by several chromatographic steps. During purification, CF II_m activity separated into two components, one essential (CF IIA_m) and one stimulatory (CF IIB_m) for the cleavage reaction. CF IIA_m fractions contain the human homologs of two yeast 3' end processing factors, Pcf11p and Clp1p, as well as cleavage factor I_m (CF I_m) and several splicing and transcription factors. We report the cloning of hClp1 and show that it is a genuine subunit of CF IIA_m. Antibodies directed against hClp1 deplete cleavage activity, but not polyadenylation activity from HeLa cell nuclear extract. hClp1 interacts with CF I_m and the cleavage and polyadenylation specificity factor CPSF, suggesting that it bridges these two 3' end processing factors within the cleavage complex.

Keywords: mRNA/pre-mRNA 3' end processing/
splicing/transcription

Introduction

Eukaryotic mRNA precursors (pre-mRNAs) undergo a number of processing steps before they are exported to the cytoplasm. Pre-mRNA processing includes the addition of a cap structure to the 5' end, the removal of internal non-coding sequences by splicing and the generation of a new 3' end by endonucleolytic cleavage and polyadenylation. *In vivo*, transcription, capping, splicing and 3' end processing of pre-mRNAs are coupled processes, but they can be analyzed individually *in vitro* (reviewed in Barabino and Keller, 1999; Minvielle-Sebastia and Keller, 1999; Zhao *et al.*, 1999; Hirose and Manley, 2000; Proudfoot, 2000).

The 3' ends of pre-mRNAs are generated in a two-step reaction. The pre-mRNA is first cleaved endonucleolytically and the upstream cleavage fragment is subsequently polyadenylated; the downstream cleavage product is degraded. The two steps of the reaction are tightly coupled *in vivo*, but can be uncoupled and assayed separately *in vitro*, which allowed the biochemical dissection of 3' end formation. Six *trans*-acting factors are required for the *in vitro* reconstitution of mammalian 3' end processing (reviewed in Wahle and Rügsegger, 1999; Zhao *et al.*, 1999). Cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulation factor (CstF) recognize the hexanucleotide AAUAAA upstream and a G/U-rich sequence element downstream of the cleavage site, respectively. In addition, the cleavage complex contains cleavage factors I_m (CF I_m) and II_m (CF II_m) and poly(A) polymerase (PAP). After the first step, CstF, CF I_m and CF II_m are thought to be released together with the downstream cleavage fragment; CPSF remains bound to the upstream cleavage product and tethers PAP to the RNA. PAP is the enzyme responsible for the addition of the poly(A) tail in a processive reaction that also requires both CPSF and poly(A)-binding protein II (PABP2).

The sequences of many of the polypeptides involved in 3' end formation are conserved from yeast to mammals, although the *cis*-acting RNA elements are different. It is unclear to what extent the polypeptides that are similar in sequence are also functional homologs (reviewed in Keller and Minvielle-Sebastia, 1997). For example, the two yeast CF IA subunits Rna14p and Rna15p (Minvielle-Sebastia *et al.*, 1997) share sequence similarity with the 77 and 64 kDa subunits of CstF, but are, in contrast to CstF, essential for both cleavage and polyadenylation (Kessler *et al.*, 1996; Amrani *et al.*, 1997; Minvielle-Sebastia *et al.*, 1997). Furthermore, the other two CF IA subunits, Pcf11p and Clp1p, are not homologous to the third subunit of CstF.

With the exception of CF II_m, all mammalian 3' end processing factors have been purified to homogeneity and cDNA clones for most of their subunits have been isolated. To characterize the last missing component of the mammalian 3' end processing machinery, we purified human CF II_m from HeLa cell nuclear extract. During purification of CF II_m, the activity separated into an essential fraction (CF IIA_m) and an apparently non-essential, but stimulatory component (CF IIB_m). The most purified CF IIA_m fractions still contained >15 polypeptides, all of which were identified by microsequencing. The CF IIA_m fraction contained the human homologs of the yeast CF IA subunits Pcf11p and Clp1p (hClp1 and hPcf11) as well as CF I_m and several splicing and transcription factors. We have isolated cDNA clones for hClp1 and have raised antibodies against the heterologously expressed protein. We show that hClp1 is a true

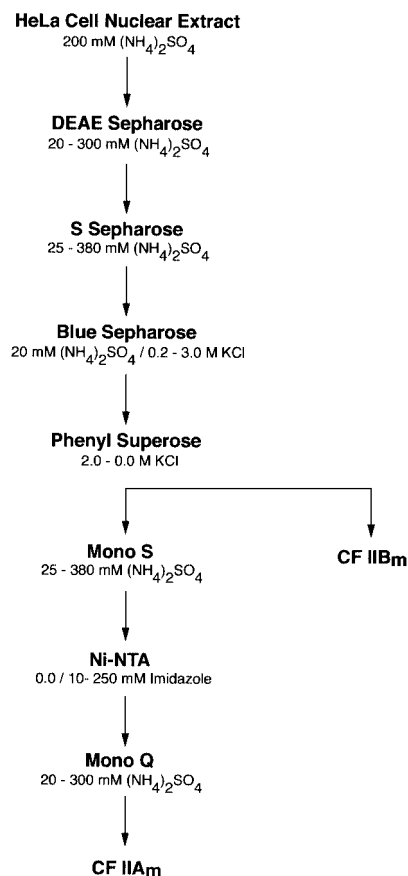


Fig. 1. Purification scheme of CF IIA_m. The chromatographic columns used to separate CF IIA_m and CF IIB_m and to purify CF IIA_m are shown schematically. The concentrations and types of salts for the preparation of HeLa cell NXT and during the fractionation are indicated below each step.

subunit of CF IIA_m, which is required for cleavage, but not for polyadenylation of pre-mRNA. In addition, we found that hC1p1 interacts directly with CF I_m and CPSF.

Results

Purification of CF IIA_m

CF IIA_m was purified from HeLa cell nuclear extract (NXT) by seven chromatographic steps (Figure 1). The purification was monitored by testing column fractions for their ability to catalyze the specific cleavage of SV40 late pre-mRNA at the poly(A) site in the presence of CF I_m, CPSF, CstF and PAP. On the DEAE column, CstF and PAP eluted in the flow-through, whereas CF IIA_m, CF I_m and CPSF bound. Both CF I_m and CPSF were found to co-elute with CF IIA_m activity on DEAE-Sepharose and Blue Sepharose columns. CF IIA_m and CF I_m could be partially separated by chromatography on S Sepharose, Mono S and phenyl Superose columns (Rüegsegger *et al.*, 1996 and results not shown). During chromatography on phenyl Superose, CF IIA_m separated into two components: CF IIA_m, which bound, and CF IIB_m, which eluted in the flow-through. Addition of CF IIB_m to *in vitro* cleavage assays with SV40 late or L3 pre-mRNA as substrate increases cleavage activity significantly, but apparently is not essential for cleavage *in vitro* as the reaction also occurs

