Biochemical characterization and three-dimensional structure analysis of the yeast cleavage and polyadenylation factor CPF

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Bertrand Paguet

Aus Frankreich

Basel, 2008

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Walter Keller and Prof. Andreas Engel.

Basel, den 22 mai 2007

Prof. Dr. Hans-Peter Hauri
Dekan der Philosophisch-
Naturwissenschaftlichen Fakultät

Table of contents

Abbreviations

3D:	three dimensions
APT:	associated with Pta1p complex
bp:	base pairs
CBP:	calmodulin binding protein
CF IA:	cleavage and factor IA
CF IB:	cleavage and factor IB
CF II:	cleavage and factor II
CFP:	cyan fluorescent protein
ChIP:	chromatin immunoprecipitation
CID:	CTD-interaction domain
CPF:	cleavage and polyadenylation factor
C-RBS:	C-terminal RNA-binding site
Cryo-EM:	cryo-electron microscopy
CTD:	C-terminal domain of the largest subunit (Rpb1p) of RNA polymerase II
DSE:	downstream element
EE:	efficiency element
EM:	electron microscopy
FRAP:	fluorescence recovery after photobleaching
FRET:	fluorescence resonance energy transfer
GFP:	green fluorescent protein
kDa:	kiloDalton
MALDI-TOF:	matrix assisted laser desorption ionization-time of flight
MDa:	megaDalton
mRNA:	messenger ribonucleic acid
MS:	mass spectrometry
MSA:	multivariate statistical analysis
NLS:	nuclear localization signal
NMD:	nonsense-mediated mRNA decay
NMR:	nuclear magnetic resonance
nt:	nucleotides

ORF:	open reading frame
PAN:	poly(A)-specific nuclease
PCR:	polymerase chain reaction
PE:	positioning element
PF I:	polyadenylation factor I
PP1:	type 1 serine/threonine protein phosphatase
Pre-mRNA:	pre-messenger RNA or primary transcript or mRNA precursor
ProtA:	protein A
RBD:	RNA-binding domain
RCT:	random conical tilt
RNA:	ribonucleic acid
RNAP II:	RNA polymerase II
RRM:	RNA recognition motif
S2:	CTD serine 2
S5:	CTD serine 5
S. cerevisiae:	*Saccharomyces cerevisiae*
SDS PAGE:	sodium dodecyl sulfate polyacrylamide gel electrophoresis
snoRNA:	small nucleolar RNA
snRNA:	small nuclear RNA
snRNPs:	small nuclear ribonucleoprotein
STEM:	scanning transmission electron microscopy
TAP:	tandem affinity purification
TEV:	tobacco etch virus
TRAMP:	Trf4p/Air1p/Mtr4p polyadenylation complex
TRP:	tryptophan
UTR:	untranslated regions
YFP:	yellow fluorescent protein
YPD:	yeast peptone dextrose
YPG:	yeast peptone galactose

Summary

In eukaryotes, protein-encoding genes are transcribed in the nucleus by RNA polymerase II (RNAP II). Before the gene transcript (pre-mRNA) is transported to the cytoplasm and can function in translation, it has to undergo three specific maturation steps: capping, splicing and 3' end processing. It is well established now that pre-mRNA processing events occur cotranscritpionally, while RNAP II is still transcribing the gene.

Pre-mRNA 3' end formation is an essential step in gene expression. With the exception of mRNAs coding for replication-dependent histone proteins, all eukaryotic pre-mRNAs are processed at their 3' end by a coupled two-step reaction that involves a specific endonucleolytic cleavage at the poly(A) site and subsequent poly(A) tail addition to the upstream cleavage fragment catalyzed by poly(A) polymerase (Pap1p). The complete maturation of the pre-mRNA 3' end is directed by the presence of *cis*-acting sequence elements on the pre-mRNA that recruit protein factors. Surprisingly, 3' end processing is accomplished by a complex protein machinery, which is highly conserved from yeast to mammals. Indeed, most of the polypeptides involved in 3' end formation in mammals have homologues in yeast. In *S. cerevisiae*, the cleavage and polyadenylation factor (CPF), cleavage factor IA (CF IA), cleavage factor IB (CF IB or Nab4p/Hrp1p) and the poly(A) binding protein (Pab1p) are required for specific and accurate 3' end processing activities.

The main pre-mRNA 3' end processing factor, CPF, is a multiprotein complex that consists of 15 polypeptides and is required for both pre-mRNA 3' end processing reactions. Most of its subunits are essential (only two are not essential) and all the components are involved in RNA recognition or protein-protein interactions within CPF or with other protein complexes, such as CF IA or RNAP II. Biochemical studies of all CPF subunits have confirmed that they belong to the complex and allowed characterization of their function in the context of pre-mRNA 3' end formation.

Affinity purification of CPF identified a novel protein that co-purified with the previously known components and was therefore proposed to be a new subunit of the complex. This putative new component is non-essential and was named Cpf11p. In this work, we showed that Cpf11p is not stably associated with CPF, as TAP tag purification of Cpf11p did not result in the purification of CPF components (Chapter

3). Furthermore, *in vitro* cleavage and polyadenylation assays performed with Cpf11p-depleted extracts did not show any defect, indicating that Cpf11p is not involved in pre-mRNA 3' end formation. We also found that expression of Cpf11p, in contrast to the expression of CPF subunits, is controlled in a sugar-dependent manner. Taken together, our results strongly suggest that Cpf11p has no function in pre-mRNA 3' end processing.

CPF plays a central role in pre-mRNA 3' end processing. Its requirement for cleavage and polyadenylation reactions is mediated by cooperative interactions with CF IA and recognition of *cis*-acting polyadenylation signals on the primary transcript. Despite every polypeptide involved in pre-mRNA 3' end processing was characterized, the mechanism by which the pre-mRNA is cleaved and polyadenylated is not known. One aim of this thesis is to provide insight into the three-dimensional structure of CPF. This would allow a better understanding in the arrangement of the subunits in the complex and provide information on the molecular mechanism of pre-mRNA 3' end processing. We have developed an efficient purification procedure that yields highly pure and active CPF. Here we report for the first time the 3D structure of the complex at a resolution of 25 Å, determined by single-particle electron microscopy on natively purified CPF using angular reconstitution and random conical tilt (Chapter 4). The 3D model reveals a rough globular shape of the complex and a strikingly large central cavity. We discuss the possibility that the inner cavity represents a reaction chamber in which pre-mRNA 3' end processing reactions could take place. Furthermore, we have determined the mass of CPF particles by scanning transmission electron microscopy (STEM) at approximately one megaDalton.

The work reported in this thesis should contribute to a better understanding of the mechanism by which pre-mRNAs are processed at their 3' end by presenting the first 3D model of the CPF complex, and by refining the protein composition of CPF. In addition, a part of this work is dedicated to investigation of *in vivo* interconnections between transcription and 3' end processing of pre-mRNAs (Chapter 2).

Chapter 1: Introduction

1.1 Three processing events lead to the maturation of messenger RNAs

In metazoans, protein-encoding genes are exclusively transcribed by the RNA polymerase II (RNAP II). The eukaryotic messenger RNA (mRNA) precursors (pre-mRNAs) are synthesized in the nucleus and have to systematically undergo three modification steps to become mature and functional: capping, splicing and 3' end processing. The mature mRNA is exported to the cytoplasm where it is translated into proteins by the ribosomes. The processing reactions do not occur independently in the nucleus but cotranscriptionally, in connection to RNAP II transcription.

The first processing event is capping that modifies the 5' triphosphate end of the primary transcript, by adding a 7-methyl guanine cap structure as soon as the nascent transcript has reached 20-25 nucleotides (nt) in length (Jove and Manley, 1984; Rasmussen and Lis, 1993). In both metazoans and unicellular eukaryotes, three enzymatic activities are required for cap formation (Figure 1.1). The 5' triphosphate end of the pre-mRNA is first hydrolyzed by a RNA 5' triphosphatase, generating a 5' diphosphate end that is then capped with GMP by a RNA guanylyltransferase, and finally methylated by a (guanine-N7) methyltransferase. In *Caenorhabditis elegans* and mammals, a bifunctional protein that contains both the RNA 5' triphosphatase and guanylyltransferase activities catalyzes the first two steps, whereas in the yeast *Saccharomyces cerevisiae* (*S. cerevisiae*), three separate proteins catalyze the three different steps: the RNA 5' triphosphatase Cet1p, RNA guanylyltransferase Ceg1p and (guanine-N7) methyltransferase Abd1p (for reviews, see Gu and Lima, 2005; Shatkin and Manley, 2000). 5' capping is an essential process and has important effects for mRNAs export and stability, translation initiation, and efficient splicing (Furuichi and Shatkin, 2000; Shibagaki et al., 1992).

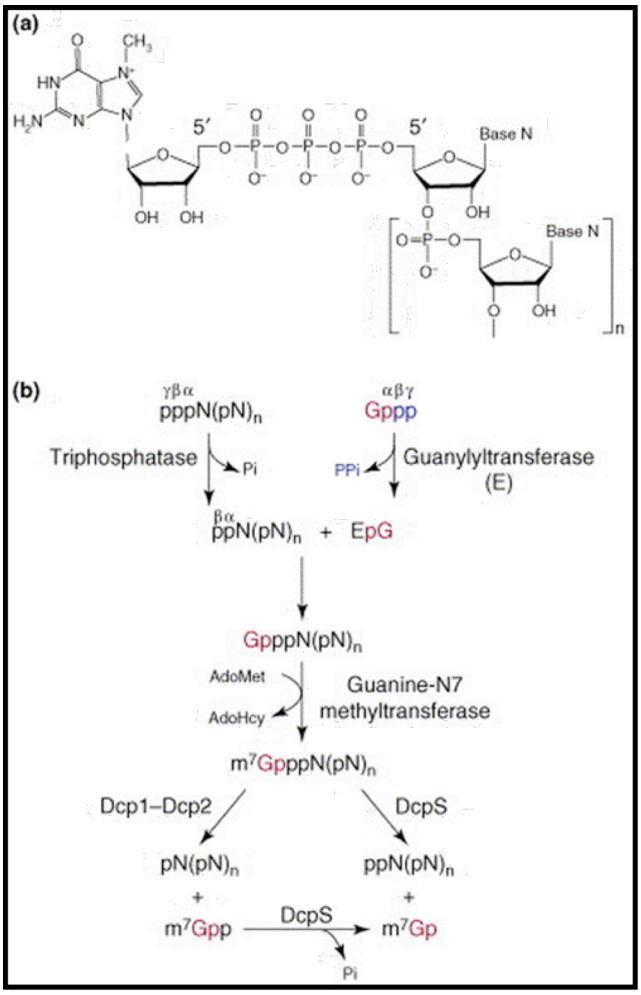

Figure 1.1. Messenger RNA cap structure and its metabolism (from Gu and Lima, 2005).
(A) Chemical structure of the mRNA cap. The cap consists of N7-methyl guanosine linked by an inverted 5'-5' triphosphate bridge to the first nucleoside of the mRNA.
(B) Synthesis and degradation of the mRNA cap (enzymes involved at each step are indicated next to the arrows. Dcps are decapping enzymes).

In eukaryotic pre-mRNAs, exons are interrupted by noncoding introns. The second processing event, splicing, removes noncoding introns from newly synthesized pre-mRNAs and joins the coding exon segments together in a two-step transesterification reaction (Kramer, 1996). During the first step, an adenosine residue located in the branchpoint sequence near the 3′ end of the intron, carries out a nucleophilic attack on the 5′ splice site (Figure 1.2). This reaction generates a free exon 1 and a lariat-exon 2 as splicing intermediates. During the second step, the exon 1 attacks at the 3′ splice site to generate the final splicing products, the spliced exons 1-2 and the lariat intron. These two catalytic steps are accomplished by the spliceosome, a large and dynamic multi-component complex consisting of the U1, U2, U4/U6, and U5 small nuclear ribonucleoproteins (snRNPs) and a multitude of non-snRNP proteins (Will and Luhrmann, 2001).

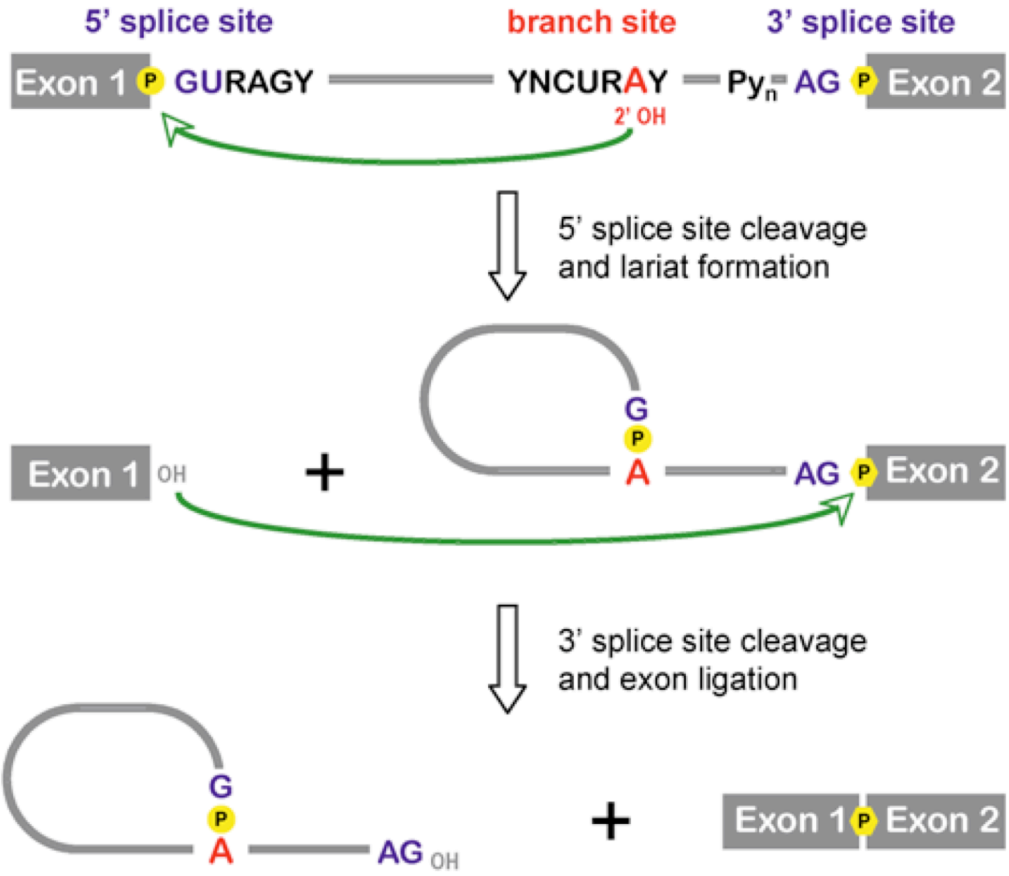

Figure 1.2. Schematic view of the two transesterification steps in pre-mRNA splicing. The splicing products are the exons joined together and the intron removed as a lariat. For explanation of the reactions, see the text.

Finally, to complete the series of pre-RNAs modifications, the majority of primary transcripts are processed at their 3' end by a coupled cleavage and polyadenylation reaction (Figure 1.3). The 3' end of mature mRNAs is generated by a site-specific endonucleolytic cleavage of an internal phosphodiester bond of the primary transcript. The upstream cleavage product generated is subsequently polyadenylated at its 3' hydroxyl end, whereas the downstream cleavage product is rapidly degraded (Wahle and Ruegsegger, 1999; Zhao et al., 1999a). In eukaryotic cells, the only known exceptions to this pathway are small nucleolar RNAs (snoRNAs) and the major histone mRNAs, which are endonucleolytically cleaved downstream of a stem-loop structure by the stem-loop binding protein (SLBP) but are not polyadenylated (Dominski and Marzluff, 1999; Perumal and Reddy, 2002). The length of the poly(A) tails depends on the species and reaches 250 nt in mammals,

whereas in yeast the poly(A) tail has an average length of about 70 nt. Poly(A) tails have important roles in mRNA stability and cytoplasmic export, and are also required for efficient translation initiation (Tarun and Sachs, 1996).

Pre-mRNA 3' end processing involves several protein factors that are well conserved from mammals to yeast, although their organization is different. In mammals, the protein factors are: the cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factors I_m and II_m (CF I_m and CF II_m), poly(A) polymerase (PAP) and nuclear poly(A) binding protein (PABPN1). In yeast, Pap1p, the enzyme that catalyzes the polyadenylation reaction, and Ysh1p, the putative endonuclease, both belong to the large multiprotein cleavage and polyadenylation factor (CPF). Other protein factors, such as cleavage factor IA (CF IA), cleavage factor IB (CF IB) and poly(A) binding protein (Pab1p) act in cooperation with CPF and recognize specific *cis*-acting sequences located on the pre-mRNA to accurately process its 3' end. 3' end processing reactions can be reconstituted *in vitro* by incubating protein cell extracts or purified factors with suitable RNA substrates (Butler and Platt, 1988; Butler et al., 1990).

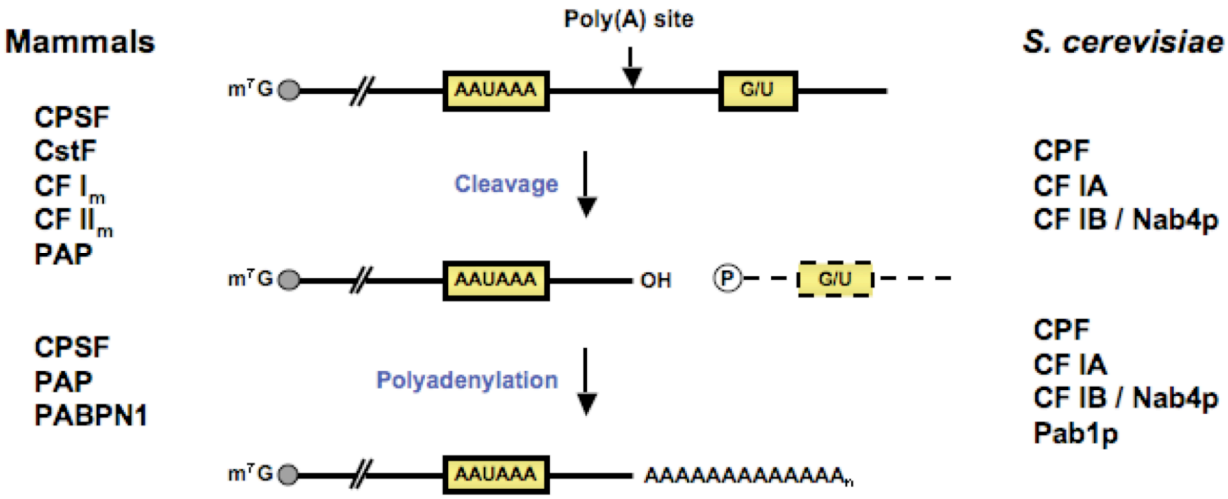

Figure 1.3. Pre-mRNA 3' end processing is a coupled two-step reaction.
The primary RNAP II transcript is endonucleolytically cleaved at the poly(A) site. Then, the 3'-hydroxyl end of the upstream cleavage product is polyadenylated, while the downstream cleavage product is degraded (dashed lines). The length of the poly(A) tails varies with the species to 70 nt in yeast and 250 nt in mammals. The protein factors responsible for each step of the reaction in mammals and in *S. cerevisiae* are indicated on the left and on the right respectively.

1.2 3' end processing of pre-mRNAs in mammals

3' end processing of pre-mRNAs is a seemingly simple two-step reaction that involves a site-specific endonucleolytic cleavage of the mRNA precursor followed by the addition of a poly(A) tail to the upstream cleavage product. Thus, only two enzymes are predicted to be necessary and sufficient to generate the 3' end of mature mRNAs. However, intensive biochemical characterization of polypeptides involved in 3' end processing in mammals and in *S. cerevisiae* has revealed an unexpectedly complex protein composition (for reviews, see Keller and Minvielle-Sebastia, 1997; Wahle and Ruegsegger, 1999; Zhao et al., 1999a). The development of *in vitro* processing assays allowed the identification of 3' end processing factors via fractionation of cell extracts from mammalian cells (Moore and Sharp, 1985) and yeast cells (Butler and Platt, 1988). Both organisms require multiprotein complexes to perform 3' end processing of pre-mRNAs and most of the mammalian polypeptides involved have yeast counterparts, indicating conservation during evolution (Keller and Minvielle-Sebastia, 1997). *Cis*-acting sequences located in the 3' untranslated region (3' UTR) of the primary transcript recruit the protein factors and guide them to the site of cleavage and polyadenylation. These sequence elements have been defined as conserved blocks of nucleotides situated in the proximity of the cleavage site.

1.2.1 Cleavage and polyadenylation signals in mammals

In mammalian cells, the *cis*-acting signals required for 3' end formation consist of a core of three sequence elements in the pre-mRNA (Figure 1.4).

The first element is the hexanucleotide AAUAAA that is located 10-30 nt upstream of the poly(A) site (also called cleavage site). This sequence is one of the most highly conserved sequence known and is found in almost all mRNAs (Proudfoot, 1991; Wahle and Kuhn, 1997). This hexanucleotide sequence is specifically bound by CPSF. AUUAAA is the only active variant that can also be found (Graber et al., 1999a). Mutagenesis studies have established the essential requirement of this sequence in both cleavage and polyadenylation (Manley et al., 1985; Wahle and Keller, 1992).