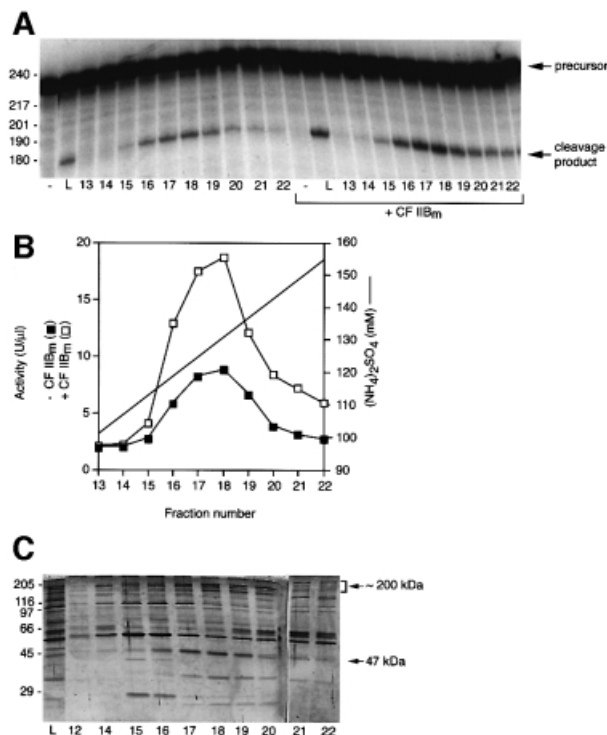


Fig. 2. Purification of CF IIA_m by Mono Q column chromatography. (A) CF IIA_m activity profile over the Mono Q column with and without the addition of CF IIB_m. Cleavage assays were carried out as described in Materials and methods for 85 min at 30°C with SV40 late pre-mRNA as substrate and 8 μl of the fractions indicated at the bottom. Samples to which 2 μl of CF IIB_m were added are marked with a bracket. Samples were analyzed on a denaturing 6% (w/v) polyacrylamide gel. Sizes of standards in nucleotides are indicated on the left. -, SV40 late pre-mRNA incubated with all protein factors except CF IIA_m; L, 4 μl of the load of the Mono Q column. (B) Quantitation of the assay shown in (A). The Mono Q fractions were assayed in the presence (open squares) or absence (filled squares) of 2 μl of CF IIB_m. Activities (U/μl) were determined as described (Rüegsegger *et al.*, 1996). (C) Silver stain of a 10% SDS-polyacrylamide gel resolving proteins from the Mono Q fractions. Aliquots of 2 μl of the fractions indicated at the bottom were loaded. The molecular masses of the size standards in kDa are indicated on the left. Polypeptides co-eluting with CF IIA_m activity are indicated with an arrowhead and a bracket. L, the load of the Mono Q column.

in its absence (Figure 2A and B). CF IIB_m has no CF IIA_m activity on its own (negative control in the presence of CF IIB_m in Figure 2A) as well as no effect on specific and unspecific polyadenylation of pre-cleaved L3 RNA (results not shown). None of the factors known to affect 3' end processing could be detected in CF IIB_m by western blot analysis with specific antibodies (results not shown; the following proteins were tested: CF I_m, CstF, CPSF, PAP, PABP1, PABP2, hC1p1, hPcf11, CBP 80 kDa, U2AF65, PTB, U1A and RNA polymerase II). Thus, CF IIB_m is most probably a new factor that stimulates the first step of 3' end processing.

Since CF IIA_m is essential for cleavage of pre-mRNA, we purified it further. Even after seven chromatographic steps, CF IIA_m still contained >15 polypeptides (Figure 2C). Among these, one polypeptide of 47 kDa and a group of large molecular mass polypeptides (140–200 kDa) co-eluted with CF IIA_m activity on a Mono Q column. CF IIA_m contained significant amounts of CF I_m, but was free of CPSF, which separated on the

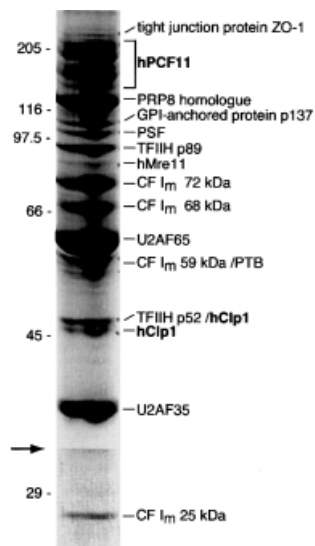


Fig. 3. Identification of the polypeptides in CF IIA_m. A peak fraction of the final Mono Q column was separated on a 10% SDS-polyacrylamide gel and the gel was stained with colloidal Coomassie. The lower part of the image (indicated with an arrow) was processed differently to compensate for the weaker staining of low molecular mass polypeptides. The molecular masses of the size standards in kDa are indicated on the left. All polypeptides were identified by mass spectrometry and are labeled accordingly on the right. Polypeptides highlighted in bold co-elute precisely with CF IIA_m activity on Mono Q. The accession numbers are: tight junction protein ZO-1, SwissProt (sp): Q07157; hPcf11, DDBJ/EMBL/GenBank (gb): AB020631; PRP8 homolog, sp: O60231; GPI-anchored protein p137, sp: Q14444; PSF, sp: P23246; TFIIH p89, sp: P19447; hMre11, sp: P49959; CF I_m 68 kDa, gb: X67337; U2AF65, sp: P26368; PTB, sp: P26599; TFIIH p52, TREMBL (tr): Q92759; hClp1, tr: Q92989; U2AF35, sp: Q01081; CF I_m 25 kDa, tr: O43809. No database accession numbers have been assigned yet to the CF I_m 59 and 72 kDa proteins.

final Mono Q column, as judged by western blot analysis and sequencing (Figure 3 and results not shown).

The purification factor and the yield of CF IIA_m activity throughout the purification could not be determined accurately because of the presence of contaminating nucleases in the Blue Sepharose column fractions and the removal of the stimulatory factor CF IIB_m upon chromatography on phenyl Superose. In addition, the presence of CPSF and/or CF I_m in CF IIA_m fractions also influences the cleavage activity. We estimated an ~170-fold increase in specific activity between the DEAE-Sepharose pool and the final Mono Q pool with a yield of ~8%. Approximately 700 µg of CF IIA_m were obtained from 3.5 g of HeLa cell NXT.

Two polypeptides in CF IIA_m are homologous to proteins involved in yeast 3' end formation

The proteins contained in a peak fraction of the final column (Mono Q) were fractionated on a 10% SDS-polyacrylamide gel and sequenced by peptide fingerprinting with MALDI-TOF-MS and with MS/MS peptide fragmentation (Figure 3). The Mono Q fraction contained the 3' end processing factor CF I_m and several known splicing factors: both subunits of U2AF, U2AF65 (Zamore *et al.*, 1992) and U2AF35 (Zuo and Maniatis, 1996), the polypyrimidine tract-binding protein (PTB; Patton *et al.*,

1991), the PTB-associated splicing factor (PSF; Patton *et al.*, 1993) and a human homolog of yeast PRP8 (Imamura *et al.*, 1998). This CF IIA_m fraction also contained the transcription factors TFIIH p52 (Marinoni *et al.*, 1997) and TFIIH p89 (Weeda *et al.*, 1990) as well as a DNA repair protein (hMre11; Petrini *et al.*, 1995). In addition, two membrane proteins, a tight junction protein (ZO-1; reviewed in Mitic and Anderson, 1998) and the GPI-anchored protein p137 (Gessler *et al.*, 1996), were identified. Although present in the peak fractions containing CF IIA_m activity, all these polypeptides eluted in a broader range than CF IIA_m.

A group of polypeptides of ~200 kDa and a 47 kDa band co-eluted precisely with CF IIA_m activity and were found to have sequence similarities to subunits of the yeast 3' end processing factor CF IA (Amrani *et al.*, 1997; Minvielle-Sebastia *et al.*, 1997): all polypeptides with masses between 140 and 200 kDa showed homology to Pcf11p and accordingly were named human Pcf11 (hPcf11). Since the relative amounts of these polypeptides varied between different Mono Q columns, whereas the patterns of polypeptides of <130 kDa were similar, the smaller polypeptides are likely to be degradation products of a larger protein. Database searches led to the identification of a partial cDNA sequence of hPcf11 (DDBJ/EMBL/GenBank accession No. AB020631). Within this part, hPcf11 and yeast Pcf11p have two common domains, which share 24 and 31% identity, respectively. Furthermore, hPcf11 contains an additional domain, which is not present in the yeast homolog (I.Kaufmann and H.de Vries, unpublished results). We are currently in the process of isolating a full-length cDNA clone encoding hPcf11.

The 47 kDa protein showed sequence similarity to yeast Clp1p (Minvielle-Sebastia *et al.*, 1997) and was therefore named human Clp1 (hClp1). The hClp1 open reading frame (ORF) was identified originally at one extremity of a chromosomal inverted insertion in the gene AF10 and was defined as 'human homologue to a hypothetical *Caenorhabditis elegans* ATP/GTP-binding protein' (HEAB; Tanabe *et al.*, 1996). hClp1 (TREMBL accession No. Q92989) has a total of 425 amino acids and a predicted mass of 47 kDa, in good agreement with the sizes of the two polypeptides identified by microsequencing. The smaller 47 kDa protein is most probably a degradation product of the larger polypeptide, since often only one 47 kDa protein was observed by Mono Q chromatography.

hClp1 is evolutionarily conserved, and homologous proteins can be found in several organisms (Figure 4). The amino acid sequence of hClp1 shows 55, 47, 34, 39, 29 and 23% identity with *Drosophila melanogaster*, *C.elegans*, *Arabidopsis thaliana* 1, *A.thaliana* 2, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* Clp1p. Of these polypeptides, only the function of *S.cerevisiae* Clp1p is known, and it was shown to be involved in 3' end processing (Minvielle-Sebastia *et al.*, 1997). All Clp1p homologs are similar over the entire length and contain the Walker A and B motifs, which have been implicated in ATP/GTP binding (Figure 4; reviewed in Walker *et al.*, 1982; Saraste *et al.*, 1990). The conservation of these motifs between species suggests that Clp1p is a nucleotide-binding protein, and preliminary data suggest that hClp1 indeed binds GTP (our unpublished UV cross-linking experiments). The amino acid

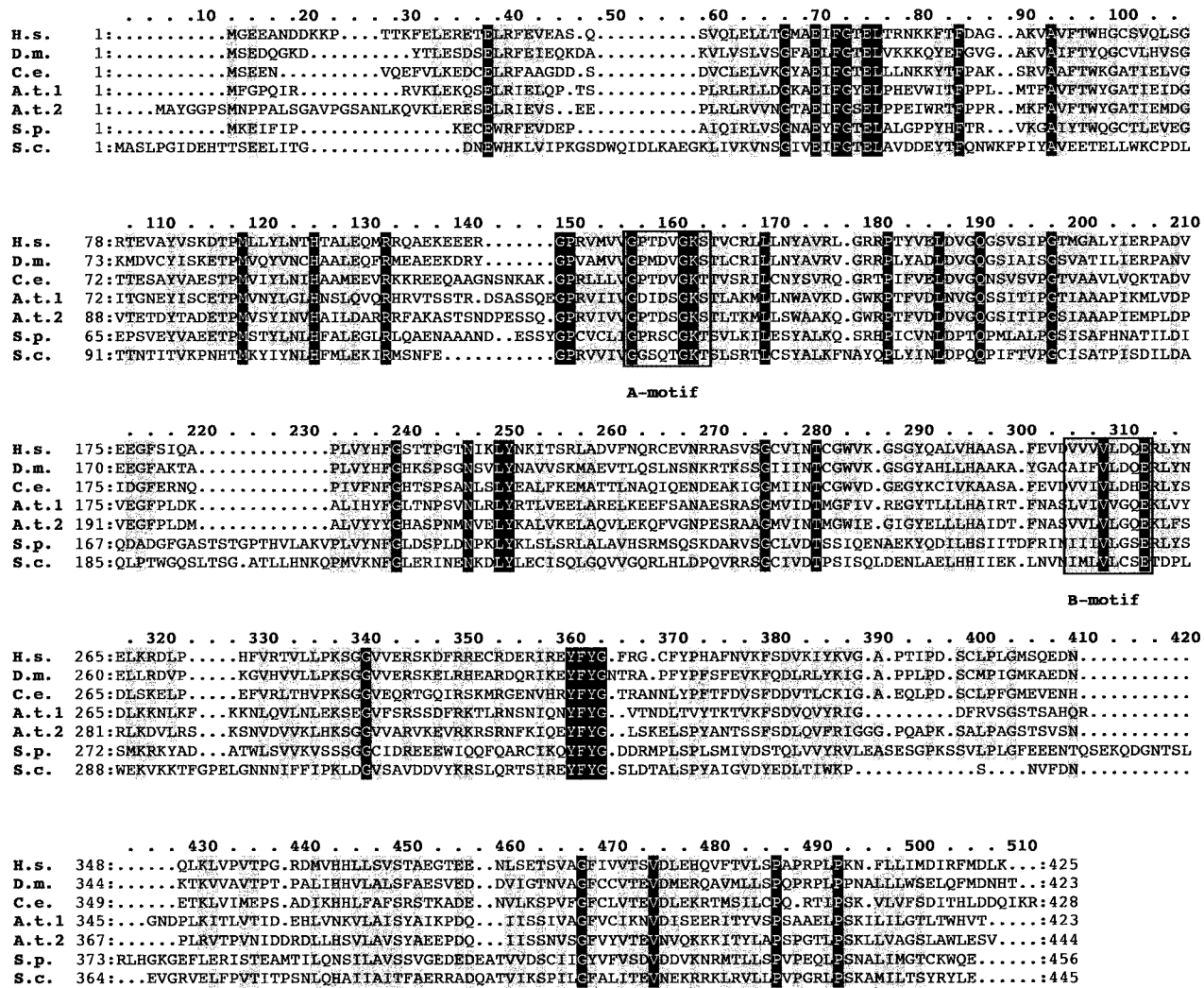


Fig. 4. hClp1 is evolutionarily conserved and has an ATP/GTP-binding motif. The alignment of *H.sapiens* (tr: Q92989), *D.melanogaster* (tr: Q9V6Q1), *C.elegans* (sp: P52874), *A.thaliana* 1 (gb: AB010077), *A.thaliana* 2 (tr: QSR06), *S.pombe* (tr: Q10299) and *S.cerevisiae* (tr: Q08685) Clp1p sequences was generated with clustalx. The black and gray boxes indicate identical and similar residues, respectively. The conserved Walker A motif with the consensus sequence -A/G-X-X-X-G-K-S/T- and the B motif are indicated.

sequence of hClp1 does not show any known nuclear localization signal. However, western blot analysis with hClp1 antiserum showed that hClp1 could be detected in nuclear, but not in cytoplasmic extract (results not shown).

hClp1 is an essential CF IIA_m subunit

We isolated the hClp1 ORF by PCR from HeLa cDNA and expressed recombinant proteins carrying either a histidine tag or a GST tag followed by a TEV protease cleavage site (GST-TEV-hClp1) in *Escherichia coli* (Nishi et al., 1998). Analysis of Mono Q fractions on a western blot with antibodies raised against the His-tagged protein (α -hClp1) showed that the 47 kDa polypeptide co-eluting with CF IIA_m activity is indeed hClp1 (Figure 5).

To test whether hClp1 is a true subunit of CF IIA_m, we carried out immunodepletion experiments. HeLa cell NXT or partially purified CF IIA_m fractions were depleted with α -hClp1 bound to protein A-Sepharose. The depleted supernatants were assayed for cleavage of SV40 late pre-mRNA. No CF IIA_m cleavage activity could be detected in

the supernatants after incubation with α -hClp1 (Figure 6A, lanes 3, 12 and 13). Full cleavage activity could be restored by the addition of purified CF IIA_m (lanes 4, 5, 14 and 15). After incubation with pre-immune serum bound to protein A-Sepharose, cleavage activity was only slightly reduced in the supernatants (lanes 2, 9, 10 and 11), which was caused by dilution. As control for the depletion of hClp1, we carried out western blot analysis, which revealed that the HeLa cell NXT as well as the CF IIA_m fractions were depleted of hClp1 with the antibody, but not after incubation with the pre-immune serum (results not shown).

From the polypeptides identified in purified CF IIA_m, CF I_m and hPcf11 co-immunoprecipitated with hClp1 from HeLa cell NXT, and CF I_m, hPcf11, U2AF and TFIIF p89 from purified CF IIA_m (results not shown and see below). Whether the other polypeptides in CF IIA also co-immunoprecipitate with hClp1 from purified fractions could not be determined for technical reasons, with the exception of ZO-1, which was never observed to co-immunoprecipitate.

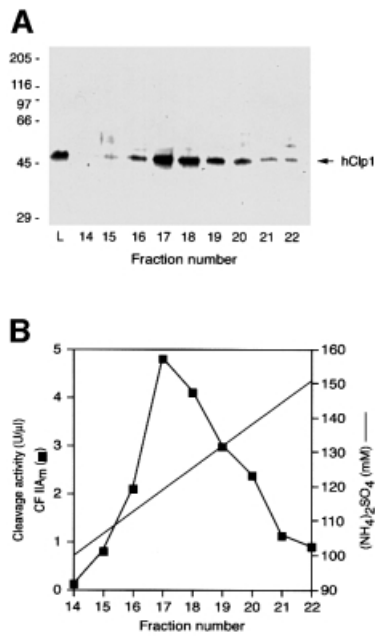


Fig. 5. hClp1 co-elutes with CF IIA_m activity from a Mono Q column. In contrast to the profile shown in Figure 2B, the load of this column was a pool from a phenyl Superose column and the Mono S and Ni-NTA steps were omitted. (A) Aliquots of 10 μ l of the column fractions indicated at the bottom were separated on a 10% SDS-polyacrylamide gel and blotted onto a nitrocellulose membrane, which was detected with α -hClp1. The molecular masses of the size standards in kDa are indicated on the left. (B) Cleavage activity profile of the Mono Q column. Assays were carried out as described in Materials and methods for 85 min at 30°C with 2 μ l of the fractions indicated at the bottom and L3 pre-mRNA as substrate. Samples were analyzed on a denaturing 6% (w/v) polyacrylamide gel and the CF IIA_m activity (filled squares) was determined as described in Materials and methods.