The downstream element (DSE) is the second element of the core

polyadenylation signal and is located 30 nt downstream of the poly(A) site. It is less conserved and two different types are found: a U-rich element with short stretches of U residues (Chou et al., 1994) and a GU-rich element with the consensus sequence YGUGUUYY (Y = pyrimidine). Point mutations or small deletions in the DSE have only weak effects, and larger deletions are required to abolish function. Nevertheless, the proximity of the DSE to the poly(A) site can affect the cleavage site position (MacDonald et al., 1994) and the efficiency of cleavage (Gil and Proudfoot, 1987). Importantly, the position of this element ensures that the transcript is not cleaved again, as it is removed during cleavage.

The third element is the poly(A) site itself. The distance between the hexanucleotide AAUAAA sequence and the DSE usually determines the selection of the cleavage site (Chen et al., 1995). In 70% of vertebrate mRNAs, endonucleolytic cleavage occurs downstream of an adenosine, whereas the preceding nucleotide is a C in 59% of the cases, defining the dinucleotide CA as the poly(A) site consensus (Sheets et al., 1990).

In addition, other sequences called auxiliary sequences can modulate the efficiency of the processing, but are not essential (Gilmartin et al., 1995). As an example, CF I_m was found to specifically bind to the sequence UGUAA surrounding the cleavage site (Brown and Gilmartin, 2003).

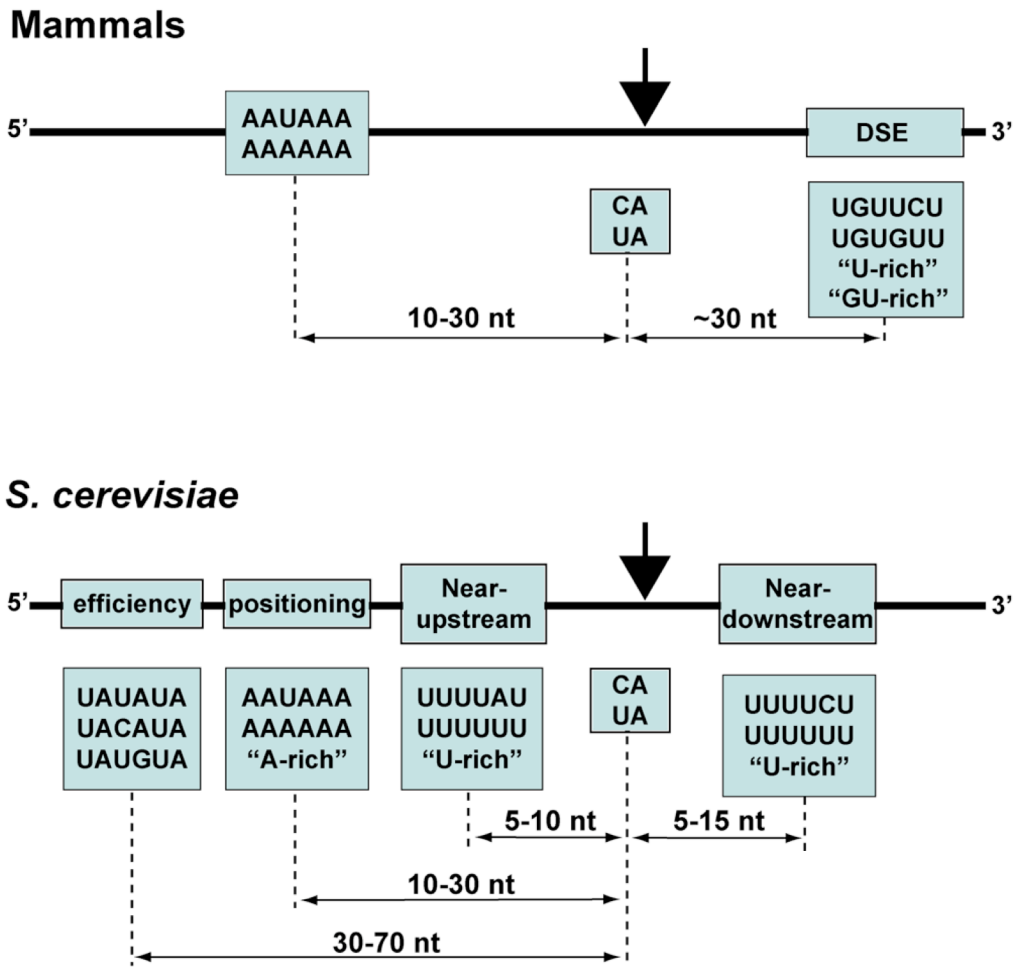

Figure 1.4. Polyadenylation regulatory signals in mammals and yeast.
AAUAAA sequence, downstream element (DSE), efficiency element, positioning element, near-upstream and near-downstream elements are indicated in boxes on the RNA (black line). The boxes below contain examples of functional sequences for each of the elements. The distances between the different sequence elements relative to the poly(A) site are indicated below. The black arrow marks the position of cleavage and polyadenylation.

1.2.2 Cleavage and polyadenylation factors in mammals

By fractionating cell extracts, six protein factors have been identified to be involved in 3' end processing of pre-mRNAs in mammals (Figure 1.5). Cleavage of mammalian pre-mRNA substrates requires the combination of five of those factors (Christofori and Keller, 1988; Takagaki et al., 1989): CPSF, CstF, CF I_m , CF II_m and PAP. For polyadenylation, CPSF, PAP and nuclear PABPN1 are required.

1.2.2.1 Cleavage and polyadenylation specificity factor

CPSF specifically binds to the highly conserved hexanucleotide AAUAAA signal located 10-30 nt upstream of the cleavage site. Purification of this complex from HeLa cells and calf thymus has led to the identification of five subunits: CPSF-160, CPSF-100, CPSF-73, CPSF-30 and hFip1, which all have homologues in yeast (Barabino et al., 1997; Bienroth et al., 1991; Jenny et al., 1994; Kaufmann et al., 2004; Murthy and Manley, 1995).

The biggest subunit CPSF-160 is responsible for the interaction with the AAUAAA sequence on the pre-mRNA (Jenny et al., 1994), and in addition interacts with PAP and the 77 kDa subunit of CstF (Murthy and Manley, 1995).

CPSF-100 and CPSF-73 share similarities with the metallo-β-lactamase / β-CASP (named for metallo-β-lactamase, CPSF, Artemis, Snm1, Pso2) enzymes involved in the cleavage of nucleic acids (Callebaut et al., 2002). Nevertheless, CPSF-100 has no endonucleolytic activity, as a part of the conserved motif necessary for activity is lacking. The function of CPSF-100 is still unknown. For a long time, CPSF-73 has been suspected to be the endonuclease responsible for cleavage of pre-mRNAs (Ryan et al., 2004). However, the first direct experimental evidence for endonucleolytic activity was provided in a recent publication together with a crystal structure of CPSF-73 at a 2.1 Å resolution (Mandel et al., 2006).

CPSF-30 contains five zinc-finger domains and a zinc knuckle (DuBois et al., 1990), both being involved in RNA-binding (Barabino et al., 1997). CFSF-30 was shown to bind to poly(U) sequences of RNA (Barabino et al., 1997).

hFip1 was recently identified as a component of CPSF (Kaufmann et al., 2004). It directly interacts with CPSF-30, CstF-77 and PAP. Furthermore, it binds preferentially to U-rich sequences on the pre-RNA and contributes to CPSF-mediated stimulation of PAP activity (Kaufmann et al., 2004).

1.2.2.2 Cleavage stimulation factor

Composed of the three polypeptides CstF-77, CstF-64 and CstF-50, the CstF factor binds to the U-rich or GU-rich sequence elements downstream of the cleavage site of the pre-mRNA (MacDonald et al., 1994; Takagaki et al., 1990).

CstF-77 contains HAT repeats that mediate protein-protein interactions (Colgan and Manley, 1997). Consistently, CstF-77 bridges the other CstF subunits and directs interaction with CPSF-160, possibly stabilizing the CPSF-CstF-RNA complex (Murthy and Manley, 1995).

CstF-64 has an RNA-binding domain that binds the DSE of pre-mRNAs (MacDonald et al., 1994), and interacts with symplekin (Hofmann et al., 2002).

CstF-50 is composed of seven WD-40 repeats, which were shown to bind the phosphorylated C-terminal domain (CTD) of the largest subunit of RNAP II (McCracken et al., 1997b).

1.2.2.3 Cleavage factors I_m and II_m

CF I_m is composed of the 25 kDa, 59 kDa and 68 kDa subunits and possibly a fourth subunit of 72 kDa (Ruegsegger et al., 1996). CF I_m has no known homologues in yeast. CF I_m 59 kDa and CF I_m 68 kDa subunits contain three distinct domains: a RNP-type RNA-binding domain at the N-terminus, a central proline-rich domain and a C-terminal RS-like domain similar to that of the SR proteins involved in splicing. The 25 kDa subunit binds to RNA and was shown to associate with the largest subunit of CF I_m, with PAP and with PABNP1 (Dettwiler et al., 2004; Kim and Lee, 2001). The 68/25 kDa heterodimer was shown to bind the sequence UGUAA that surrounds the cleavage site on the pre-mRNA (Brown and Gilmartin, 2003). The nature of the 72 kDa protein is still unclear. Furthermore, CF I_m was shown to increase stability of the CPSF-CstF-RNA complex (Ruegsegger et al., 1996).

During purification of CF II_m from HeLa cells the activity could be separated into a CF IIA_m fraction and a CF IIB_m fraction. The essential CF IIA_m contained hClp1 and hPcf11, two proteins with yeast homologues, whereas CF IIB was shown to be non-essential but stimulatory for cleavage (de Vries et al., 2000).

hPcf11 has a DSI consensus motif, which was shown to bind the CTD in yeast (Sadowski et al., 2003). hPcf11 also contains two zinc finger motifs and 30 repeats of the consensus sequence LRFDG. In addition, hPcf11 was shown to play a role in cleavage, but is dispensable for polyadenylation.

hClp1 is evolutionary conserved and contains Walker A and B motifs, suggesting that hClp1 is a nucleotide binding protein (de Vries et al., 2000).

Furthermore, hClp1 is involved in the cleavage reaction. Recently, hClp1 was proposed to be a novel human RNA-kinase that phosphorylates siRNAs and 3' exons during tRNA splicing (Leuschner et al., 2006).

1.2.2.4 Poly(A) polymerase

PAP is the enzyme that catalyzes the polyadenylation reaction on cleaved pre-mRNAs and is therefore a key player in 3' end formation of mRNAs in eukaryotes. The N-terminal part of the protein is highly conserved and contains the catalytic domain that is characteristic for the superfamily of nucleotidyltransferases (Holm and Sander, 1995; Martin and Keller, 1996). The crystal structure of the bovine PAP was solved at a resolution of 2.5 Å (Martin et al., 2000). The catalytic site contains three conserved aspartate residues that are absolutely essential for activity. Furthermore, a RNA-binding domain is present in the C-terminal part of the polymerase, overlapping with a region needed for AAUAAA-dependent activity (Thuresson et al., 1994). The C-terminal part of PAP also contains a bipartite nuclear localization signal (NLS), and a region rich in serine and threonine residues allowing regulation of PAP activity by phosphorylation (Abuodeh et al., 1998; Colgan et al., 1996). PAP alone has a low level of activity, which is enhanced when manganese is substituted for magnesium (Wahle, 1991). PAP specifically uses ATP and has no specificity for RNA. Finally, PAP specifically polyadenylates RNAs containing the AAUAAA sequence in the presence of CPSF and is also required for the cleavage of most pre-mRNAs.

1.2.2.5 Nuclear poly(A) binding protein

CPSF and PAP are sufficient to allow poly(A) addition to the precleaved RNA substrate. However, fast and processive activity requires PABPN1, which also controls poly(A) tail length (Bienroth et al., 1993). This small protein of 33 kDa contains a very acidic N-terminal domain, a very basic C-terminal domain and a single RNP domain in the center. PABPN1 interacts with CPSF-30 in vitro (Chen et al., 1999) and specifically binds to poly(A) and poly(G) with a binding site length of 9 nt (Nemeth et al., 1995).

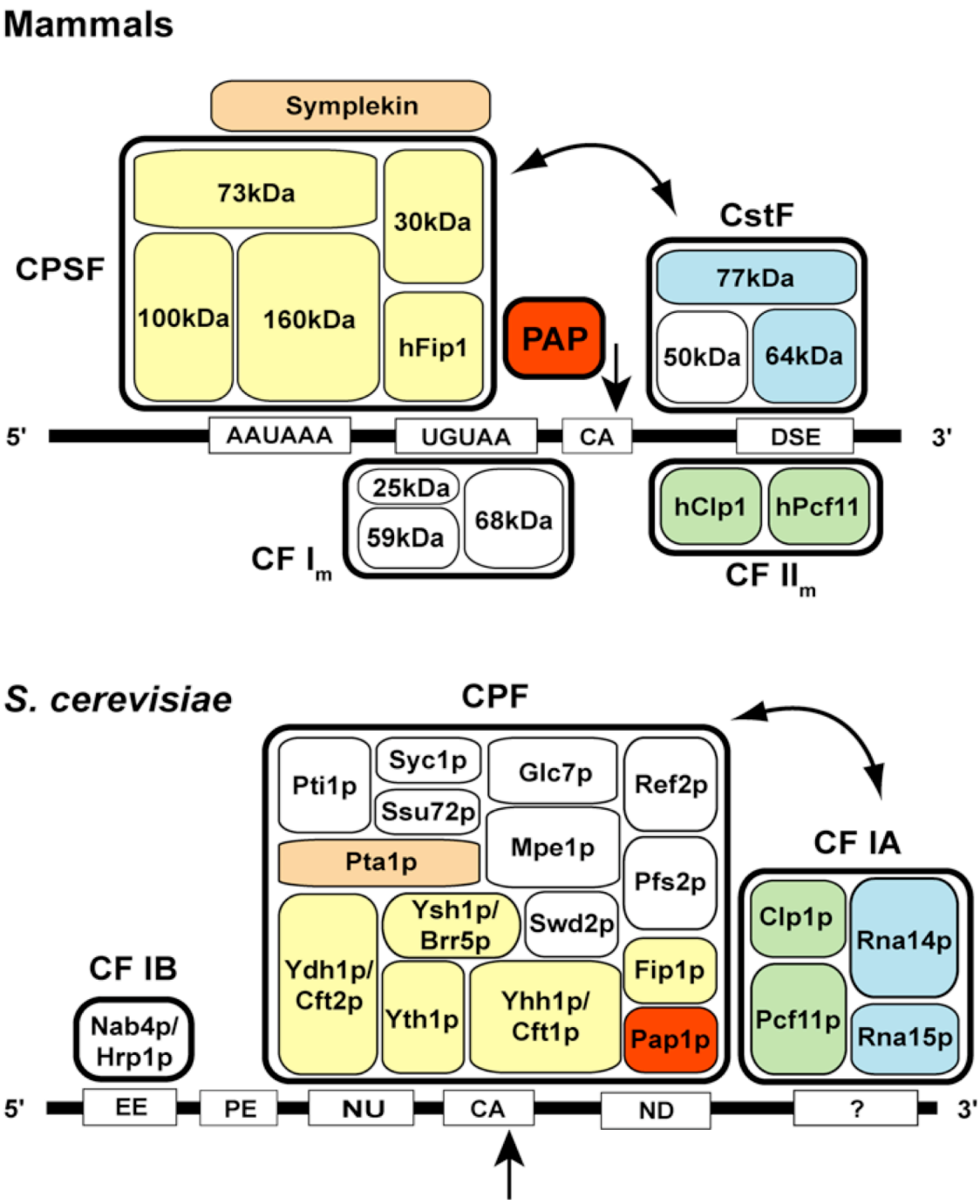

Figure 1.5. Schematic representation of the mammalian and yeast 3' end processing machineries.
Multiprotein complexes and their subunit composition are represented with their respective points of contact to the poly(A) signal elements (white rectangles) of the pre-mRNA (black line). Homologous factors between mammals and yeast are colour coded. Proteins in white have no known homologues. Site of cleavage is denoted by a black arrow. Double-headed arrows represent contacts between proteins of different factors. CA: cleavage site; DSE: downstream sequence element; EE: efficiency element; PE: positioning element; NU: near upstream element; ND: near downstream element; ?: putative signal element.

1.3 3’ end processing of pre-mRNAs in yeast

1.3.1 *Cis*-acting signals required for pre-mRNA 3’ end processing in yeast

In contrast to mammalian polyadenylation signals that are highly conserved, *cis*-acting signals in pre-mRNA 3’ end processing in the yeast *S. cerevisiae* are more degenerate and redundant. Consequently, analysis of *cis*-acting regulatory sequences has been more complicated than in mammalian cells. Nevertheless, mutagenesis studies and computational analysis of a large set of known and predicted processing sites have allowed the definition of a general pattern of *cis*-acting signals required for 3’ end processing (Graber et al., 1999b). Five blocks of sequences have been defined as follow (Figure 1.4): (i) The efficiency element (EE); (ii) The positioning element (PE); (iii) The near upstream element; (iv) The near downstream element; (v) The poly(A) site (or cleavage site) itself. It has been suggested that these *cis*-acting signals act cooperatively, since a deletion of one element only reduced the efficiency of 3’ end processing (Beyer et al., 1997; Zhao et al., 1999a). Thus, the presence of one weak signal in the 3’ UTR of a transcript can be compensated by stronger signals in other sequence elements.

The EE is typically found between 35 and 70 nt upstream of the cleavage site. The consensus sequence for this element was identified by mutagenesis studies and computational analysis as UAUAUA, as the most efficient sequence to direct 3’ end processing (Guo et al., 1995), but can be variable with U-rich stretches (Guo and Sherman, 1995). The U residues at the first and fifth positions are the most critical nucleotides in this sequence. The EE enhances cleavage and has been shown to have a more important role than the PE in the cleavage site selection (Guo and Sherman, 1996). Finally, the EE is specifically bound by CF IB, which acts in cleavage site selection (Minvielle-Sebastia et al., 1998) and functions to activate the PE.

The PE is an “A-rich” element that is typically found between 10 and 30 nt upstream of the cleavage site. As the EE, the sequence of the PE is variable. However, the most efficient motifs were found to be AAUAAA and AAAAAA. The PE is present in all organisms and directs the position of the cleavage site. It also contributes to the efficiency of the processing as single point mutations or deletion of

the motif resulted in the reduction in the use of the cleavage site (Abe et al., 1990; Hyman et al., 1991).

The near upstream and downstream elements occur within 10 nt upstream and downstream of the cleavage site (Graber et al., 1999a). They are best characterized as "U-rich elements", as the most common sequences immediately surrounding the cleavage site are long runs of uracils (UUUUUU or UUUUCU, etc). They were found to contribute to cleavage site selection and to enhance cleavage *in vitro* (Dichtl and Keller, 2001). Surprisingly, they don't have counterparts in animals (Graber et al., 1999a). Yhh1p, Ydh1p and Yth1p, three components of CPF have been shown to bind to these U-rich elements (Barabino et al., 2000; Dichtl et al., 2002a; Dichtl and Keller, 2001; Keller et al., 1991).

Mapping of the poly(A) sites of numerous genes allowed the definition of the consensus sequence as a dinucleotide that consists of a pyrimidine followed by one or more A residues (Heidmann et al., 1994; Russo et al., 1993). Cleavage takes place on the 3' side of an adenosine residue. In contrast to animal genes that use a single poly(A) site downstream of the hexanucleotide AAUAAA, numerous yeast genes use a cluster of poly(A) sites (Zhao et al., 1999a). Thus, even if the dinucleotide CA is the sequence of choice for cleavage, adjacent sub-optimal poly(A) sites, also called cryptic cleavage sites, can be used if the dinucleotide CA is mutated or destroyed. T-rich motifs are frequently found immediately before and after the poly(A) site (Graber et al., 1999b).

It is unclear whether yeast uses a specific motif downstream of the cleavage site, as in higher eukaryotes. Very few genes require a downstream sequence for efficient processing *in vivo* as exemplified by the *ADH2* transcript (Hyman et al., 1991). Nevertheless, in most of the cases, deletions of the downstream sequences have no effect (Aranda et al., 1998).

Thus, the polyadenylation signals in yeast are degenerate and redundant, as described by the sequence flexibility of the main motifs. The redundancy of the sequences provides an explanation to the fact that mutations or deletions in these motifs have only weak effects in 3' end processing. Furthermore, the *cis*-acting signals are believed to act in cooperation, as shown for some genes (like the *CYC1* gene) that use multiple weak signals to act additively to provide a strong signal (Guo et al., 1995). The presence of secondary structures in the primary transcripts may also be important, at least in some genes (Sadhale and Platt, 1992). The role of putative

RNA folding has not been intensively investigated, but together with redundancy of the regulatory sequences, it could well contribute to the complexity of polyadenylation signals in yeast.

1.3.2 *Trans*-acting factors involved in pre-mRNA 3' end processing in *S. cerevisiae*

In the yeast *S. cerevisiae*, cleavage requires CPF, CF IA and CF IB. In addition to these protein factors, Pab1p is required for specific polyadenylation. Although the polyadenylation *cis*-acting signals used by mammals and yeast are different in their sequence and organization, protein factors of the 3' end processing apparatus of both organisms exhibit high conservation (Figure 1.5 and Table 1.1).

Fractionation of yeast cell extracts have allowed the identification of the different factors involved in pre-mRNA 3' end processing (Chen and Moore, 1992; Dichtl et al., 2002b; Kessler et al., 1996; Nedea et al., 2003). Reconstitution of 3' end processing assays with partially purified chromatographic fractions have initially defined five protein factors: cleavage factors I and II (CF I and CF II), polyadenylation factor PF I, poly(A) polymerase (Pap1p) and poly(A) binding protein (Pab1p). CF II and PF I had several polypeptides in common and were later shown to be part of the same functional unit renamed CPF (Ohnacker et al., 2000). Furthermore, analysis of the protein composition of CPF has revealed the presence of Pap1p, which was shown to be much more active in the context of CPF than as a separate polypeptide. In addition, further purification of CF I led to the separation into CF IA and CF IB (Kessler et al., 1996). Thus, the contemporary view of the yeast 3' end processing machinery involves the four protein factors CPF, CF IA, CF IB and Pab1p. The large majority of genes encoding for 3' end processing polypeptides is essential for cell viability, confirming the essential function of their products in pre-mRNA 3' end formation. All of these factors have now been purified to homogeneity and are described in details below.

Factors	Size (kDa)	Genes	Sequence features	Mammalian homologue (% identity)
Cleavage factor IA **CF IA**	76	*RNA14*	WXXY	CstF-77K (24%)
	72	*PCF11*	CID	hPcf11 (26%)
	50	*CLP1*	P-loop	hClp1 (26%)
	38	*RNA15*	RBD	CstF-64K (43%)
Cleavage factor IB **CF IB**	73	*NAB4/HRP1*	RBD	hnRNPA1
Cleavage and polyadenylation factor **CPF**	150	*YHH1/CFT1*	RBD	CPSF-160K (24%)
	105	*YDH1/CFT2*		CPSF-100K (25%)
	100	*YSH1/BRR5*	HXHXDH	CPSF-73K (53%)
	85	*PTA1*		Symplekin
	64	*PAP1*	Poly(A) polymerase	PAP (47%)
	60	*REF2*		
	58	*MPE1*	Zn^{2+}-knuckle	hMpe1
	55	*FIP1*		hFip1
	53	*PFS2*	WD-40	hPfs2
	47	*PTI*		
	37	*SWD2*	WD-40	hSwd2 (35%)
	36	*GLC7*	Ser/Thr-phosphatase	PP1 (84%)
	26	*YTH1*	Zn^{2+}-finger	CPSF-30K (40%)
	23	*SSU72*	CTD-phosphatase	hSsu72 (54%)
	21	*SYC1*		

Table 1.1. Yeast pre-mRNA 3' end processing factors are evolutionary conserved.
Size (in kDa), sequence features and mammalian homologues are indicated.

1.3.2.1 The cleavage and polyadenylation factor

Comprising 15 polypeptides, including both enzymes that catalyze the cleavage and polyadenylation steps, CPF is the main 3' end processing factor. It is the yeast counterpart of the mammalian CPSF factor involved in the recognition of the highly conserved AAUAAA hexamer sequence of pre-mRNAs. CPF composition was analyzed by affinity purification combined to mass spectrometry (MS) procedures (Dichtl et al., 2002b; Gavin et al., 2002; Nedea et al., 2003) and was shown to contain the following proteins, from the largest to the smallest subunit: Yhh1p/Cft1p (150 kDa), Ydh1p/Cft2p (105 kDa), Ysh1p/Brr5p (100 kDa), Pta1p (85 kDa), Pap1p (64 kDa), Ref2p (60 kDa), Mpe1p (58 kDa), Fip1p (55 kDa), Pfs2p (53 kDa), Pti1p (47 kDa), Swd2p (37 kDa), Glc7p (36 kDa), Yth1p (26 kDa), Ssu72 (23 kDa), Syc1p (20 kDa). Recently, it was proposed that Swd2p, Glc7p, Ref2p, Pti1p, Ssu72p, Syc1p and Pta1p can form a distinct unit called APT (associated with Pta1p) that would

independently direct snoRNAs (also transcribed by RNAP II) 3' end processing (Nedea et al., 2003). Therefore, CPF can be divided in two subcomplexes, the APT and the core-CPF that contains all the other subunits.

Yhh1p, the largest subunit of CPF, is the homologue to the mammalian AAUAAA interacting protein CPSF-160. Both proteins share 24% identity and 51% similarity in the amino acid sequence level (Stumpf and Domdey, 1996). Yhh1p is a RNA-binding protein that binds the pre-mRNA at U-rich sequences surrounding the poly(A) site and is involved in poly(A) site recognition (Dichtl et al., 2002a). The RNA-binding domain is composed of β-propeller repeats in the central region of the protein. Furthermore, Yhh1p specifically interacts with the phosphorylated CTD, and was therefore proposed to have a role in the coupling of 3' end formation and transcription (Dichtl et al., 2002a). In addition, Yhh1p has been shown to interact directly with almost all of the CPF and CF IA subunits. Finally, in cell free extracts immunodepleted with antibodies directed against Yhh1p and extracts of temperature-sensitive *yhh1* mutant alleles, both cleavage and polyadenylation reactions were abolished.

Ydh1p shares significant homology to CPSF-100 (24% identity and 43% similarity), but is also related to Ysh1p and CPSF-73 (Jenny et al., 1996). Ydh1p does not contain any known RNA-binding domains. Nevertheless, Ydh1p was successfully UV cross-linked to wild type pre-mRNAs in the presence of ATP, but not to precleaved RNAs or substrates lacking the EE (Zhao et al., 1997). Furthermore, this protein was shown to bind with high affinity at U-rich sequences surrounding the cleavage site and to the EE to a weaker extent (Dichtl and Keller, 2001). Ydh1p was shown to interact with several 3' end processing polypeptides but also with the CTD of RNAP II (Kyburz et al., 2003). Recently, the crystal structure of Ydh1p at 2.5 Å resolution was reported (Mandel et al., 2006).

Ysh1p is the yeast homologue of mammalian CPSF-73 (23% identity and 48% similarity), and shares up to 53% identity in the first 500 amino acids (Chanfreau et al., 1996; Jenny et al., 1996). It is essential for both cleavage and polyadenylation activities *in vivo* and *in vitro*. Ysh1p contains a β-CASP motif and is therefore a member of the metallo-β-lactamase superfamily that comprises enzymes acting on nucleic acid substrates (Callebaut et al., 2002). Supported by the facts that β-lactamase motifs are common to metal-dependent hydrolases, and that CPSF-73 has

been reported to be the mammalian 3' end processing endonuclease (Mandel et al., 2006; Ryan et al., 2004), Ysh1p has been proposed to be the endonuclease that acts on yeast pre-mRNAs. However, a direct proof of nuclease activity is still missing.

Pta1p has been originally identified as protein encoded by a gene affecting pre-tRNA processing (O'Connor and Peebles, 1992) and has been shown to be involved in both cleavage and poly(A) addition reactions (Zhao et al., 1999b). The phosphorylation status of Pta1p is regulated by the phosphatase Glc7p and influences polyadenylation activity of CPF: phosphorylated Pta1p inhibits poly(A) addition, whereas the unphosphorylated form does not (He and Moore, 2005). Furthermore, addition of Pta1p can restore polyadenylation activity in defective Glc7-depleted extracts (He and Moore, 2005).

Pap1p belongs to the large superfamily of nucleotidyl transferases and was the first 3' end processing factor to be purified (Lingner et al., 1991a; Lingner et al., 1991b). Pap1p is the enzyme that catalyzes polyadenylation of cleaved pre-mRNAs. The mammalian and yeast proteins share 47% identity in the first 400 amino acids, a region that comprises the catalytic domain containing three aspartate residues highly conserved across species (Martin et al., 2000). The yeast Pap1p lacks a C-terminal domain that carries a nuclear localization signal and a serine/threonine rich domain in mammals. However, an RNA-binding site (C-RBS) at the C-terminus is essential for activity (Zhelkovsky et al., 1998). Moreover, Pap1p contains two Fip1p binding sites, one of them overlapping with C-RBS and allowing Fip1p to regulate the processivity of the polymerase (Zhelkovsky et al., 1998). Two other binding sites also exist, one being involved in the recognition of the last three nucleotides of the RNA primer, the other one helping the enzyme to discriminate against deoxyribonucleotide substrates. Interestingly, studies have shown that Pap1p undergoes phosphorylation and ubiquitination during the S-G2 transition of the cell cycle, and that the phosphorylated form of the enzyme was inactive (Mizrahi and Moore, 2000). In contrast to the mammalian system, Pap1p is not involved in the cleavage step (Mandart and Parker, 1995).

Ref2p is one of the only two non-essential polypeptides of the 3' end processing machinery. Ref2p is an RNA-binding protein that is directly involved in the cleavage step of 3' end processing (Russnak et al., 1995) and plays the role of negative regulator of poly(A) synthesis (Mangus et al., 2004). Besides its functions in pre-mRNA 3' end processing, Ref2p is an essential actor in snoRNAs 3' end

formation (Dheur et al., 2003). Finally, Ref2p was proposed to recruit the phosphatase Glc7p, and Swd2p in CPF (Nedea et al., 2003).

Mpe1p is an evolutionary conserved protein that is exclusively involved in polyadenylation of pre-mRNAs (Vo et al., 2001). The *MPE1* gene genetically interacts with *PCF11* and the protein may contribute to the poly(A) site selection (Vo et al., 2001). Furthermore, Mpe1p contains a RNA-binding zinc knuckle motif (CX_2 CX_4 HX_4 C). Although this zinc finger motif has been implicated in the interaction between proteins and single-stranded nucleic acids, its deletion had no effect on the capability of Mpe1p to bind RNAs.

Fip1p was originally identified as a protein interacting with Pap1p and is exclusively involved in polyadenylation (Preker et al., 1995). Indeed, Fip1p has a key regulatory function in polyadenylation by inhibiting the processive activity of Pap1p through multiple interactions (Helmling et al., 2001). Association of Fip1p with Pap1p inhibits polyadenylation activity by limiting access of the RNA substrate to the C-RBS of the polymerase. Individual functional domains of Fip1p have been identified. The Fip1p domain responsible for binding and inhibition of Pap1p comprises the amino acids 80-105, which are essential for cell viability. Furthermore, the amino acids 206-220 are responsible for interaction with Yth1p and for polyadenylation of the mRNA precursor. Finally, the region comprising amino acids 105-206 promotes limitation of RNA-binding to the C-RBS of Pap1p. Another feature of Fip1p is that it helps the recruitment of Pap1p to CPF by interacting with Yth1p (Barabino et al., 2000). Interactions with Ysh1p, Psf2p and Rna14p have also been reported (Barabino et al., 2000; Ohnacker et al., 2000; Preker et al., 1995). Finally, it was found that Fip1p can be phosphorylated by the creatine kinase CK2 (Zielinski et al., 2006), and that phosphorylation decreased the association of Fip1p with CPF (He and Moore, 2005).

Pfs2p is a 53 kDa protein that contains seven WD-40 repeats between amino acids 90 and 380 (Ohnacker et al., 2000). WD repeats, also called transducin repeats, are found in many polypeptides that are involved in diverse cellular processes and that often belong to multiprotein complexes (Smith et al., 1999). Although removal of the C-terminal extension following the last WD repeat is dispensable for cell viability, deletion of one WD repeat is lethal (Ohnacker et al., 2000). The same authors demonstrated that Pfs2p directly interacts with subunits of CPF and CF IA, and thus proposed that Pfs2p functions in the assembly and stabilization of the 3' end

processing apparatus. Consistently, it is essential for both cleavage and poly(A) addition activities.

Pti1p was originally identified by two-hybrid screen interacting with Pta1p as bait and shows significant similarities to the CF IA subunit Rna15p and the mammalian CstF-64 (W. Hübner, unpublished data). Pti1p contains a RNA recognition motif and directly binds Pta1p and Rna14p (Skaar and Greenleaf, 2002). It is not essential for 3' end processing activities, nevertheless, mutations were shown to affect the cleavage site choice (Dheur et al., 2003; Skaar and Greenleaf, 2002). Furthermore, Pti1p plays an essential role in snoRNAs 3' end formation and has been shown to uncouple cleavage and polyadenylation upon overexpression (Dheur et al., 2003). Finally, Pti1p seems to be phosphorylated upon cell growth arrest (W. Hübner, unpublished data).

Swd2p contains seven WD-40 repeats and is associated with two different multiprotein complexes: CPF and the SET1 complex (SET1C) that methylates lysine 4 of histone 3 (H3-K4). In the context of the SET1C, Swd2p is required for H3-K4 di and trimethylation that acts in regulation of gene expression (Dichtl et al., 2004; Jenuwein and Allis, 2001). In the context of CPF, Dichtl et al. (2004) have found that Swd2p was required for 3' end formation of specific mRNAs and snoRNAs. In contradiction, Cheng et al. (2004) did not find direct involvement in cleavage and polyadenylation (Cheng et al., 2004). Furthermore, it has been proposed that the roles of Swd2p in 3' end formation and in histone tail modifications are independent (Dichtl et al., 2004).

Glc7p is a type 1 serine/threonine protein phosphatase (PP1), for which it is the only isoform in yeast (Feng et al., 1991). Amino acid sequences of catalytic subunits of PP1 are highly conserved across many species (Ramaswamy et al., 1998). Like mammalian PP1 catalytic subunits, Glc7p regulates many physiological processes, including glycogen metabolism, glucose repression, transcription, membrane fusion, mitosis, sporulation, ion homeostasis, and cell wall organization (Peggie et al., 2002; Stark, 1996; Tan et al., 2003; Williams-Hart et al., 2002). Recently, Glc7p has been shown to have an important role in mRNA export by dephosphorylating Npl3p, a factor involved in mRNA transport (Gilbert and Guthrie, 2004). In the context of pre-mRNA 3' end processing, the requirement of the phosphatase for polyadenylation but not for cleavage has been demonstrated (He and Moore, 2005). Furthermore, phosphorylated Pta1p was found to be a substrate of

Glc7p within the CPF complex (He and Moore, 2005). In addition, the presence of Glc7p within CPF has suggested that the activity of the complex could be regulated by cycles of phosphorylation and dephosphorylation (He and Moore, 2005). Additional studies also highlighted the role of Glc7p in snoRNA transcription termination (S. Röck, unpublished data).

Yth1p contains five zinc fingers of the C_3-H type, among which the second and fourth zinc fingers were shown to be essential (Barabino et al., 2000; Tacahashi et al., 2003), and is 40% identical to the mammalian 30 kDa polypeptide of CPSF (Barabino et al., 1997). Yth1p is an essential RNA-binding component of the cleavage and polyadenylation factor and binds to the pre-mRNA in the vicinity of the cleavage site (Barabino et al., 2000). Yth1p also binds directly to Fip1p and Ysh1p and is essential for both steps of pre-mRNA 3'-end processing (Barabino et al., 2000; Barabino et al., 1997).

Ssu72p was initially identified based on genetic and physical interaction with the general transcription factor TFIIB and the Rpb2p subunit of RNAP II (Dichtl et al., 2002b; Pappas and Hampsey, 2000; Sun and Hampsey, 1996). Ssu72p is involved in transcription termination of both pre-mRNAs and snoRNAs (Ganem et al., 2003; Nedea et al., 2003; Steinmetz and Brow, 2003) and affects start site selection during transcription initiation. Furthermore, Ssu72p is directly involved in pre-mRNA cleavage but not in polyadenylation (He et al., 2003). In addition, evidence was provided that Ssu72p is a CTD phosphatase specific for phosphorylated serine 5 (Ganem et al., 2003; Krishnamurthy et al., 2004). Chromatin immunoprecipitation experiments also revealed that Ssu72p is present at both terminator and promoter regions of RNAP II genes (Nedea et al., 2003). Exploiting these observations, Ssu72p and its phosphatase activity were proposed to mediate gene looping (juxtaposition of promoter and terminator regions) of some yeast genes (Ansari and Hampsey, 2005).

Together with Ref2p (see above), Syc1p is one of the two non-essential protein involved in 3’ end processing in yeast (Gavin et al., 2002; Nedea et al., 2003). Its amino acid sequence is highly similar to the C-terminal region of Ysh1p that is essential for cell viability (35% identity and 52% similarity over 199 amino acids). The presence of Syc1p was shown to exacerbate the processing and growth defects of many processing mutants, suggesting a role in regulating 3’ end formation efficiency (Zhelkovsky et al., 2006).

1.3.2.2 The cleavage factor IA

CF I was originally identified as a factor required for both the cleavage and polyadenylation reactions (Chen and Moore, 1992), and was further separated into two factors, CF IA and CF IB (Kessler et al., 1996). CF IA contains four essential subunits: Rna14p (76 kDa), Pcf11p (72 kDa), Clp1p (50 kDa) and Rna15p (38 kDa). CF IB consists of the single polypeptide Nab4p. Numerous protein-protein interactions have been reported between CF IA and CPF, indicating a high degree of crosstalk between these two complexes.

Rna14p and Rna15p are tightly associated, forming an heterodimeric complex required for RNA-binding (Noble et al., 2004), and are essential for both steps of 3' end formation (Kessler et al., 1997). Rna14p has sequence homology to mammalian CstF-77 (24% identity) and contains 10 HAT repeats (Takagaki and Manley, 1994). Rna14p interacts with Pcf11p and Nab4p (Gross and Moore, 2001), but also with the CPF subunits Pfs2p and Yhh1p (Kyburz et al., 2003; Ohnacker et al., 2000).

Rna15p is the yeast homologue of CstF-64 (Takagaki and Manley, 1994) and contains a canonical single RNA recognition motif (RRM) at its N-terminus and a C-terminal region that is essential for polyadenylation and transcription termination (Minvielle-Sebastia et al., 1994). Although the protein has affinity for U-rich sequences and can be UV cross-linked to RNA substrate (Kessler et al., 1996; Minvielle-Sebastia et al., 1991; Takagaki and Manley, 1994), the sequence that is recognized by Rna15p is still unclear. However, UV cross-linked experiments suggested that Rna15p binds to the PE in the presence of Rna14p and Nab4p/Hrp1p (Gross and Moore, 2001).

Pcf11p is involved in both steps of 3' end formation and was initially identified in two-hybrid screens as a polypeptide interacting with Rna14p and Rna15p (Amrani et al., 1997). A CTD-interaction domain (CID) located at the N-terminus of Pcf11p is responsible for the direct interaction of Pcf11p with the CTD phosphorylated on serine 2 and by this couples 3' end processing and transcription termination (Barilla et al., 2001; Licatalosi et al., 2002; Sadowski et al., 2003). Furthermore, Pcf11p is involved in the dismantling of the RNAP II elongation complex (Zhang et al., 2005). Pcf11p has only low affinity for the CTD and efficient interaction probably requires multiple binding sites (Noble et al., 2005). Interestingly,

pre-mRNA 3′ end processing activities and CTD binding of Pcf11p could be functionally uncoupled from each other (Sadowski et al., 2003). Pcf11p has putative homologues in many species that all contain a N-terminal CID (Barilla et al., 2001; Steinmetz and Brow, 1996; Yuryev et al., 1996). The high-resolution X-ray structure of the Pcf11p CID in complex with a CTD-derived peptide has been reported, identifying eight α-helices in a right-handed superhelical arrangement in the CID (Meinhart and Cramer, 2004). The central region of Pcf11p contains binding sites for the three other CF IA subunits and two putative zinc-finger motifs (Amrani et al., 1997; Sadowski et al., 2003).

Clp1p is the yeast homologue of the mammalian hClp1 (de Vries et al., 2000). The amino acid sequence contains a Walker A motif that is predicted to bind adenine or guanine nucleotides, suggesting a role in nucleotide binding or catalysis. Clp1p binds to Pcf11p but not to the other CF IA subunits (Gross and Moore, 2001) and also contacts the Ysh1p subunit of CPF (Kyburz et al., 2003) that is believed to be the endonuclease responsible for pre-mRNA cleavage (Dominski et al., 2005; Ryan et al., 2004). A recent publication reports the structure of the ternary complex formed by the association of Clp1p with ATP and the Clp1p-binding region of Pcf11p (Noble et al., 2007). This structure revealed that Clp1p contains three domains: a small N-terminal β sandwich domain, a C-terminal domain containing a novel α/β-fold and a central region that binds ATP. The nucleotide binding site contains a canonical P-loop motif similar to SIMIBI-class ATP/GTPases subunits (Leipe et al., 2002). The central domain also contains the binding site to Pcf11p.