Although CF I_m co-immunoprecipitated with hClp1, addition of purified CF I_m to HeLa cell NXT depleted with α -hClp1 was unable to restore cleavage activity (results not shown). Therefore, depletion of hClp1 from HeLa cell NXT is likely to be directly responsible for the inactivation of cleavage activity, indicating that hClp1 is a genuine CF IIA_m component. However, addition of recombinant hClp1 protein alone or in combination with CF I_m to HeLa cell NXT after depletion with α -hClp1 could not rescue cleavage activity. In addition, recombinant protein was also unable to replace purified CF IIA_m in *in vitro* cleavage assays (results not shown), suggesting that hClp1 alone is not sufficient to restore cleavage activity. hClp1 is therefore probably not the only subunit of CF IIA_m. hPcf11 is the most likely candidate for a second subunit, since it is co-eluting with cleavage activity, co-immunoprecipitating with hClp1 and is homologous to the yeast 3' end processing factor Pcf11 (I.Kaufmann and H.de Vries, unpublished results).

Addition of CF IIA_m or His-tagged hClp1 to specific polyadenylation assays showed that neither CF IIA_m nor the recombinant protein has a significant effect on polyadenylation (results not shown). Consistent with these findings, depletion of hClp1 from HeLa cell NXT did not impair specific polyadenylation activity, although cleavage activity was abolished (Figure 6B).

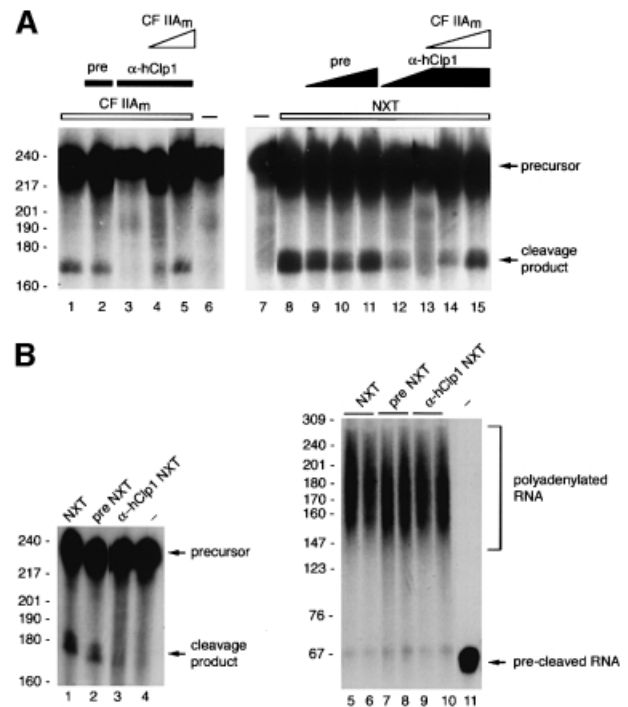


Fig. 6. Antibodies directed against hClp1 deplete CF IIA_m activity from partially purified CF IIA_m and dialyzed HeLa cell NXT. Cleavage assays were carried out as described in Materials and methods for 90 min at 30°C with SV40 late pre-mRNA as substrate. Samples were analyzed on a denaturing 6% (w/v) polyacrylamide gel. Sizes of standards in nucleotides are indicated on the left. (A) Immunodepletion experiments with partially purified CF IIA_m (lanes 1–5) and dialyzed NXT (lanes 8–15). Lanes 1–5, SV40 late pre-mRNA was incubated in the presence of CF I_m, CstF, CPSF, PAP and the CF IIA_m fraction indicated. Lane 1, 4 μ l of CF IIA_m (phenyl Superose column fraction, input of the immunodepletion experiments); lane 2, 4 μ l of CF IIA_m depleted with pre-immune serum; lane 3, 4 μ l of CF IIA_m depleted with α -hClp1; lane 4, 2 μ l of CF IIA_m depleted with α -hClp1 plus 2 μ l of CF IIA_m (phenyl Superose column fraction, same as input); lane 5, 2 μ l of CF IIA_m depleted with α -hClp1 plus 4 μ l of CF IIA_m (phenyl Superose column fraction, same as input); lanes 6 and 7, SV40 late pre-mRNA incubated in the absence of protein factors; lane 8, 3 μ l of NXT (input of the immunodepletion experiments); lanes 9–11, 3 μ l of NXT depleted with increasing amounts of pre-immune serum (5, 10 and 15 μ l of beads, respectively); lanes 12 and 13, 3 μ l of NXT depleted with increasing amounts of α -hClp1 (5 and 10 μ l of beads, respectively); lanes 14 and 15, 3 μ l of NXT depleted with α -hClp1 (10 μ l of beads) to which increasing amounts of CF IIA_m were added (2 and 4 μ l of a phenyl Superose fraction). (B) hClp1 is not essential for polyadenylation *in vitro*. Extracts depleted with pre-immune serum or α -hClp1 were tested for cleavage (lanes 1–3) and polyadenylation (lanes 5–10). Polyadenylation assays were carried out as described in Materials and methods for 30 min at 30°C with L3 pre-cleaved mRNA as substrate. All samples were analyzed on denaturing 6% (w/v) polyacrylamide gels. Lanes 1, 5 and 6, 2 μ l of dialyzed NXT (input of the immunodepletion experiments); lanes 2, 7 and 8, 2 μ l of NXT depleted with pre-immune serum; lanes 3, 9 and 10, 2 μ l of NXT depleted with α -hClp1; lane 4, SV40 late pre-mRNA; lane 11, L3 pre-cleaved mRNA.

hClp1 interacts with CF I_m and CPSF

Both CPSF and CF I_m co-purify with CF IIA_m over several chromatographic steps. To test whether hClp1 specifically interacts with CF I_m and CPSF, immunoprecipitation experiments were carried out with α -hClp1 and HeLa cell NXT (Figure 7A). The immunoprecipitated proteins were tested for the presence of CF I_m and CPSF by western blot analysis with antibodies directed against one of their