1.3.2.3 The cleavage factor IB

CF IB consists of the single polypeptide Nab4p/Hrp1p (Kessler et al., 1997; Kessler et al., 1996). The factor is required for polyadenylation but not for cleavage, however, it regulates cleavage site selection by suppressing the use of cryptic poly(A) sites (Minvielle-Sebastia et al., 1998) and specifically binds the pre-mRNA on the *cis*-acting EE (Chen and Hyman, 1998; Valentini et al., 1999). The central region of Nab4p contains two RNP-type RNA-binding domains (RBDs) arranged in tandem (Henry et al., 1996). The solution structure of Nab4p in complex with an RNA sequence $G(UA)_4$ mimicking the EE sequence showed that only six bases $(UA)_3$ form

specific contacts to the factor (Perez-Canadillas, 2006). The C-terminal part of Nab4p is rich in arginine and glycine, with putative methylation sites. Interestingly, methylation of Nab4p facilitates the protein to shuttle between the nucleus and the cytoplasm (Henry et al., 1996; Kessler et al., 1997; Shen et al., 1998). The *NAB4* gene is genetically linked to the *NPL3* gene that encodes a protein involved in mRNA export, thus suggesting a role for Nab4p in this process (Henry et al., 1996). Furthermore, Nab4p is implicated in modulating the activity of the nonsense-mediated mRNA decay (NMD) pathway, a surveillance mechanism that monitors premature translation termination and degrades aberrant mRNAs (Gonzalez et al., 2000).

1.3.2.4 The poly(A) binding protein

Polyadenylated mRNAs are bound by Pab1p that is a protein highly conserved across species. Pab1p contains four N-terminal RNA recognition motifs (RRMs), which are connected to a C-terminal helical domain via a segment rich in proline and methionine (Mangus et al., 2004; Sachs et al., 1986). The main role of Pab1p is poly(A) tail length control (Kessler et al., 1997; Minvielle-Sebastia et al., 1997). Regulation of poly(A) tail length is achieved through two different functions: Pab1p inhibits Pap1p activity by limiting its access to the pre-mRNA (Zhelkovsky et al., 1998), and recruits a poly(A)-specific nuclease, PAN (Deardorff and Sachs, 1997; Lowell et al., 1992). Association with the pre-mRNA occurs in the nucleus but also in the cytoplasm (Adam et al., 1986; Setyono and Greenberg, 1981). A minimum of 12 adenosines is required as a binding site for one polypeptide. The association of Pab1p with the poly(A) tail also leads to stimulation of mRNA translation *in vivo* and *in vitro* by interacting with eIF4G, a component of the yeast translation initiation factor (Tarun and Sachs, 1996; Tarun et al., 1997). Furthermore, Pab1p has an inhibitory function in mRNA decapping and deadenylation (Caponigro and Parker, 1995).

1.3.3 Poly(A) tail functions in yeast

Polyadenylation is an essential processing step for most eukaryotic mRNAs. Poly(A) tails are synthesized in the nucleus within a defined size range depending on the

organism: from 70 nt in yeast to 250 nt in mammals. The length of the poly(A) tract added to the messengers is tightly regulated by Pab1p. Why poly(A) tails have a defined size is not understood, however, it appears that the functions of poly(A) tails are linked to the poly(A) binding protein.

1.3.3.1 Poly(A) tails stimulate translation initiation

A combination of *in vivo* and *in vitro* data obtained during the last 20 years led to the observation that the poly(A) tail added to the 3' end of mRNAs stimulates translation initiation of the transcripts (Iizuka et al., 1994; Jacobson and Favreau, 1983; Sachs and Davis, 1989). Translation stimulation is mediated by Pab1p that acts synergistically with the 5' end cap structure to enhance the binding of the 40S small ribosomal subunit to the mRNA (Tarun and Sachs, 1995; Tarun and Sachs, 1996). The recruitment of the 40S small ribosomal subunit to the mRNA is mediated by a physical interaction between Pab1p and the translation initiation factor eIF4F. Furthermore, it was found that this interaction with Pab1p occurs through a contact to the eIF4G component of the eIF4F complex, only when Pab1p was bound to the poly(A) tail, resulting in the circularization of the mRNA molecule (Tarun and Sachs, 1996; Tarun et al., 1997). These combined cooperative interactions enhance the affinity of eIF4F for the 5' cap of the mRNA (Borman et al., 2000; Haghighat and Sonenberg, 1997; Luo and Goss, 2001) and the RNA-binding capacity of Pab1p (Le et al., 1997; Munroe and Jacobson, 1990a; Munroe and Jacobson, 1990b). Translation initiation enhancement by poly(A) tails is also found in higher eukaryotes (Munroe and Jacobson, 1990a). Poly(A) tails then also have a role in development by regulating the expression of certain gene products. Indeed, poly(A) tail shortening or extension was shown to regulate the translation level of some mRNAs (Curtis et al., 1995).

1.3.3.2 The role of Poly(A) tails in mRNA stability

The process of mRNA decay can be initiated by three distinct events: removal of the 5' cap, endonucleolytic cleavage and poly(A) shortening (Jacobson and Peltz, 1996).

Poly(A) tails have a role in the regulation of mRNA stability by preventing degradation by exonucleases, as deadenylation is often the first step of mRNA decay in yeast (Beelman and Parker, 1995; Caponigro and Parker, 1996; Sachs and Wahle, 1993). Poly(A) tail shortening in yeast is followed by decapping of the transcript by the Dcp1p-Dcp2p complex (Figure 1.1) and 5' to 3' exonucleolytic degradation by Xrn1p or the exosome (Decker and Parker, 2002).

As in translation initiation, Pab1p has a role in mRNA decay. Yeast *pab1* mutant strains contain mRNAs with abnormally long poly(A) tails, indicating that Pab1p is also involved in poly(A) tail shortening (Sachs and Davis, 1989). Furthermore, Pab1p is recruited by the poly(A) nuclease PAN that shortens the excessively long newly synthesized poly(A) tails in an mRNA-specific manner (Boeck et al., 1996; Brown et al., 1996). The 3' to 5' exoribonuclease activity of PAN was shown to be stimulated by Pab1p (Lowell et al., 1992).

1.3.3.3 Poly(A) tails are involved in mRNA export

Poly(A) tails are also implicated in mRNA export of the transcripts from the nucleus to the cytoplasm (Huang and Carmichael, 1996; Long et al., 1995). After complete maturation, mRNA molecules are translocated into the cytoplasm through the nuclear pore complex to serve as template for translation into proteins. Translation per se is an essential step in the expression of genetic information. Pab1p, which coats the nascent poly(A) tails, has been shown to interact with specific nucleoporins and the nuclear export signal export receptor Xpo1p (Allen et al., 2001; Hammell et al., 2002). Furthermore, Pab1p was found to localize in the nucleus and in the cytoplasm by immunofluorescence studies, consistently with the fact that the protein shuttles between both compartments (Anderson et al., 1993). The role of poly(A) tails in export reflects a quality control mechanism, which retains the mRNAs that were not correctly processed, and that helps the mature mRNAs to reach the cytoplasm (Hammell et al., 2002; Hilleren et al., 2001).

1.3.3.4 Poly(A) tails are involved in RNA quality control

Pap1p is the canonical nuclear poly(A) polymerase that catalyzes the polyadenylation step during pre-mRNA 3' end processing. Recent studies have shed light on another type of nuclear poly(A) polymerase in yeast belonging to the polymerase β-type nucleotidyltransferase superfamily (Aravind and Koonin, 1999), Trf4p, which is conserved throughout eukaryotes (LaCava et al., 2005; Vanacova et al., 2005). Genetic and biochemical studies of this novel poly(A) polymerase demonstrated a new function for poly(A) tails in degrading aberrant RNAs (tRNAs, rRNAs, snRNAs and some intergenic mRNAs). Trf4p stably associates with a RNA helicase, Mtr4p, and with one of two RNA-binding proteins, Air1p or Air2p, to form the Trf4p/Air1p/Mtr4p polyadenylation complex (TRAMP). It was shown that TRAMP has polymerase activity that can polyadenylate *in vitro* small RNAs, including tRNAs. Interestingly, poly(A) addition to aberrantly structured RNAs was then shown to stimulate degradation by the exosome, a complex of multiple 3'-5' exonucleases that degrade numerous RNAs (Mitchell et al., 1997), revealing a novel involvement of poly(A) tails in RNA quality control (LaCava et al., 2005; Vanacova et al., 2005). This property is reminiscent of the poly(A) tail function in prokaryotes, in which the poly(A) tail promotes the degradation of the RNA (Grunberg-Manago, 1999). Thus, Trf4p and the TRAMP complex are involved in a nuclear surveillance mechanism that prevents the export and subsequent expression, in the case of mRNAs, of misfolded RNAs. How does TRAMP discriminate a substrate that has to be degraded from a substrate that has to undergo conventional polyadenylation is still an open question. It was however suggested that the TRAMP complex could recognize structural features induced by the misfolded RNAs.

In summary, the roles of poly(A) tails illustrate the interdependence of the different steps in gene expression and the regulatory effects of these coupling events.

1.4 Interconnections between pre-mRNA processing events and RNAP II transcription

Eukaryotic gene expression is a complex stepwise process that includes transcription, pre-mRNA processing and translation into active proteins. The complexity of each

step in the pathway required them to be studied independently of the others. The establishment of *in vitro* assays first allowed to reconstitute the activity of each individual process and to characterize the proteins involved. Therefore, as the different steps were analyzed separately, gene expression has been traditionally viewed as a linear series of events, with each one going to the end before the next starts. However, recent analyses have demonstrated that many of the steps involved in this pathway are connected (Figure 1.6).

In the recent years, a growing number of genetic and biochemical studies have revealed that the protein machineries that perform the different steps in the gene expression pathway interact both physically and functionally (for reviews, see Bentley, 1999; Cramer et al., 2001a; Hirose and Manley, 2000; Proudfoot, 2000; Shatkin and Manley, 2000). By such a connection, the different processes influence one another's efficiency and specificity. In addition, this coupling ensures the efficient transfer of the genetic information from one step to the next.

All three pre-mRNA processing steps are coupled to transcription (Bentley, 2005; Cramer et al., 2001a; Hirose and Manley, 2000; Kornblihtt et al., 2004; Maniatis and Reed, 2002; Orphanides and Reinberg, 2002; Proudfoot, 2004). It has been demonstrated that the transcription machinery actively recruits the capping enzymes (Shatkin and Manley, 2000), and that splicing and 3' end formation promote transcription termination (Fong and Bentley, 2001; Zaret and Sherman, 1982). Those coupling events are mediated by the C-terminal domain of the largest subunit (Rpb1p in yeast) of RNAP II.

1.4.1 The CTD is a recruitment platform for pre-mRNA processing factors

Optimal pre-mRNA 3' end processing is achieved by coupling with RNAP II only and not other polymerases. The reason for this specificity resides in the presence of a unique and unusual domain in the largest subunit of RNAP II, called the C-terminal domain (CTD), which provides a landing pad for reversible interactions with pre-mRNA processing factors (Hirose and Manley, 1998; McCracken et al., 1997b; Proudfoot et al., 2002). The binding of specific processing factors depends on the phosphorylation state of the CTD, which is dynamic during the transcription cycle (Dahmus, 1995).

The CTD is an essential domain in RNAP II, but is absent in RNA pol I and RNA pol III. The CTD consists of multiple heptapeptide repeats whose the amino acid consensus sequence, $Y_1S_2P_3T_4S_5P_6S_7$, is identical across animals, plants and some protozoa. The number of repeats depends on the species, and is 52 in human and 26 in yeast, from which 8 are essential for yeast viability (Nonet et al., 1987). The importance of the CTD in the processing steps is illustrated by the fact that certain CTD mutations that have no effect on transcription significantly affect capping (Cho et al., 1997; McCracken et al., 1997a), splicing and 3' end polyadenylation (McCracken et al., 1997b). In crystallographic studies, the CTD appears as an unstructured tail-like extension directly adjacent to the exit groove for the pre-mRNA and is linked to Rpb1p through a flexible linker of 80 residues (Cramer et al., 2001b). This flexibility apparently facilitates interaction with the processing factors.

The CTD can undergo several kinds of modifications, such as phosphorylation, cis/trans isomerization of prolines and glycosylation (Meinhart et al., 2005). Site-specific phosphorylation and dephosphorylation is the most important modification and is critical for CTD function (Cho et al., 2001; Dahmus, 1996; Komarnitsky et al., 2000). The current paradigm is that the phosphorylation pattern of RNAP II changes during the transcription cycle and that different proteins bind to specific phosphorylated forms of the CTD. The heptapeptide repeat contains five potential phosphorylation sites: Y_1, S_2, T_4, S_5 and S_7. However, phosphorylation occurs mainly on serines at position 2 and 5 (Corden et al., 1985; Zhang and Corden, 1991), giving rise to four possible phosphorylation states of the CTD (unphosphorylated form; phosphorylated at S2; phosphorylated at S5; phosphorylated at S2 and S5). These phosphorylations involve two specific cyclin associated kinases: in yeast, the TFIIH subunit Kin28p specifically phosphorylates S5 (Feaver et al., 1991), whereas S2 is phosphorylated by Ctk1p. Several phosphatases are also known, including Ssu72p (a subunit of CPF) that is specific for S5 (Ganem et al., 2003) and Fcp1p that can dephosphorylate both S2 and S5 (Cho et al., 2001). The phosphorylation pattern is dynamic during the transcription cycle and thus controls the recruitment of processing factors. S5 posphorylation occurs early in the transcription cycle and promotes promoter release, the formation of the elongation complex, and leads to the recruitment of capping enzymes (Cho et al., 1997; Ho et al., 1998; Komarnitsky et al., 2000; McCracken et al., 1997a). In contrast, posphorylation of S2 is predominant when RNAP II transcribes the gene, including at the 3' end, and

triggers association with splicing and 3' end processing factors (Cho et al., 1997; Komarnitsky et al., 2000). Furthermore, the dephosphorylated form stimulates the recruitment of RNAP II in the preinitiation complex (Cho et al., 1999). Thus, relating to the posphorylation pattern, a "CTD code" can define the position of RNAP II within the transcription cycle (Buratowski, 2003).

The CTD code consists of all the possible states that the CTD can adopt. Taking into account four different phosphorylation states and four possible proline configurations (*cis* or *trans* conformations of P_3 and P_6), a combination of 16 different states is obtained for only one repeat (Buratowski, 2003). Moreover, the complexity of the CTD code is extended by the facts that a functional unit of the CTD is a pair of repeats (Stiller and Cook, 2004), that the residue Y_1 can also be phosphorylated (Baskaran et al., 1993), and by possible CTD glycosylation (Kelly et al., 1993).

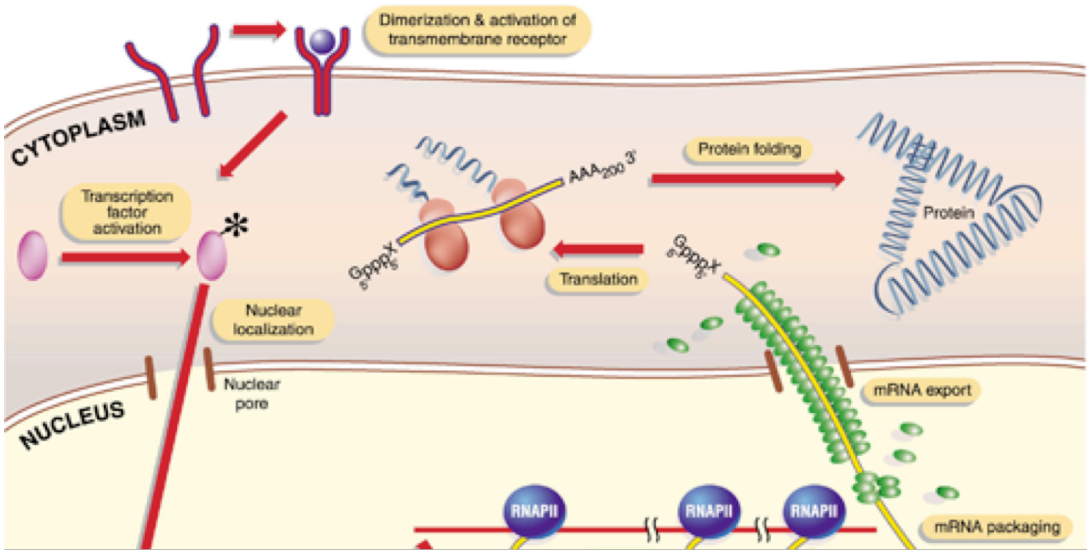

Figure 1.6. Transcriptional coupling of pre-mRNA processing events (from Orphanides et al., 2002).
The transcription cycle is depicted starting at the bottom left. Each maturation step (capping, splicing and 3' end processing) is physically and functionally connected to transcription.

1.4.2 Coupling between capping and transcription

Capping of pre-mRNAs occurs cotranscriptionally soon after transcription initiation, when the nascent transcript has reached 20-40 nt in length (Coppola et al., 1983; Rasmussen and Lis, 1993). The cap structure plays an important roles in mRNA stability by preventing 5'-3' exonucleolytic degradation, in stimulating splicing and 3' end formation, but also in transport and cytoplasmic translation (Lewis et al., 1995).

The capping enzymes are in fact the first processing factors to be recruited to the CTD during the transcription cycle. Phosphorylation of the CTD on S5 by Kin28p is required for the binding of the three capping enzymes (Cho et al., 1997; Komarnitsky et al., 2000; McCracken et al., 1997a). Moreover, activity of the mammalian guanylyltransferase is significantly enhanced upon CTD binding, as the Km for GTP was reduced (Ho and Shuman, 1999). After addition of the cap, the dephosphorylation of S5 in the early transcription elongation phase is coupled to the release of the capping enzymes. As shown by chromatin immunoprecipitation experiments (ChIP), Ceg1p-Cet1p is rapidly removed, whereas the methyltransferase, Abd1p, remains unexpectedly associated with RNAP II even at the 3' end of the genes (Komarnitsky et al., 2000; Schroeder et al., 2000). This might be due to the other function of Abd1p in the stimulation of transcription elongation (Schroeder et al., 2004), whereas the RNA triphosphatase, Cet1p, was shown to inhibit re-initiation of transcription (Myers et al., 2002).

Hence, capping enzymes not only fulfil the first pre-mRNA processing step, but are also able to modulate transcription activity. Therefore, they are though to act as a checkpoint to make sure that the cap has been added to the primary transcript before going on in the gene expression pathway.

1.4.3 Coupling between splicing and transcription

Most eukaryotic genes are interrupted by noncoding sequences, also called introns. To generate a functional mRNA, introns have to be removed and exons joined together during the splicing reaction (for a review, see Kramer, 1996). Like capping enzymes, splicing factors are recruited to the nascent transcript to cotranscriptionally remove introns (Beyer et al., 1981), and can enhance transcription elongation (Fong and

Zhou, 2001). Of the three pre-mRNA processing events, splicing is the most complicated, as reflected by the involvement and the stepwise integration of five snRNPs composed of U1, U2, U4, U5, and U6 small nuclear RNAs (snRNAs) and more than 200 proteins in the dynamic spliceosome complex (Jurica and Moore, 2003). Consistent with this complexity, the recruitment of the splicing apparatus is still largely misunderstood, and the role of the CTD less fully documented than for capping. A link between splicing and transcription was demonstrated by the coimmunoprecipitation of spliceosomal snRNPs and RNAP II by using a specific antibody for the phosphorylated CTD. The phospho-CTD was shown to stimulate the spliceosome assembly *in vitro* and *in vivo*, whereas the unphosphorylated form inhibited the reaction (Fong and Bentley, 2001; Hirose et al., 1999). The simplest situation is found in the budding yeast *S. cerevisiae*, where only 3% of the genes contain introns (Ares et al., 1999; Lopez and Seraphin, 1999). Moreover, in most cases, these genes have only one single and short intron.

In yeast, U1snRNP, which binds to the 5' splice site, interacts with the transcription elongation complex in the intron and remains bound to the downstream exon, suggesting that U1snRNP is recruited to the intronic RNA rather than to the RNAP II CTD (Kotovic et al., 2003). In contradiction with this observation, U1snRNP was also found on some genes lacking introns, raising the hypothesis that splicing factors can be recruited either by binding to the pre-mRNA or to the transcription machinery. Consistently, the association of the U1snRNP protein Prp40p, of another splicing factor Sub2p and of a set of splicing-related factors (SCAFs) with the phosphorylated CTD was reported (Abruzzi et al., 2004; Corden and Patturajan, 1997; Morris and Greenleaf, 2000). These interactions suggest a role for these proteins in linking RNAP II transcription to splicing.

However, it is still questionable whether the functional link between transcription and splicing observed in mammals can be extended to yeast. Because very few genes have more than one intron, it is conceivable that yeast does not use a spliceosome recycling mechanism, as metazoans require to tackle their highly fragmented genes. Indeed, most of the data indicating interactions between splicing components and RNAP II come from mammalian cells studies (Carty and Greenleaf, 2002; Emili et al., 2002; Mortillaro et al., 1996; Robert et al., 2002). ChIP experiments suggest that different snRNPs are recruited at different times during elongation. Furthermore, the CTD has only very little influence on splicing in yeast

(Licatalosi et al., 2002), whereas the phosphorylated CTD enhances the reaction in mammals (Bird et al., 2004; Millhouse and Manley, 2005).

1.4.4 Coupling between 3' end formation and transcription

Most pre-mRNAs are processed at their 3' end by a two-step reaction that cleaves at the poly(A) site and then polyadenylates the upstream cleavage product. This processing phase is carried out by the multiprotein apparatus described above. Like the two other processing steps, 3' end formation is intimately coupled to transcription. Indeed, it was shown that correct transcription termination requires an intact poly(A) site (Whitelaw and Proudfoot, 1986; Zaret and Sherman, 1982), and that deletion of the RNAP II CTD has a detrimental effect on 3' end formation (McCracken et al., 1997b).

Surprisingly, the recruitment of polyadenylation factors not only occurs at the 3' end, but also at the 5' end and through the entire gene. Indeed, ChIP experiments indicated that this recruitment is progressively increased from the promoter to the end of the gene, suggesting that 3' end processing factors associate with transcribing RNAP II from the 5' end to the 3' end of the gene (Calvo and Manley, 2005; Kim et al., 2004a; Licatalosi et al., 2002). The unexpected presence of 3' end processing factors at the promoter region is correlated to some early transcriptional events. Rna15p (a CF IA subunit) interacts with the coactivator Sub1p that has a role in the definition of the transcriptional start site (Calvo and Manley, 2005) and Ssu72p (a CPF subunit) interacts with the general transcription factor TF IIB (Calvo and Manley, 2005; Krishnamurthy et al., 2004). In addition, Ssu72p, which is a CTD phosphatase specific for S5, has been involved in 3' end formation and transcription termination (Dichtl et al., 2002b; Ganem et al., 2003; Krishnamurthy et al., 2004). The interactions between components of the initiation and termination transcription machineries suggest that both transcription phases are connected, possibly to facilitate reinitiation of transcription.

Strikingly, this distribution of 3' end processing factors over the gene is similar to that of S2 phosphorylation of the CTD, which is high at the 3' end region (Komarnitsky et al., 2000). CTD phosphorylation of S2 by Ctk1p plays an important role in 3' end maturation. Indeed, a strong connection has been established between

the CTD phosphorylated on S2 and the recruitment of 3' end formation factors. Among evidences, overexpression of the CPF subunit Pti1p was shown to suppress the mutant phenotype of strains that carry a deletion of the *CTK1* gene (Skaar and Greenleaf, 2002). Moreover, the recruitment of polyadenylation factors to 3' end regions of genes was disrupted in yeast strains lacking Ctk1p (Ahn et al., 2004). Furthermore, S2 phosphorylation significantly increased the *in vitro* CTD binding of Pcf11p, a CF IA component (Licatalosi et al., 2002). The crystal structure of this interaction shows that there is no direct contact between the S2 phosphate group and Pcf11p, but instead that the interaction induces a stable conformation of the structure (Meinhart and Cramer, 2004). Pcf11p contains a specific CTD interaction domain (CID) that mediates the binding to the CTD. Other 3' end processing factors were also shown to preferentially bind to the phosphorylated CTD, as Yhh1p (Dichtl et al., 2002a), Ydh1p (Kyburz et al., 2003), Pta1p (Rodriguez et al., 2000), Rna14p (Barilla et al., 2001) and Nab4p (Komarnitsky et al., 2000), but also to the CTD without known preference for a phosphorylated state, as the CPF subunits Pap1p, Pfs2p, Yth1p and the CF IA subunits Rna14p and Rna15p. Interestingly, both unphosphorylated and phosphorylated forms of the CTD are able to activate 3' end cleavage even in the absence of on-going transcription (Hirose and Manley, 1998). Since RNAP II undergoes changes in its phosphorylation status when the end of the gene is reached, this absence of discrimination may allow the 3' end processing factors to stay attached to the polymerase to complete their task.

The final step in the production of a mature mRNA is transcription termination that is critical for successful gene expression. It involves the release of the transcript from the site of transcription and the release of RNAP II from the DNA template, so that the polymerase can be recycled for a new round of transcription. A functional link between pre-mRNA 3' end processing and transcription termination was initially demonstrated by the finding that both a functional poly(A) signal (Zaret and Sherman, 1982) and 3' end processing factors (Birse et al., 1998; Hirose and Manley, 1998; McCracken et al., 1997b) are required for correct transcription termination (reviewed in Proudfoot, 1989). These data were obtained by using transcription run-on analysis showing that mutations in the poly(A) signal led both to loss of mRNA and to the continued transcription of the 3' flanking regions of the gene (Connelly and Manley, 1988; Whitelaw and Proudfoot, 1986). Hence, loss of the poly(A) signal leads to read-through into downstream genes by polymerases that have

failed to terminate at upstream genes. This is of particular importance in yeast where genes are closely spaced. Furthermore, mutant strains defective in the expression of subunits of CPF and CF IA were tested for defects in RNAP II termination also by run-on analysis. Rna14p, Rna15p and Pcf11p subunits of CF IA (Birse et al., 1998) and CPF subunit Yhh1p were found to be required for transcription termination (Dichtl et al., 2002a). It is striking that some domains of these factors are devoted to transcription termination, whereas others are specific for 3' end formation. Thus, Rna15p and its *S. pombe* homologue *ctf1* contain a domain which when deleted leads to termination defect, but has no effect on polyadenylation (Aranda and Proudfoot, 2001). In addition, Yhh1p binds to the phospho-CTD via a domain, which is different than the one used for RNA binding. Moreover, Pcf11p possesses a specific CTD-binding domain (CID) in which mutations block transcription termination, but have no effects on pre-mRNA 3' end processing (Sadowski et al., 2003). Finally, even the strength of the poly(A) signal was shown to directly influence transcription termination efficiency (Osheim et al., 1999).

The mechanism of RNAP II termination is not well defined, and two alternative models have been proposed for coupling termination to 3' end processing: the "antiterminator" model (or allosteric model) and the "torpedo" model (for reviews, see Buratowski, 2005; Luo and Bentley, 2004). The antiterminator model proposes that conformational changes in RNAP II or loss of an antiterminator factor are mediated by the recognition of the poly(A) site, leading to gradual termination (Calvo and Manley, 2001; Greenblatt et al., 1993; Logan et al., 1987). The torpedo model suggests that endonucleolytic cleavage of the pre-mRNA at the poly(A) site is required to initiate termination. The downstream cleavage product, still attached to the elongating polymerase, is degraded by a 5'-3' exonuclease that targets RNAP II. In this model, the exonuclease catches up the polymerase and provokes termination (Connelly and Manley, 1988). Unlike the antiterminator model, endonucleolytic cleavage is an obligatory step in the torpedo model.

There is evidence for and against both models. Supporting the antiterminator model, the presence of some 3' end processing factors interacting with the CTD could induce conformational changes in the polymerase (Calvo and Manley, 2005; Kim et al., 2004a; Licatalosi et al., 2002). The CTD phosphorylated on S2 provides a kind of platform on which protein factors can bind, which could transmit a termination signal to the polymerase. Furthermore, it was shown in both mammals and yeast that optimal

termination requires the CTD. Also consistent with the antiterminator model is the bi-functional Pcf11p subunit of CF IA. Mutant strains deficient for CTD binding were shown to impair termination whereas cleavage activity remained unaffected (Sadowski et al., 2003). In contrast, mutant strains carrying a disrupted poly(A) site but an intact CID still supported termination. These observations indicate that pre-mRNA 3' end processing and termination functions of Pcf11p can be uncoupled (Sadowski et al., 2003). Thus, cleavage is not required for its termination function. Moreover, Pcf11p can dismantle transcription elongation complexes *in vitro* in a poly(A) site-independent manner, probably by bridging the CTD to the RNA and thus inducing a conformational change in the polymerase (Zhang et al., 2005).

On the other hand, a strong evidence for the torpedo model resides in the finding that the yeast 5'-3' exonuclease Rat1p and the mammalian hXrn2, are required for transcription termination (Kim et al., 2004b; West et al., 2004). In *rat1* mutants, cleavage and polyadenylation of pre-mRNAs are not affected, whereas a strong defect is observed in transcription termination. Furthermore, loss of Rat1p stabilizes the downstream cleavage products, suggesting that Rat1p is the exonuclease that usually degrades RNAs downstream of the poly(A) site.

A recent study showed that Rat1p contributes to the cotranscriptional degradation of the nascent RNA, which is however not sufficient to direct transcription termination (Luo et al., 2006). Moreover, it showed that Rat1p has a role in both 3' end formation and termination processes, as it allows recruitment of 3' end processing factors. Finally, recruitment of Rat1p itself required the presence of 3' end processing factors. Thus, the results of this study suggest a combined antiterminator/torpedo model in which Rat1p acts as a component of the 3' end processing machinery.

The complexity of the eukaryotic gene expression is reflected by the fact that the different mechanisms of the pathway are interconnected. These coupling events make sure that extensive controls ensure accurate and efficient expression of the genes and appropriate processing of the mRNAs.

1.5 Aims of this thesis

The past few years have brought progress in determining the detailed biochemistry of pre-mRNA 3' end processing and its connection with other processes of mRNA synthesis. Despite all of these efforts, the molecular mechanism by which pre-mRNAs are cleaved and polyadenylated is still unclear. One aim of this thesis is to provide structural information of the yeast 3' end processing machinery, in particular CPF, to better understand the mechanism of action of the complex. Another question is the characterization of Cpf11p, a putative new subunit of CPF. Finally, many *in vitro* data demonstrate that pre-mRNA 3' end formation and transcription are coupled, but very little is known about the *in vivo* situation. Another aim of this thesis is to investigate directly in living cells the relevance of the coupling between the two processes.

Chapter 2: Investigating *in vivo* coupling between yeast pre-mRNA 3' end processing factors and the CTD of RNAP II

This part of our work will not be treated as a complete chapter, since we are still in the process of constructing the appropriate tools that are required for the fluorescence microscopy studies. Instead, we present the yeast strains we have generated and show the obstacles we encountered to obtain fluorescent signals strong enough to perform fluorescence resonance energy transfer (FRET) and fluorescence recovery after photobleaching (FRAP) in yeast.

2.1 Introduction

Transcription and 3' end formation of pre-mRNAs transcribed by RNAP II are tightly coupled processes. First studies revealed that the poly(A) signal, which directs 3' end cleavage of the pre-mRNA, also acts as a transcription termination signal for RNAP II (Connelly and Manley, 1988; Whitelaw and Proudfoot, 1986; Zaret and Sherman, 1982). In addition, yeast strains with temperature-sensitive mutations in cleavage factors were shown to have a defect in terminating transcription at non-permissive temperatures, indicating that 3' end processing factors are also required for termination. Pcf11p, Rna14p and Rna15p subunits of CF IA and Yhh1p subunit of CPF were shown to be required for transcription termination, supporting the functional connection between 3' end formation factors and RNAP II (Birse et al., 1998; Dichtl et al., 2002a).

The connection between the transcription complex and the 3' end processing machinery is mediated by the C-terminal domain (CTD) of the largest subunit (Rpb1p) of RNAP II (McCracken et al., 1998). Deletion of the CTD strongly inhibits cleavage of pre-mRNAs in metazoan cells. In yeast, it reduces the efficiency of cleavage and the length of poly(A) tails (Licatalosi et al., 2002). The CTD is an unstructured domain that protrudes from RNAP II and consists of several heptad repeats of the highly conserved sequence YSPTSPS. A key feature of the CTD is the presence of serine residues at positions 2 and 5 of the heptapeptide repeat, which can be phosphorylated by Ctk1p and Kin28p, respectively. The phosphorylation status is critical for CTD function, as phospho-CTD directly activates pre-mRNA processing

reactions and acts as a platform where processing factors can bind (Hirose and Manley, 2000; Proudfoot and O'Sullivan, 2002; Proudfoot et al., 2002). CF IA subunits Pcf11p, Rna15p and Rna14p and CPF components Yhh1p, Ydh1p and Pta1p specifically interact with phospho-CTD (Ansari and Hampsey, 2005; Barilla et al., 2001; Dichtl et al., 2002a; Kyburz et al., 2003; Rodriguez et al., 2000; Steinmetz and Brow, 1996).

Pcf11p has a specific N-terminal CTD-binding domain (CID). Mutations in this domain affect transcription termination but not 3' end processing of the pre-mRNA, thus it has been proposed that both functions can be uncoupled (Sadowski et al., 2003). Via its CID, Pcf11p has been shown to preferentially bind to the CTD phosphorylated on serine 2 *in vitro* (Licatalosi et al., 2002). A co-crystal structure of this interaction has been reported and shows that there is no direct contact between Pcf11p and the phospho-serine 2 (Meinhart and Cramer, 2004). Instead, the CTD adopts a stabilized conformation that fits onto the Pcf11p surface.

Yhh1p is the yeast homologue of the mammalian AAUAAA interacting protein CPSF-160. Yhh1p is an RNA-binding protein, which participates in the poly(A) site recognition by binding sequences surrounding the cleavage site (Dichtl et al., 2002a). The RNA-binding domain is composed of β-propeller repeats. Yhh1p was shown to specifically bind to the phosphorylated CTD *in vitro* and to be involved in transcription termination (Dichtl et al., 2002a).

Surprisingly, the recruitment of 3' end processing factors to RNAP II not only occurs at the 3' end, but also at the 5' end and throughout the length of the genes during transcription (Kim et al., 2004a; Licatalosi et al., 2002). A yeast two-hybrid screen identified an interaction between the 3' end processing factor Rna15p and the transcriptional coactivator Sub1p (Calvo and Manley, 2001). Furthermore, a network of interactions between 3' end processing and general transcription factors was found at the promoter region, as TF IIB was shown to interact both with Ssu72p, a subunit of CPF, and Sub1p (Sun and Hampsey, 1996; Wu et al., 1999). The multitude of protein-protein interactions between components of the pre-mRNA 3' end formation machinery and transcription complexes at all stages (initiation, elongation and termination) indicates that the two processes are physically and functionally connected.

So far, most of the data supporting coupling between transcription and 3' end formation come from *in vitro* studies, but there is little evidence for contact between

the protein factors involved in both processes *in vivo*. Our aim is to provide direct evidence that both protein machineries truly interact *in vivo*, by using fluorescence microscopy based techniques, such as FRET and FRAP. We first want to investigate the association between the CTD and the 3' end processing factors Yhh1p and Pcf11p in living cells, as these interactions are already well described *in vitro*. The confirmation of these associations *in vivo* would validate our approach, which would then be used to test interactions of other pre-mRNA 3' end processing factors with the CTD *in vivo*.

2.2 Experimental design

2.2.1 Using fluorescence microscopy to investigate protein-protein interactions

FRET is a powerful technique for measuring interactions between two proteins directly in living cells. FRET involves the non-radiative transfer of photon energy from a fluorescent molecule (fluorophore) in an excited state (the donor) to another fluorophore (the acceptor) when both are located within close proximity. For this transfer to take place, the donor and acceptor molecules must be less than 10 nm apart and the emission spectrum of the donor fluorophore must overlap with the excitation spectrum of the acceptor fluorophore. The technique has grown in popularity thanks to the emergence of green fluorescent protein (GFP) mutants with blue, yellow or red-shifted spectral properties. FRAP is another live-cell imaging technique that provides information about the mobility and dynamics of fluorescent molecules by measuring the recovery of fluorescence in a defined region after a bleaching event. Interestingly, FRAP can also be used to address questions about protein interactions in living cells. If both transcription and 3' end processing machineries interact, FRAP is predicted to show that the proteins belonging to these apparatus behave similarly or follow the same movements in the cell. Thus, combination of both techniques provides an interesting tool to analyze *in vivo* protein-protein interactions.

Using both techniques involves the covalent fusion of the proteins of interest with compatible fluorescent proteins visible in a fluorescent light microscope. The protein fusions were genetically obtained by inserting the coding sequences of the fluorophores directly into the genome of a host cell, in-frame to and immediately

downstream of the open reading frame (ORF) of interest. With this strategy, wild type levels of protein expression are minimally perturbed. For FRAP, we used *Aequorea victoria* GFP (S65T), an enhanced form of GFP (Tsien, 1998), to tag several proteins involved in transcription (Rpb1p, Rpb3p, Spt5p; For a better understanding, Rpb1p will be referred as CTD in the following text) and pre-mRNA 3' end processing (Yhh1p, Ysh1p, Pta1p, Pap1p, Mpe1p, Pfs2p, Glc7p, Pti1p, Pcf11p, Rna14p and Nab4p). As a pair of fluorophores suitable for FRET, we used cyan fluorescent protein (CFP) as the donor (Zacharias et al., 2002) and yellow fluorescent protein (YFP) as the acceptor (Tsien, 1998). CFP and YFP molecules both derive from GFP by several point mutations, changing the emission color to blue and yellow, respectively. We also used other GFP variants such as Cerulean (CFP variant), Venus (YFP variant) and mCherry (red variant).

FRET and FRAP methods require strong fluorescent signals that are difficult to obtain in yeast, as the yeast cell is not a system of choice for fluorescence microscopy visualization. First, yeast cells are rather small (few micrometers in diameter) in comparison to mammalian cells, and consequently more difficult to observe through a microscope. Second, 3' end processing proteins are expressed at low levels in yeast cells. Many of the proteins, including essential proteins and most transcription factors, are present at levels that are not readily detectable by conventional methods (Ghaemmaghami et al., 2003). According to the study of Ghaemmaghami et al. (2003), the abundance of proteins ranges from fewer than 50 to more than 10^6 molecules per cell, and only ~600 copies of Yhh1p and 2800 copies of Pcf11p were found per cell. Thus, detection of the protein fusions requires strong fluorescence signals. Third, unlike mammalian cells, yeast cells display strong autofluorescence, which results in high background during observation through the fluorescence microscope. In addition, CFP is known to have a dim fluorescence, which results in a low signal-to-noise ratio. In accordance with these observations, the fluorescence signals generated by the strains we made were mostly too weak for FRET and FRAP analyses. To improve the strength of fluorescence signals, we also tried to use additional fluorophores (Cerulean, Venus and mCherry), tagged some proteins of interest with tandems of fluorophores or tried to use a galactose inducible system to overexpress the proteins to be tested. However, all our attempts to get sufficient fluorescence suitable for FRET and FRAP failed so far.

2.2.2 Generation of yeast strains for testing *in vivo* association of the CTD with Yhh1p and Pcf11p by FRET

A first series of double-tagged yeast strains was constructed in order to carry out FRET studies. To maximize chances to obtain suitable fluorescent signals, we generated homozygous diploid strains by mating haploids that were modified for both genes of interest: α-mated haploid cells containing CTD-YFP and Pcf11-CFP (or Yhh1-CFP) were crossed with a-mated haploid cells containing the same fusions. The reason for this resides in the facts that diploid strains are larger than haploid cells, and that two copies of the modified genes should lead to two times stronger fluorescence than only one copy. Thus, homozygous diploid strains (FRET strains) containing the following fusions were generated: CTD-YFP/Pcf11-CFP to investigate *in vivo* association between the CTD and Pcf11p; CTD-YFP/Yhh1-CFP to investigate *in vivo* association between the CTD and Yhh1p; CTD-YFP-CFP as control for FRET measurement.

To verify that the growth of FRET cells was not affected by the insertion of the fluorophores, cell cultures were prepared at the standard temperature 30ºC, at low (15ºC) and high (37ºC) temperatures. No defect in cell growth was observed at the three temperatures, indicating that the insertion of the fluorophores did not affect cell metabolism (Figure 2.1).

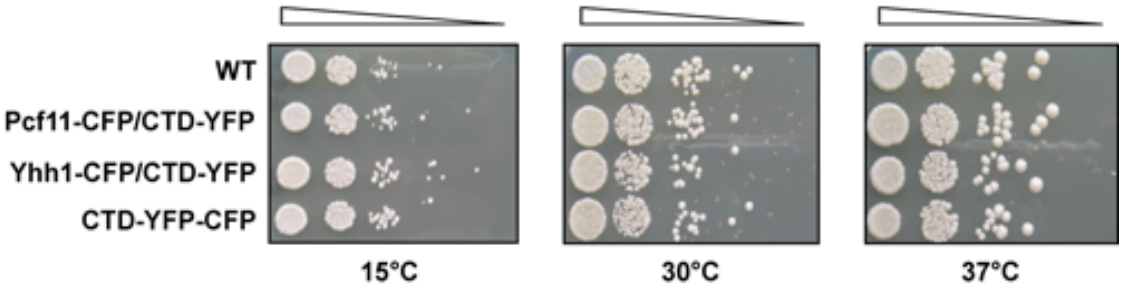

Figure 2.1. The insertion of the fluorophores did not affect cell growth.
Ten-fold serial dilutions of the strains indicated on the left were spotted on YPD agar medium and incubated at the temperatures indicated at the bottom of the panels for two days at 30°C and 37°C or four days at 15°C.

Furthermore, Western blot analysis was used to verify that the protein fusions were correctly expressed (Figure 2.2). The CTD-YFP fusion was well immunodetected in protein extracts prepared from the FRET strains by using a

specific antibody for the CTD phosphorylated on serine 2 (H5, from Covance). The presence of YFP was confirmed with an anti-GFP antibody (JL-8, from BD Biosciences), which also recognizes the other fluorophores, as they are all GFP variants. These observations indicate that the biggest subunit of RNAP II is well expressed in the cells. In contrast, neither Pcf11-CFP nor Yhh1-CFP could be detected with anti-GFP and protein-specific antibodies, most likely owing to the low amount of protein expressed, as mentioned above. However, as Yhh1p and Pcf11p are essential proteins, cells would not survive in their absence, suggesting that CFP-fusions were also expressed.