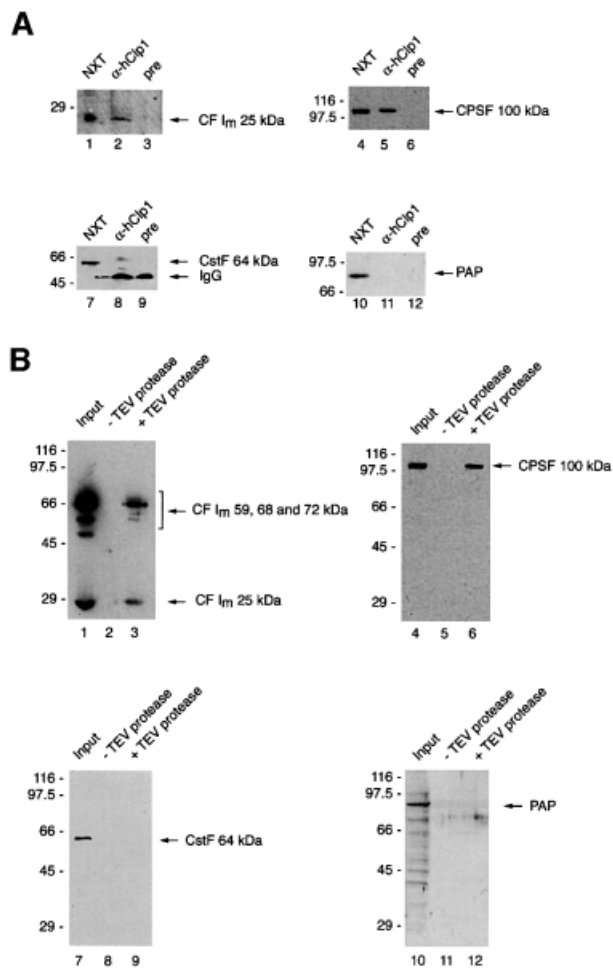


Fig. 7. CF I_m and CPSF interact with hClp1. (A) Western blot analysis of immunoprecipitations with α -hClp1 and pre-immune serum, respectively. Proteins were separated on 10% SDS-polyacrylamide gels, blotted onto nitrocellulose membranes, and the blots were probed with α -CF I_m^{25 kDa}, α -CPSF^{100 kDa}, α -CstF^{64 kDa} or α -PAP, respectively. Lane 1, 1 μ l of NXT (10% of the input of the immunoprecipitations); lane 2, 40% of the immunoprecipitate (α -hClp1); lane 3, 40% of the immunoprecipitate (pre); lanes 4, 7 and 10, 3 μ l of NXT (3% of the input of the immunoprecipitations); lanes 5, 8 and 11, 50% of the immunoprecipitate (α -hClp1); lanes 6, 9 and 12, 50% of the immunoprecipitate (pre). (B) Binding experiments with GST-TEV-tagged hClp1 were carried out as described in Materials and methods. The samples were separated on a 10% SDS-polyacrylamide gel and blotted onto a nitrocellulose membrane, which was probed with α -CF I_m^{68 kDa} and α -CF I_m^{25 kDa}, α -CPSF^{100 kDa}, α -CstF^{64 kDa} or α -PAP, respectively. Lane 1, 5 μ l of CF I_m (30% of input); lane 2, 14 μ l of the mock digest (28%); lane 3, 14 μ l of the digest with TEV protease (28%); lane 4, 5 μ l of CPSF (30% of input); lane 7, 5 μ l of CstF (10% of input); lane 10, 3 μ l of PAP (10% of input); lanes 5, 8 and 11, 14 μ l of the mock digest (20%); lanes 6, 9 and 12, 14 μ l of the digest with TEV protease (20%). The molecular masses of the size standards in kDa are indicated on the left.

constituent subunits. α -hClp1 precipitated CF I_m and CPSF along with hClp1 (Figure 7A, lanes 2 and 5), suggesting that these factors interact with each other either directly or indirectly. These interactions were not mediated by RNA, since both factors were also immunoprecipitated from HeLa cell NXT treated with RNase A and micrococcal nuclease (results not shown).

To test for a direct physical interaction, recombinant hClp1 carrying a GST tag followed by a TEV protease

cleavage site was incubated with purified CF I_m or with purified CPSF and then bound to glutathione-Sepharose. Subsequent incubation of the extensively washed beads with TEV protease released not only hClp1, but also CF I_m (Figure 7B, lane 3) and CPSF (Figure 7B, lane 6). Immunodepletion of CF I_m fractions with 68 kDa α -CF I_m strongly reduced their activity, which might be due to partial co-immunodepletion of CF I_m with CF I_m (results not shown). From these results, we conclude that hClp1 tethers CF I_m and CPSF to CF IIA_m. CstF and PAP were not immunoprecipitated with α -hClp1 from HeLa cell NXT (Figure 7A). In line with these results, purified CstF and bovine PAP did not bind to GST-TEV-tagged hClp1 (Figure 7B).

Taken together, these results strongly suggest that hClp1 is a true subunit of mammalian CF IIA_m, which is essential for cleavage, but not for polyadenylation of pre-mRNA and which makes protein-protein contacts to CF I_m and CPSF.

Discussion

CF I_m separates into two factors

We report the partial purification of CF I_m from HeLa cell NXT. Upon fractionation, CF I_m separates into two factors, CF IIA_m and CF IIB_m. CF IIA_m is an essential cleavage factor, whereas CF IIB_m is an apparently non-essential stimulatory cleavage activity (Figure 2). CF IIB_m has no effect on polyadenylation and contains no factor previously shown to be involved in 3' end processing. We conclude from these results that CF IIB_m is a new 3' end processing factor. Purified CF IIA_m contains polypeptides homologous to subunits of the yeast 3' end processing factor CF IA, hPcf11 and hClp1. In addition, significant amounts of CF I_m as well as several splicing factors and components of the transcription machinery were present in the purified CF IIA_m preparation (see below).

hClp1 and hPcf11 are subunits of CF IIA_m

The fact that multiple polypeptides were identified as hClp1 and hPcf11 by protein sequencing suggests that the proteins were either partially degraded during purification or that they are modified post-translationally. It is also possible that isoforms of hPcf11 and hClp1 are generated by alternative splicing. Alternative splicing of 3' end processing factors would not be unprecedented (for example, see Zhao and Manley, 1996).

Several lines of evidence indicate that hPcf11 and hClp1 are true subunits of CF IIA_m. First, these are the only two polypeptides that co-eluted precisely with CF IIA_m activity on all chromatographic columns. All other polypeptides, although present in significant amounts in purified CF IIA_m, can be partially separated from CF IIA_m activity. This suggests that they may be associated with CF IIA_m, but are not essential for CF IIA_m activity. Secondly, many components of the yeast and mammalian 3' end processing apparatus are conserved. Therefore, the fact that the two proteins share significant sequence similarities with the Pcf11p and Clp1p subunits of yeast CF IA (Kessler *et al.*, 1996; Minvielle-Sebastia *et al.*, 1997) strongly suggests that the two human homologs are part of the 3' end processing apparatus in mammals. Intriguingly, yeast Clp1p and Pcf11p interact with each

other in a two-hybrid assay (Uetz *et al.*, 2000). The presence of both mammalian homologs in the same complex throughout all purification steps suggests a direct interaction between the two polypeptides. Thirdly, hClp1 interacts with the known 3' end processing factors CF I_m and CPSF. Finally, immunodepletion of CF IIA_m fractions with a polyclonal antibody directed against recombinant hClp1 abolishes CF IIA_m activity.

hClp1 bridges two 3' end processing factors

Results from immunoprecipitation experiments and from interaction studies with purified components suggest that hClp1 interacts directly with CF I_m and CPSF, but not with CstF or PAP (Figure 7). Based on these data and in analogy to the yeast CF IA factor, we propose that core CF IIA_m is composed of hPcf11 and hClp1, which are likely to interact directly, and that hClp1 contacts CF I_m and CPSF within the cleavage complex.

While CF IIA_m contains no CPSF after seven chromatographic steps, we were unable to obtain CF IIA_m free of CF I_m, although the bulk of CF I_m is separated from CF IIA_m during purification. It is therefore likely that CF I_m is more abundant than CF IIA_m in cells. This may also be true for CPSF since it co-immunoprecipitates with CF IIA_m from HeLa cell NXT but activity can be restored by adding CF IIA_m that is essentially free of CPSF.

HeLa cell NXT depleted of hClp1 with polyclonal antibodies is not significantly reduced in polyadenylation activity. Therefore, mammalian CF IIA_m is essential for cleavage, but not for polyadenylation of pre-mRNA. This is in contrast to yeast CF IA, which is essential for both cleavage and polyadenylation (Kessler *et al.*, 1996; Minvielle-Sebastia *et al.*, 1997). Thus, despite the sequence conservation of the polypeptides, the function of the homologous proteins seems to differ between these organisms. This is also true for the other two subunits of yeast CF IA: their mammalian homologs are subunits of CstF. CstF is essential for cleavage of all pre-mRNAs, but has also been shown to stimulate polyadenylation of at least one transcript (Moreira *et al.*, 1998). It is therefore conceivable that during evolution certain polypeptides became essential for cleavage only but still stimulate polyadenylation. Whether this is a more general phenomenon remains to be shown.

Although we believe we now know the amino acid sequences of all polypeptides involved, the identity of the endonuclease remains elusive. Most polypeptides are apparently involved in the recognition of the *cis*-acting elements or in holding the complex together. It is therefore conceivable that the endonuclease is present as a substoichiometric 'contamination' in one of the purified factors or has not been recognized as the endonuclease.

CF IIA_m and CF I_m may connect 3' end processing to transcription and splicing

In addition to the proteins discussed above, which are directly involved in pre-mRNA 3' end formation, CF IIA_m fractions also contain two transcription factors, TFIIF p52 and TFIIF p89, subunits of the TFIIF complex. Another subunit of TFIIF, CDK7, is a cyclin-dependent kinase that has been shown to phosphorylate the C-terminal domain (CTD) of RNA polymerase II (Lu *et al.*, 1992).