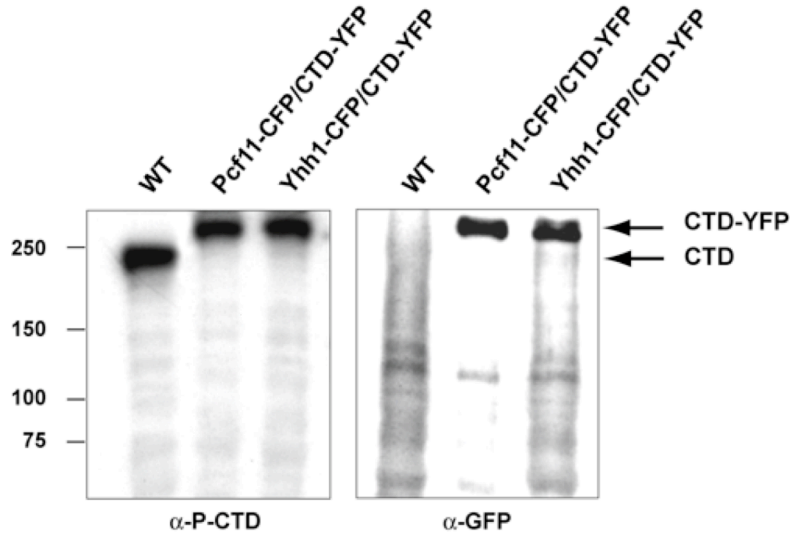

Figure 2.2. CTD-YFP is expressed in FRET cells.
Protein extracts prepared from FRET strains indicated on top were analyzed by Western blotting to verify the expression of the protein fusions. Proteins detected are indicated on the right of the panel (CTD refers to Rpb1p and was used instead of Rpb1p for a better understanding). The antibodies used are indicated at the bottom (α-P-CTD: antibody for CTD phosphorylated on serine 2). Molecular weight marker is indicated on the left (in kDa).

The observation of these yeast strains with the fluorescent microscope revealed a strong yellow nuclear fluorescent signal corresponding to CTD-YFP but a weak signal from CFP, only little stronger than the background fluorescence (Figure 2.4).

To improve the fluorescence signal from CFP, we used a variant form called Cerulean (Rizzo et al., 2004), which is 2.5-fold brighter than CFP. Cerulean was genetically fused to the C-terminus of the CTD. YFP, which is a brighter fluorophore relative to Cerulean, was attached to the C-terminus of Yhh1p and Pcf11p, expressed

at a lower level than the CTD. Western blotting analysis of these strains, containing Pcf11-YFP/CTD-Cerulean or Yhh1-YFP/CTD-Cerulean constructs, showed that CTD-Cerulean was readily detected with the anti-GFP and the anti-serine 2 phospho CTD antibodies (Figure 2.3). In contrast, Pcf11-YFP was only weakly detected and no signal was seen for Yhh1-YFP, confirming that these two proteins are expressed at a much lower level than the CTD. The microscopic inspection of these cells revealed that they show only weak nuclear fluorescence from YFP and a brighter fluorescence from Cerulean (Figure 2.4). However, fluorescence signals were not strong enough for FRET studies.

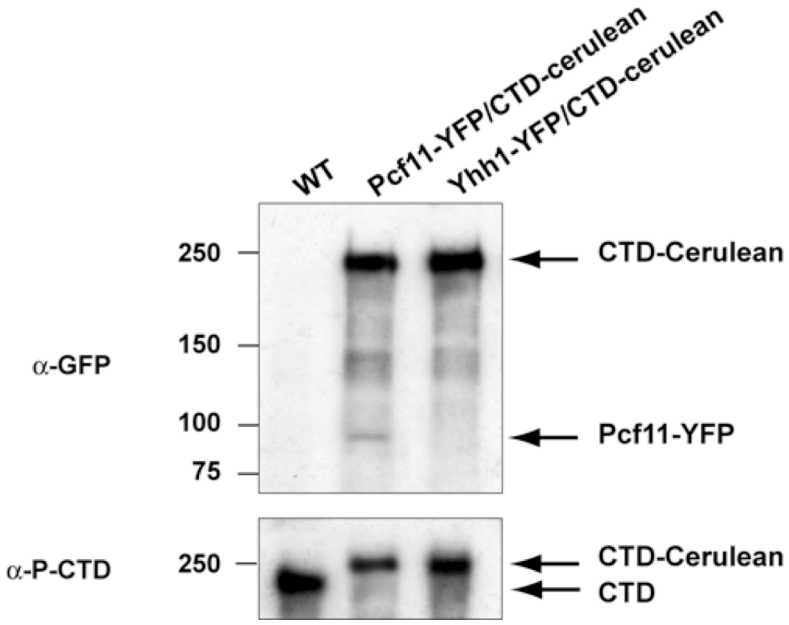

Figure 2.3. Western blotting analysis of strains containing Cerulean protein fusions.
Protein extracts from strains indicated on top were analyzed by Western blot. Antibodies used are indicated on the left and proteins detected on the right of the panels.

In another attempt to improve fluorescence, YFP was replaced by Venus, a brighter variant (Nagai et al., 2002), and new strains containing Pcf11-Venus/CTD-Cerulean or Yhh1-Venus/CTD-Cerulean were constructed. Nevertheless, also these improved variants did not show sufficient fluorescence for FRET experiments (data not shown). Unable to obtain enough fluorescence to test *in vivo* interactions between our proteins of interest by FRET, we then tried to make strains allowing inducible overexpression of the 3' end processing factors. The genes *YHH1, PCF11, RNA14* and *NAB4* were cloned into a plasmid allowing the expression of N-terminally GFP-tagged protein fusions from a galactose-inducible promoter (pGREG576; Jansen et

al., 2005). GFP and mCherry (Shaner et al., 2004), a red variant, is another pair of fluorescent proteins suitable for FRET. New yeast strains containing chromosomal CTD-mCherry fusions were transformed with inducible plasmids carrying the genes of 3' end processing factors Yhh1p, Pcf11p, Rna14p and Nab4p. These strains were not tested yet but are now available for fluorescence microscopy studies.

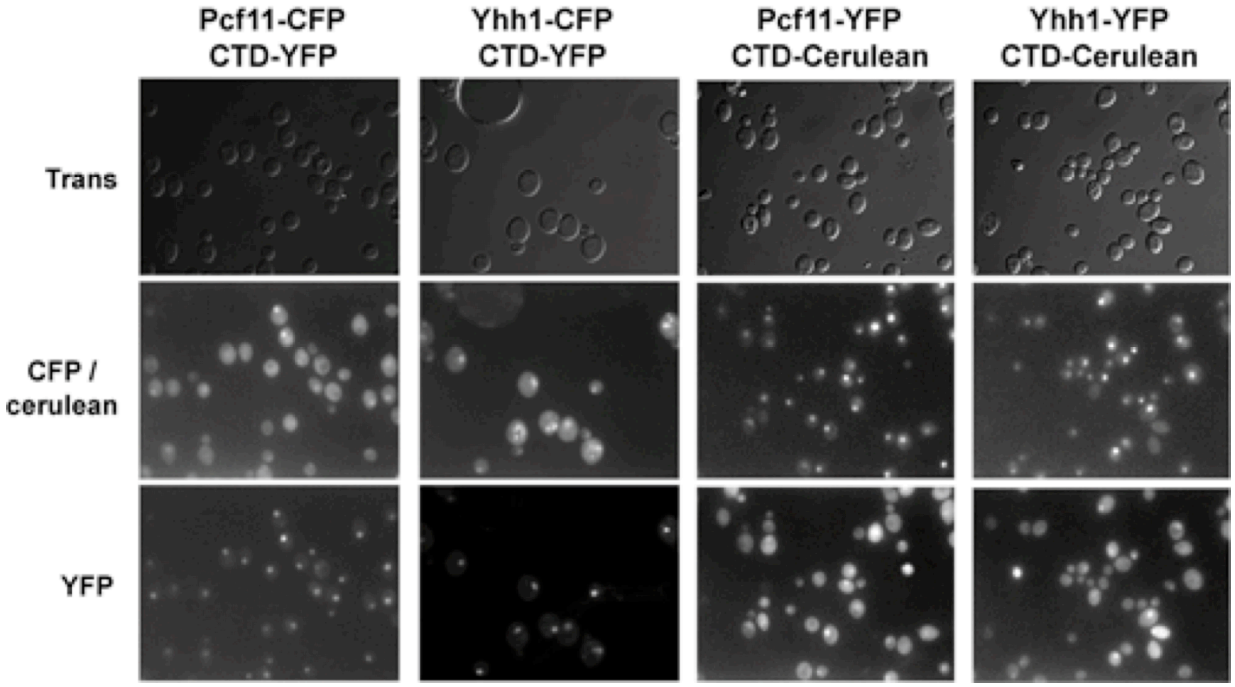

Figure 2.4. Microscopic analysis of live yeast strains expressing FRET constructs.
Representative images of strains indicated on top of the panels showing a strong nuclear fluorescence of the CTD fusions and a weak fluorescence of Pcf11p and Yhh1p fusions. Trans: transmission channel; CFP/Cerulean: fluorescence channel for CFP or Cerulean; YFP: fluorescence channel for YFP.

2.2.3 Yeast strains expressing GFP-tagged versions of 3' end processing and transcription factors for FRAP analysis

Combined with FRET experiments, FRAP analysis can also be used to analyze protein-protein interactions. Similar characteristics in the way different proteins move would suggest that they interact or belong to one large functional unit *in vivo*. A collection of yeast haploid strains expressing full-length, chromosomally tagged GFP fusion proteins was constructed to analyze localization of proteins in cell

compartments (Huh et al., 2003). Exploiting this library, we tried to perform FRAP analysis using several strains expressing GFP fusions of 3' end processing factors and proteins of the transcription machinery (Table 2.1).

GFP-tagged Proteins	Description
Rpb1p	Largest subunit of RNA pol II Carries the CTD
Rpb3p	Third subunit of RNA pol II
Spt5p	Transcription elongation factor
Pcf11p	CFIA subunit
Rna14p	CFIA subunit
Nab4p	CFIB subunit
Yhh1p	CPF subunit
Pfs2p	CPF subunit
Ysh1p	CPF subunit
Pap1p	CPF subunit
Mpe1p	CPF subunit
Pta1p	CPF subunit
Pti1p	CPF subunit
Glc7p	CPF subunit

Table 2.1. Transcription and 3' end processing proteins used for FRAP studies.

Microscopic observation of these strains revealed that all the strains expressing GFP fusions with 3' end processing factors showed too weak fluorescence signals for FRAP analysis (Figure 2.5). Pta1p, Pti1p and Pcf11p fusions with GFP showed weak fluorescence in the nucleus, whereas fluorescence of the other strains containing pre-mRNA 3' end processing protein fusions was too low to be precisely localized. In contrast, GFP fusions with proteins belonging to the transcription complex (CTD, Rpb3p, Spt5p) showed quite strong nuclear fluorescence. This fluorescence pattern could be correlated with Western blotting analysis showing that the proteins having the strongest fluorescence were also the proteins expressed at the highest level in the cells (Figure 2.6).

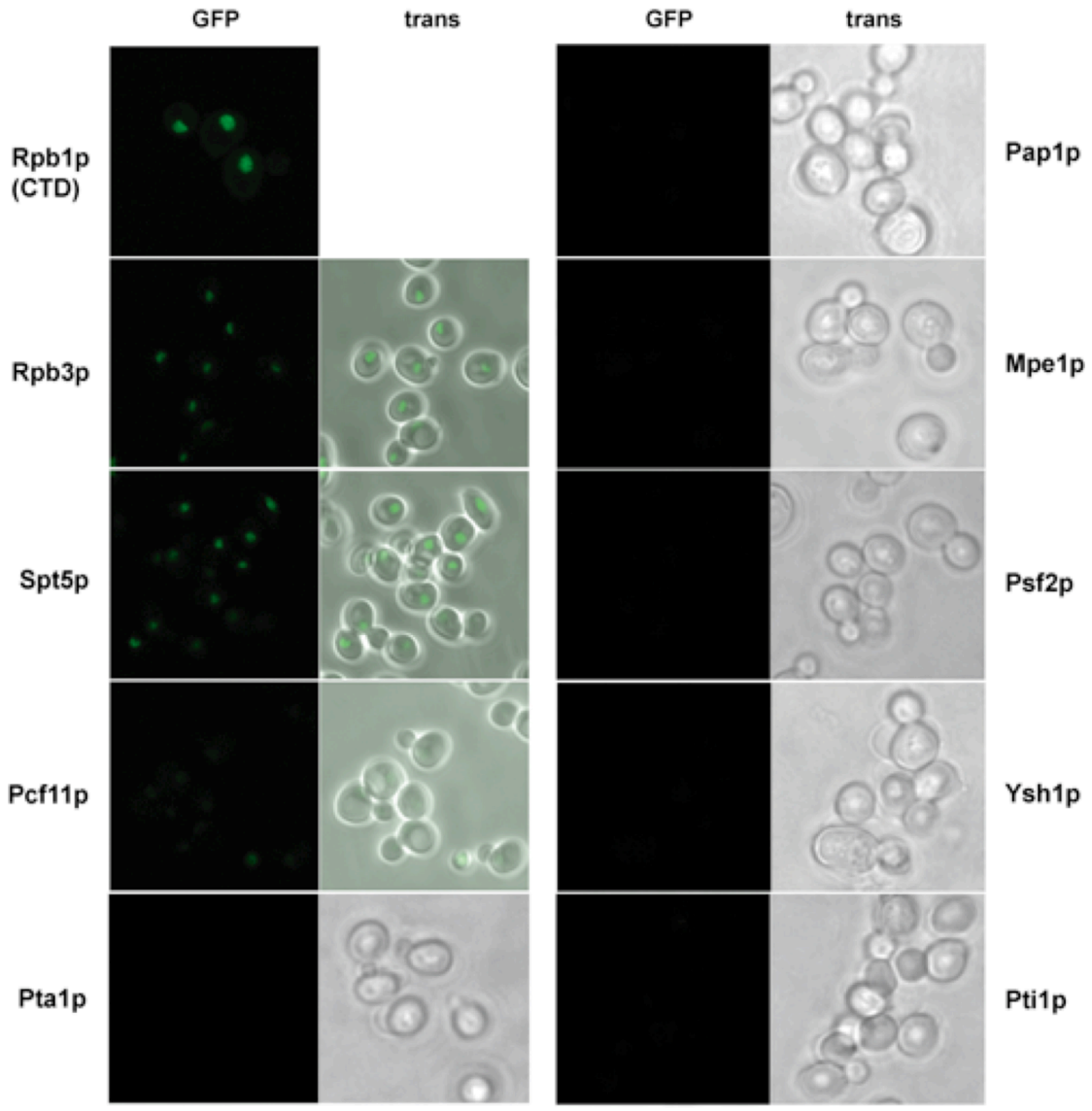

Figure 2.5. Microscopic observation of live yeast strains expressing FRAP constructs. Proteins tagged with GFP are indicated on the left and the right of the images. Fluorescent (GFP, excitation 488 nm/emission 505-530 nm) and transmission channels (trans) were used to visualize cells, as indicated on the top of the panels.

We also verified that the insertion of GFP tags did not affect the ability of the cells to perform pre-mRNA 3' end processing. Therefore, we tested protein extracts prepared from GFP strains for their ability to cleave and polyadenylate pre-mRNA substrates *in vitro*. 3' end processing activities can be reconstituted *in vitro* using

protein extracts or purified 3' end processing factors with a synthetic RNA substrate that was *in vitro* transcribed and radioactively labeled by incorporation of ^{32}P α-UTP. For the cleavage assay, we used the *CYC1* precursor that corresponds to the 3' untranslated region (3' UTR) of the iso-1-cytochrome c gene, which contains all the *cis*-acting sequence elements required for proper 3' end formation (Butler and Platt, 1988). For the *in vitro* polyadenylation assay, we used the pre-cleaved *CYC1* (pre-*CYC1)* precursor that corresponds to the upstream cleavage product of *CYC1*. Most of the extracts tested were competent in both pre-mRNA 3' end processing activities, except for Rna14-GFP and Ysh1-GFP extracts (Figure 2.7). The Rna14-GFP extract displayed weak cleavage activity and almost no polyadenylation. Surprisingly, the Ysh1-GFP extract showed a strong defect in polyadenylation, whereas cleavage was normal, although Ysh1p is believed to be the endonuclease (Mandel et al., 2006; Ryan et al., 2004). In both cases, it is possible that the addition of the GFP tag disrupts some important interactions or prevents conformational changes required for correct 3' end formation activities.

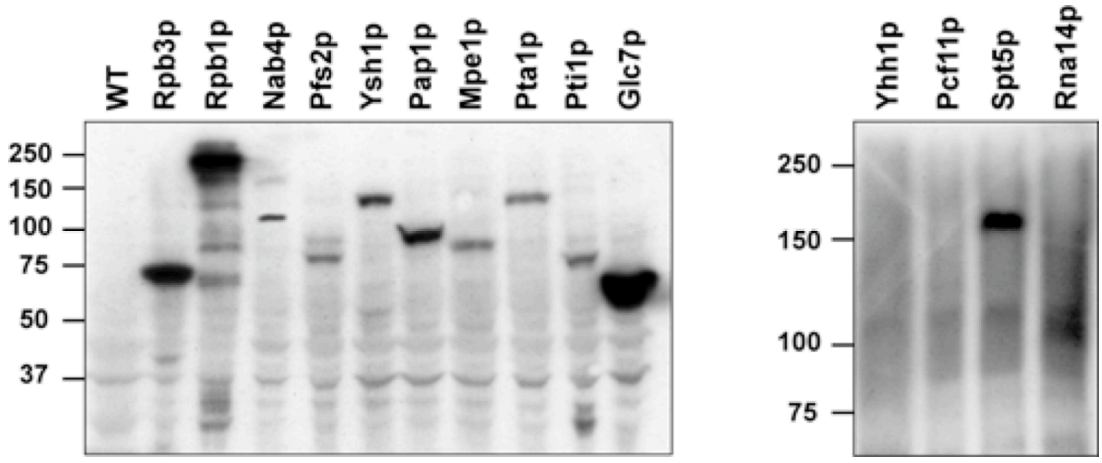

Figure 2.6. Western blotting analysis of GFP protein fusions in FRAP cells.
GFP antibody was used to verify the expression of GFP protein fusions in the protein extracts prepared from the GFP strains indicated on top. WT: wild type (no GFP-containing strain). The molecular weight marker is indicated on the left (in kDa).

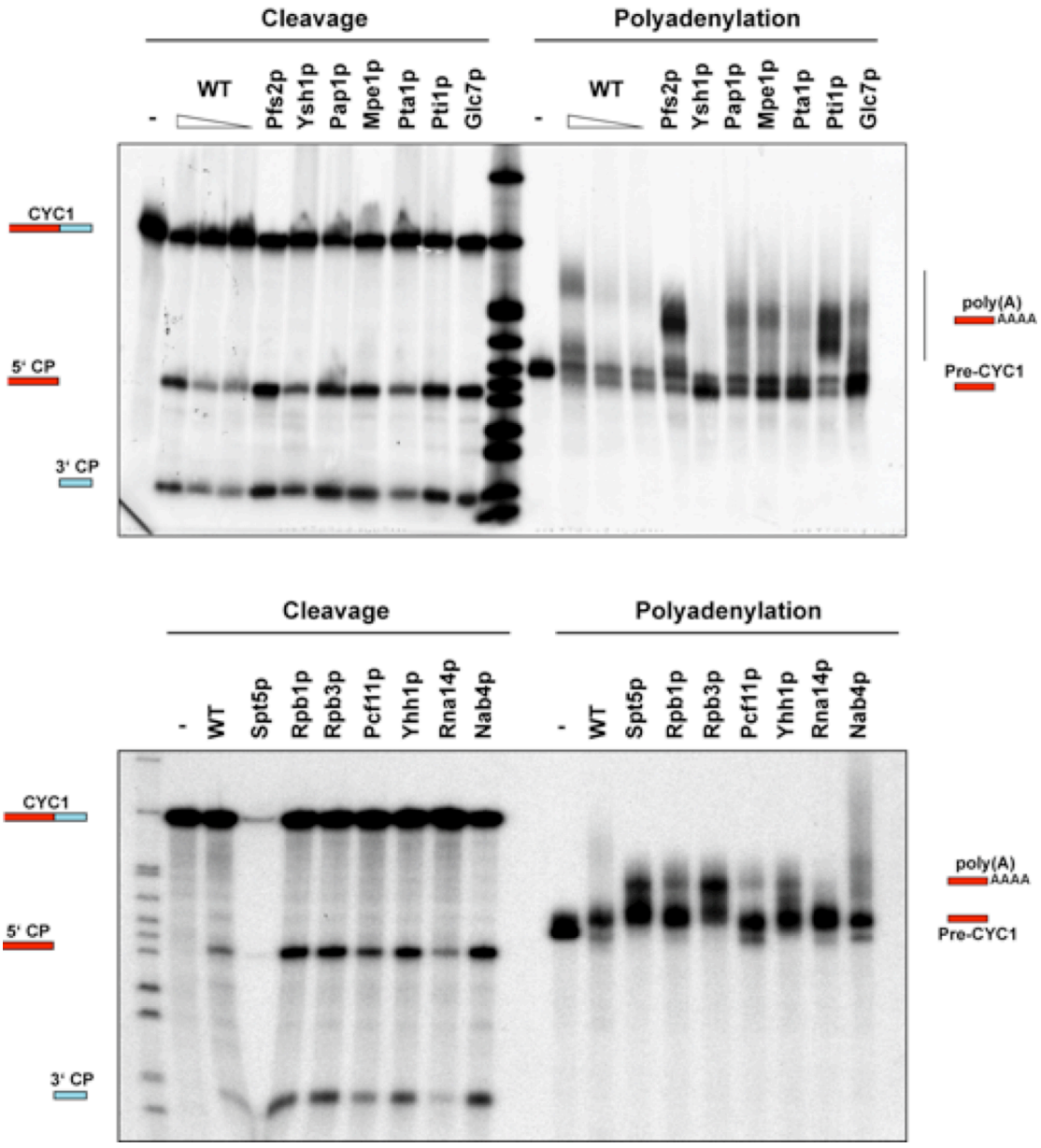

Figure 2.7. *In vitro* analysis of 3' end processing activities of GFP strains.
Protein extracts from the strains indicated on top of the panels were prepared and tested *in vitro* for cleavage and polyadenylation activities. RNA substrate used for cleavage (*CYC1*), 5' and 3' cleavage products (5' and 3' CP) are indicated on the left. RNA substrate used for polyadenylation (Pre-CYC1) and polyadenylation products [poly(A)] are indicated on the right. -: mock-treated sample; WT: wild type protein extract.

To obtain stronger fluorescent signals, we constructed diploids of the concerned strains. The mating type of the original haploid strains was switched and both a- and α-types were mated. Nevertheless, the new diploid strains did not show

stronger fluorescent signals. Another approach consisting of the addition of a second GFP tag to the pre-existing protein fusions in the original strains also did not give a stronger signal.

As endogenous expression of 3' end processing factors is too weak to provide sufficient fluorescence, overexpression of these proteins could be a successful alternative. For this, we started to clone the genes encoding for the proteins to be tested into a galactose-inducible expression vector (pGREG576), permitting the formation of amino-terminal fusions to GFP (Jansen et al., 2005).

2.2.4 Experimental procedures

Genotypes of yeast strains used in this study

FRET strains

CEN.PK2: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1* (Wach et al., 1997)

YBP11: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 PCF11-CFP/PCF11-CFP CTD-YFP/CTD-YFP*

YBP12: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 YHH1-CFP/YHH1-CFP CTD-YFP/CTD-YFP*

YBP17: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 PCF11-CERULEAN/PCF11-CERULEAN CTD-YFP/CTD-YFP*

YBP18: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 YHH1-CERULEAN/YHH1-CERULEAN CTD-YFP/CTD-YFP*

YBP23: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 PCF11-YFP/PCF11-YFP CTD-CERULEAN/CTD-CERULEAN*

YBP24: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 YHH1-YFP/YHH1-YFP CTD-CERULEAN/CTD-CERULEAN*

YBP29: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 PCF11-VENUS/PCF11-VENUS CTD-CERULEAN/CTD-CERULEAN*

YBP30: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3Δ1*/*his3Δ1 YHH1-VENUS /YHH1-VENUS CTD-CERULEAN/CTD-CERULEAN*

YBP31: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3*Δ*1*/*his3*Δ*1 CTD-mCherry* [pGREG576-*PCF11*]

YBP32: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3*Δ*1*/*his3*Δ*1 CTD-mCherry* [pGREG576-*YHH1*]

YBP33: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3*Δ*1*/*his3*Δ*1 CTD-mCherry* [pGREG576-*RNA14*]

YBP34: *Mat* **a**/α *ura3-52*/*ura3-52 leu2-3,112*/*leu2-3,112 trp1-289*/*trp1-289 his3*Δ*1*/*his3*Δ*1 CTD-mCherry* [pGREG576-*NAB4*]

FRAP strains

ATCC 201388: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*

YBP40: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB1-GFP*

YBP41: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RPB3-GFP*

YBP42: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SPT5-GFP*

YBP43: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YHH1-GFP*

YBP44: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 YSH1-GFP*

YBP45: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 PTA1-GFP*

YBP46: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 PAP1-GFP*

YBP47: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 MPE1-GFP*

YBP48: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 PFS2-GFP*

YBP49: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 PTI1-GFP*

YBP50: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 GLC7-GFP*

YBP51: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 PCF11-GFP*

YBP52: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 RNA14-GFP*

YBP53: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 NAB4-GFP*

YBP54-YBP67: isogenic to YBP40-YBP53, but (2n)

YBP68-YBP81: isogenic to YBP40-YBP53, but with *GFP-GFP* instead of *GFP*

YBP82: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pGREG576-*PCF11*]

YBP83: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pGREG576-*YHH1*]

YBP84: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pGREG576-*RNA14*]

YBP85: *MAT***a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pGREG576-*NAB4*]

Plasmids used in this study

pDH18: pYFP-CFP-his3 (Yeast Resource Center, University of Washington)

pDH5: pYFP-his3 (Yeast Resource Center, University of Washington)

pDH3: pCFP-kanamycin (Yeast Resource Center, University of Washington)

pBS10: pCerulean-hygromycin (Yeast Resource Center, University of Washington)

pBS7: pVenus-kanamycin (Yeast Resource Center, University of Washington)

pGREG576-PCF11: pGAL-GFP-PCF11

pGREG576-YHH1: pGAL-GFP-YHH1

pGREG576-RNA14: pGAL-GFP-RNA14

pGREG576-NAB4: pGAL-GFP-NAB4

pGREG plasmids derived from pGREG576: N-ter GFP, URA3 (Jansen et al., 2005)

pFA6a-mCherrry-his3MX6 (Scott Emr's laboratory)

pFA6a-GFP(S65T)-his3MX (Longtine et al., 1998)

pGAL-HO: HO endonuclease expression vector

Oligonucleotides used in this study

For FRET constructs

YHH1-FOR: 5'-ataaatattgaattttcaatgagatctttatgccagggtaagggtcgacggatccccggg-3'

YHH1-REV: 5'-ggttcttattcactactcgaatgcctagtgatagatagatgatcgatgaattcgagctcg-3'

PCF11-FOR: 5'-ggcaaggtcggtttggatgacttaaagaaattggtcacaaaaggtcgacggatccccggg-3'

PCF11-REV: 5'-aatgagcaacccatcaaataccaagaacaaagaatttaataatcgatgaattcgagctcg-3'

CTD-FOR: 5'-aagcaagacgaacaaaagcataatgaaaatgaaaattccagaggtcgacggatccccggg-3'

CTD-REV: 5'-ttatgatacaaaaataaactagaaatattatttgagaaactatcgatgaattcgagctcg-3'

These oligonucleotide primers are suitable with GFP and its different variants used in this study, such as CFP, YFP, Cerulean, Venus and mCherry.

For verification of FRET constructs

YHH1-for1: 5'-aaaccctcgcatggaaagat-3'

YHH1-rev4: 5'-taaaggcaccgtttgcactt-3'

CFP-rev2: 5'-ccccaagtcaaagtagtgac-3'

CFP-for3: 5'-ttttcaagaacttgtcatttgtatag-3'

PCF11-for1: 5'-acggggaaaatatgttgttg-3'

PCF11-rev4: 5'-cagtaacgaagtgtaccctg-3'

CTD-forA: 5'-catgccagaacaaaaaataa-3'

CTD-revD: 5’-ccaatgctctgcggatataa-3’

YFP-rev2: 5’-catggcactcttgaaaaagt-3’

YFP-forC: 5’-agcgtagttgacttgtcggg-3’

YFP-CFP-forC: 5’-aagcgtagttgacttgtcgg-3’

mCherry-rev2: 5’-tgtcggcggggtgcttcacg-3’

mCherry-for3: 5’-agcgtagttgacttgtcggg-3’

For FRAP constructs

5'-GFP-GFP: 5’-attacacatggcatggatgaactatacaaaggcgccagtaaaggagaagaactttt-3’

3'-GFP-GFP: 5’-agaaattcgcttatttagaagtggcgcgccctatttgtatagttcatcca-3’

YHH1-rec1: 5’-gaattcgatatcaagcttatcgataccgtcgacaatgaatgtatatgatgatgt-3’

YHH1-rec2: 5’-gcgtgacataactaattacatgactcgaggtcgactcacttaccctggcataaag-3’

PCF11-rec1: 5’-gaattcgatatcaagcttatcgataccgtcgacaatggatcacgacacagaagt-3’

PCF11-rec2: 5’-gcgtgacataactaattacatgactcgaggtcgacttattttgtgaccaatttct-3’

RNA14-rec1: 5’-gaattcgatatcaagcttatcgataccgtcgacaatgtccagctctacgactcc-3’

RNA14-rec2: 5’-gcgtgacataactaattacatgactcgaggtcgacttaacctgacttggtgctct-3’

NAB4-rec1: 5’-gaattcgatatcaagcttatcgataccgtcgacaatgagctctgacgaagaaga-3’

NAB4-rec2: 5’-gcgtgacataactaattacatgactcgaggtcgacttacctattatatggatggt-3’

For verification of FRAP constructs

GFP(S65T)-rev2: 5’-cttgaaaagcattgaacacc-3’

GFP(S65T)-for3: 5’-gaagatggaagcgttcaact-3’

One4all-A: 5’-gaattcgatatcaagcttatcgatacc-3’

One4all-D: 5’-gcgtcccaaaaccttctcaagcaag-3’

Yhh1-rec-B: 5’-gattcatgctatggattaat-3’

Yhh1-rec-C: 5’-tgcggcaacagatgctgaca-3’

Pcf11-rec-B: 5’-ttagcctttcactagttaaa-3’

Pcf11-rec-C: 5’-atcacaaaacacagcaaata-3’

Rna14-rec-B: 5’-atggcttccactgctctaaa-3’

Rna14-rec-C: 5’-tatttgaaacatcagtggaa-3’

Nab4-rec-B: 5’-tatcttccgtagtgtcccaa-3’

Nab4-rec-C: 5’-cggttacaacccaatgatga-3’

Construction of fluorophores-tagged yeast strains

To construct chromosomally fluorophore-tagged strains, pairs of gene-specific oligonucleotide primers were synthesized, each of which had been designed to share complementary sequences to the fluorophore tag-marker cassette at the 3' end and contain 40-45 base pairs (bp) of homology with a specific gene of interest to allow in-frame fusion of the fluorophore tag at the C-terminal coding region of the gene. Gene-specific cassettes containing a C-terminally positioned fluorophore tag were then generated by PCR using appropriate plasmid templates, which contain the *his3,* kanamycin or hygromycin genes and permit selection of transformed strains in histidine-free, kanamycin or hygromycin-containing media (Longtine et al., 1998). The haploid parent yeast strain CEN.PK2 was transformed with the PCR products, and strains were selected in SD medium (synthetic medium plus glucose) lacking histidine. Insertion of the cassette by homologous recombination was verified by genomic PCR of samples from individual colonies with two pairs of control primers. For each pair, one primer is internal to the fluorophore tag and the other is ORF-specific, both being designed to produce a product of approximately 500 bp.

Mating type switching

To switch the mating type, haploid cells were transformed with the pGAL-HO plasmid according the lithium acetate method (Gietz et al., 1995). Cells were grown over night on selective medium (0.67% yeast nitrogen base without amino acids, 2% glucose) supplemented with appropriate amino acids. Medium was removed and replaced by YPG-lactate medium (1% yeast extract, 2% bacto-peptone, 3% lactate, 2% galactose) for 3h to induce expression of HO endonuclease. 2% glucose was then added in the cultures to down-regulate HO induction and prevent re-cutting of the MAT site. Cells were finally plated on the appropriate agar selective medium and mating type switching of single colonies was verified by mating with α- and a-tester strains.

Droplet test

Yeast strains were grown overnight on YPD medium (1% yeast extract, 2% bacto-peptone, 2% glucose) and ten-fold serial dilutions were prepared from an initial OD_{600} of 0.1. 5µl of each dilution was then spotted on YPD agar-medium and incubated at 30°C and 37°C for two days, or at 15°C for four days.

Microscopy

FRAP cells were loaded on coverslips and images were acquired on a Zeiss LSM 510 confocal microscope using the PlanApochromat 63×/1.4 objective. GFP fluorescence was detected using the 488-nm line of the argon ion laser.

FRET cells were loaded on coverslips, which were placed on the stage of an inverted microscope (Leica DMIRBE, Leica France, Rueil Malmaison, France). A mercury lamp (50 W) was the excitation source for fluorescence. Cells were imaged through a 100x oil objective (Leica PL Fluotar, NA=1.3). The fluorescence corresponding to the different tagged proteins was selected using one of the filter set combinations (Omega; Optophotonics, Eaubonne, France): 440AF35 excitation filter/460DRLP dichroic beam splitter/480AF30 emission filter (CFP); 480AF30 excitation filter/520DRLP dichroic beam splitter/540AF40 emission filter (YFP) and 535DF35 excitation filter/570DRLP dichroic beam splitter/OG590 emission filter (mCherry). The steady-state fluorescence images were acquired with a cooled slow scan high-resolution (1024 x 1024 pixels) charge-coupled device (CCD) camera (Silar, St Petersburg, Russia).

Extract preparation and Western blot analysis

Protein extracts from yeast cells were prepared as described previously (Ohnacker et al., 2000). Western blot analysis was performed according to standard procedures. Briefly, protein samples were separated on either 10 or 12% SDS polyacrylamide gels (Laemmli, 1970). After transfer of proteins to nitrocellulose membranes, the blots were treated according to standard procedures. Primary antibodies used were anti-GFP (JL-8, BD Biosciences) diluted 1:2000 in blotting buffer and anti-phospho serine 2-CTD (H5, Covance) diluted 1:5000 in blotting buffer. Peroxidase-conjugated swine anti-mouse immunoglobulins (DAKO) were used as secondary antibody and chemiluminescence (ECL kit, Amersham) was used for detection.

***In vitro* cleavage and polyadenylation assays**

^{32}P-labelled *CYC1* RNA was prepared as run-off transcripts and gel purified as described previously (Minvielle-Sebastia et al., 1994; Preker et al., 1995).

A standard in vitro processing reaction was carried out in a 25 μl reaction volume containing up to 5 μl of protein extract or purified protein factors. For

polyadenylation assay, the reaction mixture contained 2% polyethylene glycol, 75 mM potassium glutamate, 0.01% NP-40, 2 mM ATP, 1.8 mM magnesium acetate, 20 mM creatine phosphate, 1 mM DTT, 0.2 mg/ml creatine kinase and 0.2 units of RNAguard (Amersham Biosciences). The volume was brought to 25 μl upon addition of buffer E-50 (20 mM Tris-HCl pH 7.9, 0.2 mM EDTA, 50 mM Kac, 0.02% NP-40, 10% glycerol) if was necessary. When the cleavage reaction was assayed, EDTA was used instead of magnesium acetate to prevent poly(A) addition, and degradation of the downstream cleavage fragment (Chen and Moore, 1992). The reactions were incubated 1h at the temperatures indicated in the figure legends. They were then stopped by addition of 75 μl of a stop mix (100 mM Tris-HCl pH7.9, 12.6 mM EDTA, 150 mM NaCl, 1% SDS, 0.2 μg/μl of proteinase K and 0.05 μg/μl glycogen) and incubated for 1h at 42°C. The RNAs were recovered by ethanol precipitation, dried and resuspended in 6 μl of formamide buffer (95% formamide, 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF). Samples were heated at 95°C for 5 min and loaded on a denaturing 6% polyacrylamide gel. The gel was exposed overnight to an X-ray film at -70°C.

Chapter 3: Characterization of Cpf11p, a potential new subunit of CPF

3.1 Introduction

In eukaryotes, accurate processing of the 3' end of most pre-mRNAs is accomplished by a tightly coupled two-step reaction. The first step consists of a site specific endonucleolytic cleavage of the primary transcript, catalyzed by Ysh1p (Mandel et al., 2006; Ryan et al., 2004). Subsequently, the upstream cleavage product is polyadenylated by poly(A) polymerase (Pap1p; Colgan and Manley, 1997; Keller and Minvielle-Sebastia, 1997; Zhao et al., 1999a). Although the process requires only two enzymatic reactions, the specific reaction involves a complex protein machinery that is highly conserved from yeast to mammals (see figure 1.5; For reviews, see Keller and Minvielle-Sebastia, 1997; Wahle and Ruegsegger, 1999; Zhao et al., 1999a). Numerous *trans*-acting factors specifically bind to *cis*-acting elements located in the 3' untranslated region (3' UTR) of the primary transcript to ensure accurate 3' end processing (Dichtl and Keller, 2001; Graber, 2003; Zhao et al., 1999a). Interestingly, they not only exhibit similarities in their sequences but also display equivalent functions in pre-mRNA 3' end maturation (Keller and Minvielle-Sebastia, 1997; Manley and Takagaki, 1996). In contrast, *cis*-acting polyadenylation signals are more conserved in mammals than in yeast, where they are redundant and degenerate.

The separation of the yeast processing factors can be obtained by fractionation of whole-cell extracts. Pre-mRNA 3' end processing in yeast involves four protein factors that can be separated by column chromatography: CPF, which is a multi-protein complex combining the two originally identified 3' end processing factors CF II and PF I (Chen and Moore, 1992; Stumpf and Domdey, 1996; Zhao et al., 1997), CF IA, CF IB and finally Pab1p. The complete pre-mRNA 3' end processing activity can be reconstituted *in vitro* upon combination of these functionally distinct factors (Chen and Moore, 1992). The use of an *in vitro* system has revealed that the cleavage and polyadenylation activities can be physically separated. CPF, CF IA and CF IB are required and sufficient for site-specific cleavage activity, whereas the addition of Pab1p is necessary for specific polyadenylation to occur (Kessler et al., 1996; Minvielle-Sebastia et al., 1997; Preker et al., 1997).

By affinity purification via tagging of one of the CPF subunits, an attempt to

find new components of CPF identified six novels proteins in addition to the nine previously described components (Dichtl et al., 2002b). The role of these new polypeptides, Ref2p, Pti1p, Swd2p, Glc7p, Ssu72p and Cpf11p in pre-mRNA 3' end formation was then analyzed. The non-essential Ref2p (RNA end formation 2) protein is an RNA-binding protein, and was shown to stimulate processing at weak poly(A) sites (Russnak et al., 1995). A possible involvement of Ref2p in the regulation of the poly(A) tail length has also been proposed (Mangus et al., 2004). Another novel protein, Pti1p, was originally isolated as an interactor with *PTA1* in a yeast two-hybrid screen (Hübner and Keller, unpublished data). A role for Pti1p in 3' end processing has been proposed as alterations in the cleavage site choice have been revealed in *pti1* mutant strains (Skaar and Greenleaf, 2002). Interestingly, while Pti1p and Ref2p display only minor roles in pre-mRNA 3' end processing, both have been shown to be involved in cleavage-dependent/polyadenylation-independent 3' end formation of snoRNAs (Dheur et al., 2003; Nedea et al., 2003). Pti1p probably uncouples cleavage and polyadenylation and may function in coordination with the Nrd1p-dependent pathway for 3' end formation of the snoRNA transcripts (Dheur et al., 2003). Also involved in snoRNAs 3' end formation, Swd2p is an essential WD-40 repeat protein, as temperature sensitive *swd2* strains were shown to be defective in 3' end formation of specific mRNAs and snoRNAs (Dichtl et al., 2004). Moreover, Swd2p plays a role in RNAP II transcription termination (Cheng et al., 2004), a process which is dependent on the association of CPF and CF IA with elongating RNAP II and the recognition of the poly(A) signals by these factors (Barilla et al., 2001; Birse et al., 1998; Dichtl et al., 2002b). Hence, as a component of CPF, Swd2p could help to coordinate the activities of both the transcription and 3' end processing machineries. Interestingly, Swd2p has been found to be physically associated not only with CPF but also with the SET1 complex (SET1C), which functions in the methylation of histone 3 lysine 4 (Miller et al., 2001; Nagy et al., 2002; Roguev et al., 2001). It was proposed that Swd2p is required for the stability and the integrity of SET1C and that it is probably part of the core complex (Dehe et al., 2006). Hence, Swd2p has a dual function in pre-mRNA processing and histone tail modification. However, the roles of Swd2p as a component of CPF or of SET1C are functionally independent (Dichtl et al., 2004). Another protein involved in regulating a wide range of cellular processes in yeast and also found in CPF is Glc7p, the catalytic subunit of protein phosphatase 1. Recently, it has been demonstrated that a subunit of CPF,

Pta1p, is a substrate of Glc7p (He and Moore, 2005), suggesting that CPF activity can be autoregulated by cycles of phosphorylation and dephosphorylation. A second phosphatase in CPF is Ssu72p, an essential tyrosine phosphatase that specifically dephosphorylates serine 5 of the C-terminal domain of largest subunit of RNAP II (Krishnamurthy et al., 2004). Ssu72p is stably associated with CPF and bridges the CPF subunits Pta1p and Ydh1p/Cft2p, the general transcription factor TFIIB, and RNAP II via Rpb2p (Dichtl et al., 2002b; Ganem et al., 2003). The last protein identified is a novel 20 kDa protein that was named Cpf11p. It is encoded by the *YDL094c* gene, which has been shown to be non-essential since a systematic deletion of this gene is not lethal for the cell (Giaever et al., 2002). Instead of Cpf11p and not identified by Dichtl et al. (2002b), Syc1p is another 20 kDa non-essential protein that was found by other groups to copurify with CPF (Gavin et al., 2002; He et al., 2003; Nedea et al., 2003). Syc1p is highly homologous to the C-terminal domain of Ysh1p, the endonuclease subunit of CPF. Thus, except for Cpf11p, all of the newly identified polypeptides have been studied to a greater detail and have been shown to have a role in pre-mRNAs 3' end processing. Herein we report our investigations aimed to verify the presence of Cpf11p in CPF and to characterize the role of this putative new CPF subunit in the context of 3' end formation. We showed that mutant strains depleted for Cpf11p still efficiently processed RNA substrates at their 3' end. We could also demonstrate by TAP tag purification that Cpf11p does not co-purify the known components of CPF, indicating that Cpf11p is unlikely to be an integral constituent of the CPF complex.