Intriguingly, addition of RNA polymerase II stimulates cleavage of pre-mRNA *in vitro* in the absence of transcription (Hirose and Manley, 1998). Whether the association of CF IIA_m with two TFIIF subunits is functionally significant and whether it modulates the kinase activity of TFIIF remain to be investigated, but this would fit well with the findings that transcription and 3' end processing are tightly coupled *in vivo*.

Kleiman and Manley (1999) have recently identified the BRCA1-associated protein BARD1 based on its ability to interact with the 50 kDa subunit of CstF. BARD1 inhibits the first step of 3' end processing. It binds to the CTD and co-localizes with DNA repair factors *in vivo*. These authors proposed that BARD1 might prevent premature polyadenylation of nascent transcripts associated with RNA polymerases stalled at sites of DNA damage. The presence of two TFIIF subunits and the BRCA1-associated protein hMre11 in CF IIA_m supports this idea, since both TFIIF and hMre11 are involved in DNA repair (reviewed in Petrini *et al.*, 1995; Frit *et al.*, 1999).

Results from both *in vitro* and *in vivo* experiments supporting physical and functional interactions between the splicing and polyadenylation machineries have accumulated over the last few years (reviewed in Barabino and Keller, 1999; Zhao *et al.*, 1999). The splicing factors PTB, PSF, a PRP8 homolog as well as both U2AF subunits were identified in the CF IIA_m fraction. In addition to the fact that CF I_m, U2AF and PTB partially co-purify with CF IIA_m, the two splicing factors also co-immunoprecipitate with CF I_m from HeLa cell NXT (results not shown). Since SR proteins have been shown to interact with each other via their RS domains (Wu and Maniatis, 1993), sequence elements that are also present in both U2AF and CF I_m, a direct interaction of U2AF and CF I_m is likely. CF I_m might therefore physically link the splicing and 3' end processing machineries by interacting simultaneously with U2AF (and possibly other spliceosomal SR proteins) and hClp1. Thus, CF IIA_m and CF I_m might be involved in the coupling of transcription and splicing with 3' end processing.

While the presence of splicing and transcription factors in CF IIA_m fits well with our current understanding of RNA maturation, the significance of the presence of two membrane proteins, the GPI-anchored protein p137 and the tight junction protein ZO-1, is less clear. It remains to be shown whether the presence of ZO-1 and the GPI-anchored protein in purified CF IIA_m is coincidental or whether these proteins add a new aspect to the production of mRNA.

Materials and methods

Cleavage and polyadenylation assays

All proteins for *in vitro* assays were purified as described (Rüegsegger *et al.*, 1996). RNAs used as substrates for cleavage assays were prepared by *in vitro* transcription (Rüegsegger *et al.*, 1996). Cleavage assays were carried out and quantitated (units) according to Rüegsegger *et al.* (1996, 1998). One unit of CF IIA_m activity corresponds to 1 fmol of upstream cleavage product formed during the time indicated. Purification factor and yield of the CF IIA_m purification were estimated as described previously (Rüegsegger *et al.*, 1996). Specific *in vitro* polyadenylation assays were carried out with CPSF and PAP in the presence of MgCl₂ (Wahle, 1991). No PABP2 was added to the polyadenylation assays.

SDS-PAGE and western blot analysis

SDS-PAGE and western blot analysis were carried out according to standard procedures and the proteins on the blots were detected with the ECL system (Amersham). The antibodies directed against 68 kDa CF I_m (α -CF I_m^{68 kDa}), 25 kDa CF I_m (α -CF I_m^{25 kDa}), 100 kDa CPSF (α -CPSF^{100 kDa}), 73 kDa CPSF (α -CPSF^{73 kDa}) and 30 kDa CPSF (α -CPSF^{30 kDa}) were described previously (Jenny *et al.*, 1994, 1996; Barabino *et al.*, 1997; Rügsegger *et al.*, 1998). Antibodies directed against 64 kDa CstF (α -CstF^{64 kDa}) and PAP (α -PAP) were kindly provided by Clinton McDonald and Georges Martin, respectively.

Purification of CF II_m

The purification of CF II_m is shown schematically in Figure 1. All procedures were carried out at 0–4°C and the loads were spun immediately before applying to the columns. Between purification steps, the samples were frozen in liquid nitrogen and stored at –80°C. All buffers used during purification contained 50 mM Tris–HCl pH 7.9, 0.4 mM Na₂EDTA pH 8.0, 0.02% (v/v) NP-40, 10% (v/v) glycerol, 3 mM MgCl₂, 0.5 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF; Serva), 0.4 µg/ml leupeptin hemisulfate (Fluka), 0.7 µg/ml pepstatin (Bachem) and ammonium sulfate (Life Technologies Inc.) or KCl at the concentrations indicated. Column fractions were assayed for CF II_m, CF IIA_m and CF IIB_m activities as appropriate. In addition, many column fractions were tested for the presence of CF I_m and CPSF by western blot analysis with α -CF I_m^{68 kDa} and α -CPSF^{73 kDa} or α -CPSF^{30 kDa}, respectively. HeLa cell NXT was prepared as described (Rügsegger *et al.*, 1996), but buffers contained 3 mM MgCl₂ instead of 1.5 mM. A 390 ml aliquot of NXT was diluted with buffer without ammonium sulfate to the conductivity of a buffer containing 30 mM ammonium sulfate. A 4.1 l aliquot (3.5 g of protein) of diluted NXT was applied to a DEAE–Sepharose fast flow column (Pharmacia) of 920 ml equilibrated in 20 mM ammonium sulfate buffer. The column was washed with 1 column volume of the same buffer and developed with a gradient (3.5 column volumes) of 20–300 mM ammonium sulfate at 1 l/h.

Fractions containing CF I_m and CF II_m, eluting between 90 and 150 mM ammonium sulfate, were pooled (980 ml, 1.5 g of protein) and dialyzed for 8 h against 3 × 15 l of buffer containing 25 mM ammonium sulfate and 20 mM HEPES–KOH pH 7.9 instead of Tris–HCl pH 7.9. The dialyzed pool was loaded onto three S Sepharose columns (Pharmacia) of 35, 75 and 150 ml, respectively, equilibrated in the dialysis buffer. The columns were washed with 1.4 column volumes of the same buffer and eluted with gradients (5 column volumes) of 25–380 mM ammonium sulfate at 1.5 column volumes/h. The CF II_m fractions from all three columns, eluting between 150 and 260 mM ammonium sulfate, were combined (900 ml, 300 mg of protein) and dialyzed against 16 l of buffer containing 20 mM ammonium sulfate three times for 2 h. The dialyzed pool was loaded onto a 150 ml Blue Sepharose column (Bienroth *et al.*, 1991) at 1.5 column volumes/h. The column was washed with 3 column volumes of the dialysis buffer and developed with a gradient of 0.2–3.0 M KCl (5 column volumes). Fractions showing CF II_m activity were pooled (0.2–1.4 M KCl; 170 ml, 128 mg of protein) and dialyzed for 6 h against 5 l of 2.0 M KCl buffer. The pool was separated into two portions of 85 ml (64 mg of protein) and each portion was applied separately to an 8 ml phenyl Superose HR 10/10 FPLC column (Pharmacia) equilibrated in the dialysis buffer. The column was washed with 10 ml of the same buffer and was developed with a 100 ml gradient from 2 to 0 M KCl at 0.7 ml/min.

On the phenyl Superose column, CF II_m separated into two activities, CF IIA_m (gradient) and CF IIB_m (flow-through). CF IIB_m fractions were not purified further. Fractions from both columns containing CF IIA_m activity were combined (1.4–0.3 M KCl; 60 ml, 40 mg of protein) and dialyzed for 3 h against 5 l of buffer containing 25 mM ammonium sulfate and 20 mM HEPES–KOH pH 7.9 instead of Tris–HCl pH 7.9. The dialyzed material was loaded onto two 1 ml Mono S HR 5/5 FPLC columns (Pharmacia) equilibrated in the dialysis buffer. The columns were washed with 6 column volumes of the same buffer and eluted with gradients of 30 ml from 25 to 380 mM ammonium sulfate at 0.5 ml/min. The fractions from both columns containing CF IIA_m activity were pooled (140–260 mM ammonium sulfate; 30 ml, 12 mg of protein) and dialyzed for 2 h against 2 l of buffer with 50 mM KCl containing no EDTA and no DTT. The dialyzed pool was incubated on a rotating wheel for 2 h with 2 ml of Ni–NTA (Superflow, Qiagen) equilibrated in the same buffer. The resin was first washed with 8 ml of 50 mM KCl buffer, then twice with 0.5 ml of 50 mM KCl buffer containing 10 mM imidazole and packed into a column. A 7 ml gradient of 10–250 mM imidazole was applied at a flow rate of 0.1 ml/min. Active fractions, eluting between 60 and 250 mM imidazole, were pooled (5 ml, 3 mg of protein) and dialyzed against 1 l of

20 mM ammonium sulfate buffer for 2 h. The dialyzed pool was applied to a 1 ml Mono Q HR 5/5 FPLC column (Pharmacia) equilibrated in 20 mM ammonium sulfate buffer. The column was washed with 5 column volumes of the same buffer and developed with a gradient (35 column volumes) of 20–300 mM ammonium sulfate buffer at 0.5 ml/min. CF IIA_m eluted between 120 and 190 mM ammonium sulfate. These fractions were used to identify the polypeptides in CF IIA_m by mass spectrometry and western blot analysis (α -CF I_m^{68 kDa}). Fractions from a small-scale purification similar to the one described here are shown in Figure 2. Their protein composition is very similar to that of the fraction shown in Figure 3 and the apparent higher complexity of the fraction shown in Figure 3 is only due to different protocols used for the analysis by SDS–PAGE.

Identification of polypeptides by mass spectrometry

A peak fraction of the final Mono Q column was loaded onto a 10% SDS–polyacrylamide gel. All polypeptides were digested *in situ* with endoproteinase Lys-C or trypsin (Fountoulakis and Langen, 1997).

Nanoelectrospray tandem mass spectrometry. Unseparated peptide mixtures obtained after tryptic digestions of unknown proteins were analyzed by nanoelectrospray tandem mass spectrometry (Wilm and Mann, 1996). Tryptic peptides were extracted as described (Fountoulakis and Langen, 1997), desalted, concentrated further on a pulled capillary containing ~100 nl of POROS R2 reverse phase material (Perseptive Biosystems) and eluted with 1 µl of 60% methanol in 5% formic acid directly into the nanoelectrospray needle. Electrospray mass spectra were acquired on an API 365 triple quadrupole mass spectrometer (Sciex) equipped with a nanoES source (Wilm and Mann, 1996). Q1 scans were carried out with a 0.2 Da mass step. For operation in the MS/MS mode, Q1 was set to transmit a mass window of 2 Da and spectra were accumulated with 0.2 Da mass steps. Resolution was set so that the fragmented masses could be assigned better than 1 Da.

Matrix-assisted laser desorption ionization mass spectroscopy. MALDI-MS analysis was carried out as described (Fountoulakis and Langen, 1997) with minor modifications. Briefly, the spots were excised, destained with 30% acetonitrile in 0.1 M ammonium bicarbonate and washed with water. The gel pieces were reswollen with 4 µl of 3 mM Tris–HCl pH 8.8 containing 0.2 µg of Lys-C (WAKO) after drying in a speedvac evaporator. The digestion was carried out for 3–12 h at room temperature. A 7 µl aliquot of 50% acetonitrile containing 0.3% trifluoroacetic acid (TFA) was added, and the content was vortexed, centrifuged for 2 min and sonicated for 5 min. A 1 µl aliquot from the separated liquid was mixed with 1 µl of saturated α -cyano cinnamic acid in 50% acetonitrile/0.1% TFA in water and applied to the MALDI target. The samples were analyzed in a time-of-flight Bruker mass spectrometer (Reflex 3) equipped with a reflector and delayed extraction. An acceleration voltage of 20 kV was used. Calibration was internal to the samples. Des-Arg-1 bradykinin (Sigma) and ACTH (18–38) (Sigma) were used as standard peptides. The peptide masses obtained for each individual polypeptide were used to search SwissProt, TREMBL and DDBJ/EMBL/GenBank databases.

Isolation of cDNAs, expression of recombinant proteins and preparation of polyclonal antibodies

The ORF coding for hClp1 was amplified from HeLa cDNA by PCR with the primers 5'-GGAATTCATATGGGAGAAGAGGCTTA-3' (2.8 pmol) and 5'-CGGGATCCTACTTCAGATCCATGAA-3' (5 pmol), which introduced *Nde*I and *Bam*HI sites at the 5' and 3' ends, respectively. The PCR products were ligated into pCR2.1 (Invitrogen) and sequenced. Correct full-length inserts were cloned into the *Nde*I and *Bam*HI sites of pGDV1-TEV (from Bernhard Dichtl; for N-terminal GST–TEV-tagged hClp1) or pGM10-(His₆) (from Georges Martin; for N-terminal His-tagged hClp1). Both recombinant proteins were expressed in *E. coli* BL21 (DE3) pLysS cells. Cells were grown in 2 × YT (GST–TEV-tagged hClp1) and LB (His-tagged hClp1), respectively. Protein expression was induced at an OD₆₀₀ of 0.7 by adding isopropyl- β -D-thiogalactopyranoside to a final concentration of 0.8 mM and the cells were harvested 2–3 h after induction. All solutions used for the purification of the recombinant proteins contained 0.5 mM PMSF, 0.4 µg/ml leupeptin hemisulfate and 0.7 µg/ml pepstatin. To purify GST–TEV-tagged hClp1, extracts prepared in phosphate-buffered saline (PBS) were incubated with reduced glutathione–Sepharose (Pharmacia), the matrix was washed three times with 10 vols of PBS and the protein was eluted with 10 mM reduced glutathione in 50 mM Tris–HCl pH 7.9. The cells expressing the His-tagged protein were resuspended in 100 ml of 50 mM Tris–HCl

pH 7.9, 10% (v/v) glycerol, 100 mM KCl. His-tagged hClp1 was purified with Ni-NTA (Qiagen) under native conditions according to the manufacturer's protocols. For the immunization of rabbits, His-tagged hClp1 was separated on a 10% SDS–polyacrylamide gel, excised and eluted. Two New Zealand white rabbits were injected with 100 µg of protein in 0.45 ml of SDS–PAGE running buffer mixed with 0.5 ml of Specol (ID-DLO). After six injections, the rabbits were terminally bled and antiserum prepared according to standard procedures.

Immunodepletion and immunoprecipitation experiments

For immunodepletion experiments, 2 ml of α -hClp1 or pre-immune serum were covalently coupled to 300 µl of protein A–Sepharose CL-4B (Pharmacia; Harlow and Lane, 1988). A phenyl Superose CF IIA_m fraction or HeLa cell NXT was dialyzed against buffer with 20 mM ammonium sulfate. Then 50 µl of the dialyzed fraction or 100 µl of NXT were incubated for 2 h at 4°C with variable amounts (5, 10 and 15 µl) of the beads equilibrated in the same buffer and the supernatants were assayed for CF IIA_m activity.

Immunoprecipitations from HeLa cell NXT with hClp1 or pre-immune serum were carried out in 150 mM KCl, 100 mM Tris–HCl pH 7.9, 0.01% NP-40. HeLa cell NXT was incubated for 2 h with the respective antibody covalently coupled to protein A–Sepharose CL-4B. The beads were washed three times in the same buffer and proteins were eluted either with 3 M MgCl₂ or by boiling in SDS–PAGE loading buffer. After separation on 10% SDS–polyacrylamide gels, polypeptides were blotted onto nitrocellulose membranes and the blots were probed with the antibodies indicated.

Binding experiments with GST–TEV-tagged hClp1

The purified 3' end processing factor indicated was incubated with GST–TEV-tagged hClp1 in PBS containing 50 µg/ml bovine serum albumin (BSA) for 1 h at 25°C on a rotating wheel. The mixture was then incubated with 20 µl of glutathione–Sepharose (Pharmacia) for 1 h at 4°C. The slurry was centrifuged and the supernatant removed. The glutathione–Sepharose beads were washed twice with 0.4 ml of 50 mM Tris–HCl pH 7.9, 150 mM KCl, 0.1% NP-40, 50 µg/ml BSA, followed by two washes with 0.4 ml of 50 mM Tris–HCl pH 7.9, 1 mM EDTA pH 8.0, 50 µg/ml BSA. The matrix was resuspended in 70 µl of 50 mM Tris–HCl pH 7.9, 1 mM EDTA pH 8.0. Twenty units of TEV protease (Gibco-BRL) were added to half of the beads, while the other half was left untreated, and both samples were incubated for 2 h at 25°C. The eluates were analyzed by western blotting.

Acknowledgements

We are grateful to Andreas Jenny, Georges Martin and Clinton McDonald for providing antibodies as specified in Materials and methods. We would like to thank Uwe Kühn and Elmar Wahle for their advice on polyadenylation assays. We also thank Silvia Barabino for providing CPSF, Georges Martin for PAP and pGM10–(His₆), Bernhard Dichtl for pGDV1–TEV, and Andrea Büerer for growing HeLa cells. We are especially grateful to Daniel Roeder for sequencing CF IIA_m by mass spectrometry. We thank Silvia Barabino, Bernhard Dichtl, André Gerber, Georges Martin and Martin Ohnacker for critical reading of the manuscript. This work was supported by the University of Basel, the Swiss National Science Foundation, the European Community (via the Bundesamt für Bildung und Wissenschaft, Bern) and the Louis-Jeantet Foundation for Medicine.

References

Amrani, N., Minet, M., Wyers, F., Dufour, M.E., Aggerbeck, L.P. and Lacroute, F. (1997) PCF11 encodes a third protein component of yeast cleavage and polyadenylation factor I. *Mol. Cell. Biol.*, **17**, 1102–1109.

Barabino, S.M. and Keller, W. (1999) Last but not least: regulated poly(A) tail formation. *Cell*, **99**, 9–11.

Barabino, S.M., Hübner, W., Jenny, A., Minvielle-Sebastia, L. and Keller, W. (1997) The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev.*, **11**, 1703–1716.

Bienroth, S., Wahle, E., Suter-Crazzolara, C. and Keller, W. (1991) Purification of the cleavage and polyadenylation factor involved in the 3'-end processing of messenger RNA precursors. *J. Biol. Chem.*, **266**, 19768–19776.

Fountoulakis, M. and Langen, H. (1997) Identification of proteins by matrix-assisted laser desorption ionization-mass spectrometry following in-gel digestion in low-salt, nonvolatile buffer and simplified peptide recovery. *Anal. Biochem.*, **250**, 153–156.

Frit, P., Bergmann, E. and Egly, J.M. (1999) Transcription factor IIIH: a key player in cellular response to DNA damage. *Biochimie*, **81**, 27–38.

Gessler, M., Klamt, B., Tsaoussidou, S., Ellis, J.A. and Luzio, J.P. (1996) The gene encoding the GPI-anchored membrane protein p137GPI (M11S1) maps to human chromosome 11p13 and is highly conserved in the mouse. *Genomics*, **32**, 169–170.

Harlow, E. and Lane, D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hirose, Y. and Manley, J.L. (1998) RNA polymerase is an essential mRNA polyadenylation factor. *Nature*, **395**, 93–96.

Hirose, Y. and Manley, J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.*, **14**, 1415–1429.

Imamura, O., Saiki, K., Tani, T., Ohshima, Y., Sugawara, Y. and Furuichi, Y. (1998) Cloning and characterisation of a human DEAH-box RNA helicase, a functional homolog of fission yeast Cdc28/Prp8. *Nucleic Acids Res.*, **26**, 2063–2068.

Jenny, A., Hauri, H.P. and Keller, W. (1994) Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit. *Mol. Cell. Biol.*, **14**, 8183–8190.

Jenny, A., Minvielle-Sebastia, L., Preker, P.J. and Keller, W. (1996) Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science*, **274**, 1514–1517.

Keller, W. and Minvielle-Sebastia, L. (1997) A comparison of mammalian and yeast pre-mRNA 3'-end processing. *Curr. Opin. Cell Biol.*, **9**, 329–336.

Kessler, M.M., Zhao, J. and Moore, C.L. (1996) Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3'-ends. *J. Biol. Chem.*, **271**, 27167–27175.

Kleiman, F.E. and Manley, J.L. (1999) Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF-50. *Science*, **285**, 1576–1579.

Lu, H., Zewel, L., Fisher, L., Egly, J.M. and Reinberg, D. (1992) Human general transcription factor IIIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature*, **358**, 641–645.

Marinoni, J.C. et al. (1997) Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH. *EMBO J.*, **16**, 1093–1102.

Minvielle-Sebastia, L. and Keller, W. (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.*, **11**, 352–357.

Minvielle-Sebastia, L., Preker, P.J., Wiederkehr, T., Strahm, Y. and Keller, W. (1997) The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in pre-messenger RNA 3'-end formation. *Proc. Natl. Acad. Sci. USA*, **94**, 7897–7902.

Mitic, L.L. and Anderson, J.M. (1998) Molecular architecture of tight junctions. *Annu. Rev. Physiol.*, **60**, 121–142.

Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L. and Proudfoot, N.J. (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev.*, **12**, 2522–2534.

Nishi, N., Morino, S., Tomoo, K., Youtani, T. and Ishida, T. (1998) Expression of a synthetic gene for initiation factor 4E-binding protein 1 in *Escherichia coli* and its interaction with eIF-4E and eIF-4E×m⁷GTP complex. *J. Biochem. (Tokyo)*, **123**, 157–161.

Patton, J.G., Mayer, S.A., Tempst, P. and Nadal-Ginard, B. (1991) Characterization and molecular cloning of polypyrimidine-tract binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes Dev.*, **5**, 1237–1251.

Patton, J.G., Porro, E.B., Galceran, J., Tempst, P. and Nadal-Ginard, B. (1993) Cloning and characterization of PSF, a novel pre-mRNA splicing factor. *Genes Dev.*, **7**, 393–406.

Petrini, J.H., Walsh, M.E., DiMare, C., Chen, X.N., Korenberg, J.R. and Weaver, D.T. (1995) Isolation and characterization of the human MRE11 homologue. *Genomics*, **29**, 80–86.

Proudfoot, N. (2000) Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.*, **25**, 290–293.

Rüegsegger, U., Beyer, K. and Keller, W. (1996) Purification and characterization of human cleavage factor I_m involved in the 3'-end processing of messenger RNA precursors. *J. Biol. Chem.*, **271**, 6107–6113.

- Rüegsegger,U., Blank,D. and Keller,W. (1998) Human pre-mRNA cleavage factor I_m is related to spliceosomal SR proteins and can be reconstituted *in vitro* from recombinant subunits. *Mol. Cell*, **1**, 243–253.
- Saraste,M., Sibbald,P.R. and Wittinghofer,A. (1990) The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.*, **15**, 430–434.
- Tanabe,S., Bohlander,S.K., Vignon,C.V., Espinosa,R., Zhao,N., Strissel,P.L., Zeleznik-Le,N.J. and Rowley,J.D. (1996) AF10 is split by MLL and HEAB, a human homologue to a putative *Caenorhabditis elegans* ATP/GTP-binding protein in an invins (10;11)(p12;q23q12). *Blood*, **88**, 3535–3545.
- Uetz,P. *et al.* (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
- Wahle,E. (1991) A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell*, **66**, 759–768.
- Wahle,E. and Rüegsegger,U. (1999) 3′-end processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.*, **23**, 277–295.
- Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.*, **1**, 945–951.
- Weeda,G., van Ham,R.C.A., Masurel,R., Westerveld,A., Odijk,H., de Wit,J., Bootsma,D., van der Eb,A.J. and Hoeijmakers,J.H. (1990) Molecular cloning and biological characterization of the human excision repair gene *ERCC-3*. *Mol. Cell. Biol.*, **10**, 2570–2581.
- Wilm,M. and Mann,M. (1996) Analytical properties of the nano-electrospray ion source. *Anal. Chem.*, **68**, 1–8.
- Wu,J.Y. and Maniatis,T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.
- Zamore,P.D., Patton,J.G. and Green,M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature*, **355**, 609–614.
- Zhao,J., Hyman,L. and Moore,C. (1999) Formation of mRNA 3′-ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**, 405–445.
- Zhao,W. and Manley,J.L. (1996) Complex alternative RNA processing generates an unexpected diversity of poly(A) polymerase isoforms. *Mol. Cell. Biol.*, **16**, 2378–2386.
- Zuo,P. and Maniatis,T. (1996) The splicing factor U2AF35 mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.*, **10**, 1356–1368.

*Received July 28, 2000; revised September 1, 2000;
accepted September 5, 2000*