3.2 Results

3.2.1 Cpf11p is a non-essential protein with no homologues in other organisms

The 20 kDa protein that copurified with CPF in the experiment presented in (Dichtl et al., 2002b) was analyzed by mass spectrometry. Results indicated a protein that corresponds to the gene product of the open reading frame *YDL094c*, named *CPF11* (for cleavage and polyadenylation factor number eleven). This gene, which is located on chromosome IV, encodes a predicted protein composed of 169 amino acids and with a molecular weight of 19.6 kDa. To analyze the importance of the protein for cell

viability, it is sufficient to generate null mutants of the gene of interest. A systematic gene-deletion approach of *S. cerevisiae* has revealed that *CPF11*-depleted strains are still viable, indicating that this gene is not essential (Giaever et al., 2002). According to the fact that this is a newly discovered protein, nothing was known about the molecular function and the cellular localization of Cpf11p when we initiated this work.

In order to make functional prediction for the uncharacterized Cpf11p and to analyze whether sequence similarities can be found with polypeptides in other species, we performed sequence homology searches. We used the BLASTP and TBLASTN algorithms (Altschul et al., 1997) of the NCBI server (http://www.ncbi.nlm.nih.gov/blast) to analyze the *S. cerevisiae* Cpf11p protein sequence. Results revealed that Cpf11p shares 38% identity in amino acids with another uncharacterized *S. cerevisiae* protein, encoded by the *YDL096c* ORF, also located on chromosome IV. Most of the CPF components have homology counterparts in other species (Keller and Minvielle-Sebastia, 1997; Manley and Takagaki, 1996). Surprisingly, no polypeptide sequences in other species were found to have significant similarities with *CPF11* (data not shown). Furthermore, with the exception of a central predicted transmembrane domain, no similarity with any known conserved domain was revealed, thus providing no information for a potential function of the protein. Taken together, these data indicate that Cpf11p is not conserved throughout evolution and suggest that the protein does not carry an essential function in the cells.

3.2.2 Cpf11p is not involved in pre-mRNA 3' end processing

A powerful way to determine gene function is the phenotypic analysis of mutants in which the gene has been deleted. In order to investigate the role of Cpf11p in 3' end processing of pre-mRNAs, we used the yeast strain BY4742 in which the gene *CPF11* was replaced by a kanamycine gene (referred as *cpf11*Δ). The null mutant of *CPF11* was generated by a PCR-based approach in which a kanamycine marker cassette, carrying flanking short sequences that are homologous to the target locus, replaced the *CPF11* target gene by exploiting the high rate of homologous recombination in yeast.

3’ end processing activities can be reconstituted *in vitro* using protein extracts or purified 3’ end processing factors with a synthetic RNA substrate that was *in vitro* transcribed and radioactively labeled by incorporation of ^{32}P α-UTP (Butler and Platt, 1988). We prepared protein extracts from the *cpf11*Δ strain and tested them for their ability to cleave and polyadenylate *in vitro* the RNA substrates *CYC1* and pre-*CYC1*, respectively, at 15°C, 30°C and 37°C. The absence of Cpf11p did not affect pre-mRNA 3’ end processing *in vitro* because protein extracts from mutant strains displayed wild type-like cleavage and polyadenylation activities (Figure 3.1). These results suggest that Cpf11p is not required for pre-mRNA 3’ end formation *in vitro*.

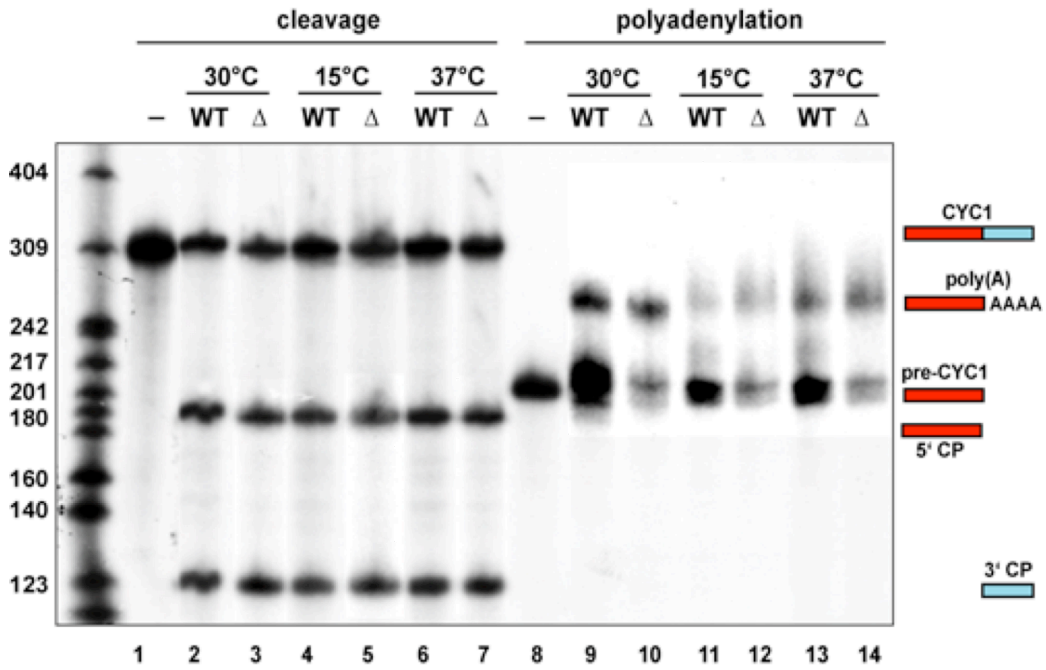

Figure 3.1. The *cpf11*Δ strain is not deficient in pre-mRNA 3’ end processing.
In vitro cleavage (lanes 1-7) and polyadenylation assays (lanes 8-14) were carried out with the *CYC1* and pre-*CYC1* RNA precursors, respectively, in the presence of wild type (WT) or Cpf11p lacking protein extracts (Δ) at the temperatures indicated on top of the figure. The RNA precursors (*CYC1* and pre-*CYC1*), 5’ cleavage product (5’ CP), 3’ cleavage product (3’ CP) and polyadenylated product [poly(A)] are indicated on the right. The molecular marker (in number of nt) is shown on the left.

3.2.3 Cpf11p is not a new component of the cleavage and polyadenylation factor

The identification of Cpf11p by MS in CPF does not necessarily mean that this protein is a stable part of the complex. We have previously shown that the protein has no influence on pre-mRNA 3’ end processing activities (see 3.2.2) and thus could

very well be a contaminant of the purification. To test whether Cpf11p stably interacts with CPF, we constructed TAP-tagged (TAP: tandem affinity purification) versions of Cpf11p by inserting the TAP tag to the C-terminal end of the protein.

3.2.3.1 Cell viability of Cpf11-TAP strains is not affected by the insertion of the TAP tag

We have used the BMA41 (Baudin-Baillieu et al., 1997) haploid yeast cells to construct strains expressing a C-terminally TAP-tagged version of Cpf11p (see experimental procedures). The TAP tag method is a very useful procedure that allows fast purification of macromolecular complexes under native conditions and when expressed at their natural level (Puig et al., 2001; Rigaut et al., 1999). Combined with mass spectrometry, the TAP strategy permits the identification of proteins interacting with the protein of interest. Several clones were obtained from which only three (Cpf11-TAP-1, Cpf11-TAP-2 and Cpf11-TAP-3) were used for further analysis.

In order to control whether the insertion of the TAP tag downstream of *CPF11* did affect the cells in their ability to grow, wild type and modified strains were grown at different temperatures, in the presence of glucose or galactose. Serial dilutions of the cultures were prepared and spotted on YPD (yeast peptone dextrose; rich medium plus glucose) and YPG (yeast peptone galactose; rich medium plus galactose) agar plates, which were incubated at 30°C and 37°C for two days, or at 15°C for five days. As shown in figure 3.2, the insertion of the TAP tag did not affect cell growth of the modified strains as no growth defect was observed in comparison to the wild type cells. This data indicate that the TAP tag cassette integrated into the genome has no undesirable effects on the host cells and resulted in the generation of new viable yeast strains that harbor a C-terminal TAP-tagged version of *CPF11*.

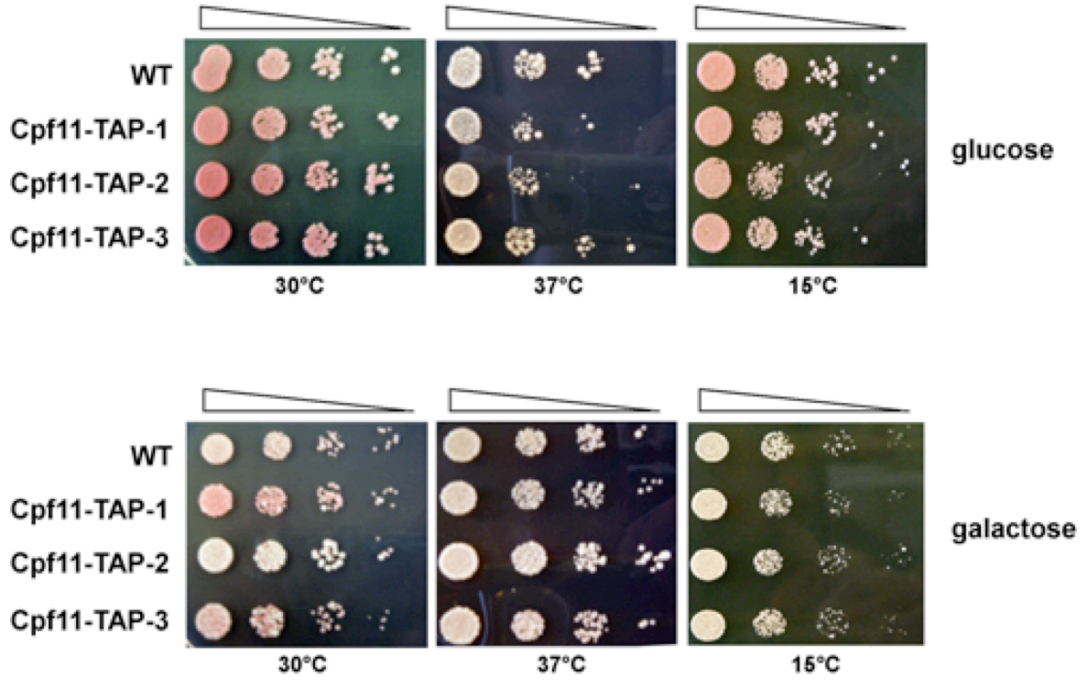

Figure 3.2. TAP tagging of Cpf11p did not affect cell growth.
Ten-fold serial dilutions of the strains indicated on the left were spotted on YPD or YPG agar medium and incubated at the temperatures indicated at the bottom of the panels for two days (30°C and 37°C) or five days (15°C).

3.2.3.2 Expression of Cpf11-TAP is regulated in a sugar-dependent manner

The fact that the TAP tag is inserted at the appropriate site does not necessarily mean that the fusion protein will be correctly expressed by the modified strains. Hence, we wanted to compare the expression level of Cpf11-TAP in modified strains with the endogenous expression level of Cpf11p in WT strains. This approach however was not feasible, as no antibody raised against Cpf11p was available and as several attempts to produce antisera were not successful. Thus, the only conceivable approach was to verify that Cpf11-TAP is indeed expressed in the modified strains. For this, we took advantage of the presence of protein A (ProtA) domains in the TAP tag to test the expression of Cpf11-TAP by Western blotting using an antibody raised against ProtA. We therefore prepared total protein extracts from the modified strains Cpf11-TAP-1, Cpf11-TAP-2 and Cpf11-TAP-3. Despite repeated attempts, we could not detect any signal that would suggest that Cpf11-TAP is indeed expressed (Figure 3.3 lanes 2-4). We consequently tested different growth conditions under the assumption

that the expression of Cpf11p might be conditional. For example, Cpf11p could be expressed in response to a stress or an environmental change. The strain Cpf11-TAP-1 was grown at different conditions that generate a stress for the cells: (i) heat shock at 42°C was tested to verify whether Cpf11p is anyhow related to heat shock proteins; (ii) growth in the presence of potassium acetate, which forces cells to sporulate; (iii) cells were grown over a long incubation time of five-days to examine the effect of a lack of oxygen; (iv) finally, cells were grown in the presence of galactose instead of glucose to analyze the effect of a change in the carbon source (Figure 3.3 lanes 5-8). Surprisingly, a clear signal in Western blot with at a size that corresponds to Cpf11-TAP (about 35-40 kDa due to the fusion with the TAP tag) was detected with protein extracts prepared from cultures grown in the presence of galactose only (Figure 3.3 lane 8). In contrast, no signals were detected in the other extracts (Figure 3.3 lanes 5-7), suggesting that only the type of sugar present in the medium enables regulation of Cpf11p expression. Thus, one possibility could be that galactose induces Cpf11p expression. In the contrary, the presence of glucose would silence the expression of the gene, or would keep it at a basal level that would render the protein undetectable by Western blotting. So far, none of the CPF subunits showed such a regulation. This finding raises the question whether this type of regulation is biologically relevant.

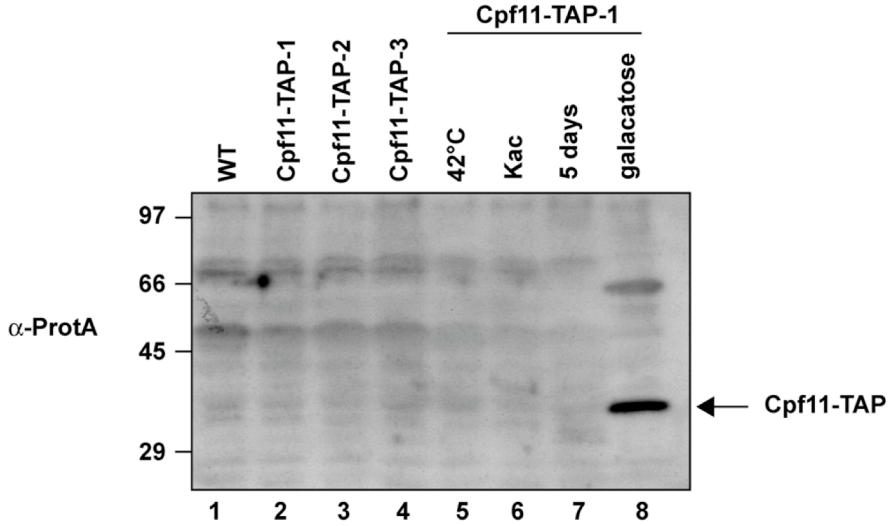

Figure 3.3. Cpf11-TAP expression is induced by galactose.
Protein extracts prepared from the strains indicated on top were analyzed by Western blotting with an antibody directed against ProtA. Yeast cells anlyzed in lanes 1-7 were grown in the presence of glucose, whereas galactose was used in lane 8. The strain Cpf11-TAP-1 was grown at different conditions: lane 5, the culture was subjected to a one-hour heat shock at 42°C; lane 6, potassium acetate was added to the medium; lane 7, cells were grown for five days; lane 8, galactose was used as the carbon source instead of glucose. The protein detected is indicated on the right of the panel and the molecular weight marker on the left (in kDa).

3.2.3.3 Affinity purification of Cpf11-TAP does not pull down CPF

To determine whether Cpf11p is an integral component of the CPF complex, we affinity purified TAP-tagged Cpf11p. Because Cpf11p expression seems to be dependent on the nature of the sugar used as carbon source, purifications of Cpf11-TAP were performed in parallel using protein extracts prepared from the strain Cpf11-TAP-1 that was grown either in the presence of glucose or in the presence of galactose. If Cpf11p is a genuine subunit of CPF, purification of the TAP-tagged version of the protein should result in the co-purification of the other components (at least of the strongest interacting partners) of the 3' end processing factor, as it is the case when the purification is performed from the ProtA-Pfs2p strain (see 4.2.1 of Chapter 4 and figure 3.4A lanes 3-4). Surprisingly, purifications from Cpf11-TAP-1 protein extracts prepared from cells that were grown either in the presence of glucose or galactose, failed to pull down CPF subunits in both cases (Figure 3.4A lanes 5-8). None of the bands corresponding to the CPF subunits were detected when the final eluates were analyzed by gel electrophoresis followed by silverstaining. In contrast, purification from protein extracts containing ProtA-Pfs2p, which was used as positive control, resulted in the elution of the whole CPF complex (Figure 3.4A lanes 3-4). It has to be noticed here that Yth1p stains poorly with silver (Preker et al., 1997) and is therefore not visible on the protein gel presented in figure 3.4A.

Yet, purification of Cpf11-TAP yielded some proteins that are different from those characterized in CPF (Figure 3.4A lanes 5-8). The pattern of bands obtained with the strain Cpf11-TAP-1 appeared to be significantly different from the mock purification obtained with protein extract from a wild type strain (Figure 3.4A, compare lanes 5-8 with 1-2). Indeed, four polypeptides were found to coelute specifically with Cpf11-TAP (Figure 3.4A lane 7, asterisks). Interestingly, these proteins were detected when the cells were grown both with glucose or galactose, but signals were more weakly stained when glucose was used, indicating lower amounts of proteins (compare the intensity of the bands in lanes 7-8 and 5-6). Furthermore, Cpf11-CBP, which is the remaining part of the fusion protein after cleavage by the TEV protease (see experimental procedures), could not be detected on the silverstained gel: it is most likely hidden by the strong TEV protease signal which runs approximately at the same size at ~25 kDa (Figure 3.4A lanes 7-8).

To confirm the observations of figure 3.3A, we have searched for the presence of several subunits of CPF in the Cpf11-TAP eluates by Western blotting. For this, specific antibodies against Ysh1p, Pfs2p and Ssu72p were used to screen for the presence of the corresponding proteins. Pfs2p was readily detected in protein extracts from the Cpf11-TAP strains (Figure 3.4B lanes 1-3). The shift in size of the band corresponding to the immunodetection of Pfs2p in the ProtA-Pfs2p protein extract is due to the presence of the ProtA tag (Figure 3.4B lane 1). The signal for Ysh1p was weak but clearly visible (Figure 3.4B lanes 1-3). In contrast, Ssu72p was not immunodetected neither in ProtA-Pfs2p protein extracts nor in Cpf11-TAP protein extracts (Figure 3.4B lanes 1-3). Actually, we observed that the Ssu72p antibody was rather inefficient to detect Ssu72p in protein extracts. Indeed, while this antibody did not immunodetect Ssu72p in ProtA-Pfs2p protein extract (Figure 3.4B lane 1), it nicely revealed the presence of Ssu72p in the eluate obtained after purification from the same extract (Figure 3.4B lane 4). Ysh1p, Pfs2p and Ssu72p were readily immunodetected in the eluate obtained from the positive control ProtA-Pfs2p protein extract, indicating that CPF was efficiently purified by this procedure (Figure 3.4B, lane 4). Moreover, the IgG matrix was tested after the elution to verify that no material was still bound to it. No protein was detected from the matrix used for the purification from Cpf11-TAP-1 protein extracts (Figure 3.4B lanes 8-9), whereas only very weak immunolabeling was still detectable in the fraction obtained from ProtA-Pfs2p extracts, showing that CPF was not totally recovered by TEV proteolytic cleavage (Figure 3.4B lane 7). In contrast, none of the tested CPF subunits could be observed in the eluates obtained from Cpf11-TAP cells that were grown with glucose or with galactose (Figure 3.4B, compare lanes 5-6 with lane 4). This correlates well with the results described in figure 3.4A. To further look for the presence of Cpf11p in the eluates of the purifications obtained from Cpf11-TAP-1 extracts, we have used an antibody that specifically recognizes the CBP part of the TAP tag when cleavage by the TEV protease as occured. As shown in figure 3.4D, a signal corresponding well to Cpf11-CBP was immunodetected in eluates obtained from galactose extracts (lane 4). The same but weaker band was also present in the glucose fraction (lane 3). We have previously shown that Cpf11-TAP was not detected in Cpf11-TAP-1 extracts prepared from cells grown in the presence of glucose (Figures 3.3 and 3.4C, lane 3). Thus, the result shown in figure 3.4D, lane 3, suggests that Cpf11-TAP

expression is not completely silenced by glucose, but instead repressed to a basal level in the absence of galactose.

Taken together, these results confirm: (i) our finding that Cpf11-TAP expression is induced by galactose; (ii) our hypothesis that Cpf11-TAP is expressed at a basal level in the presence of glucose. More importantly, these results show that Cpf11-TAP is not associated with CPF, but instead with other (unkown) proteins.

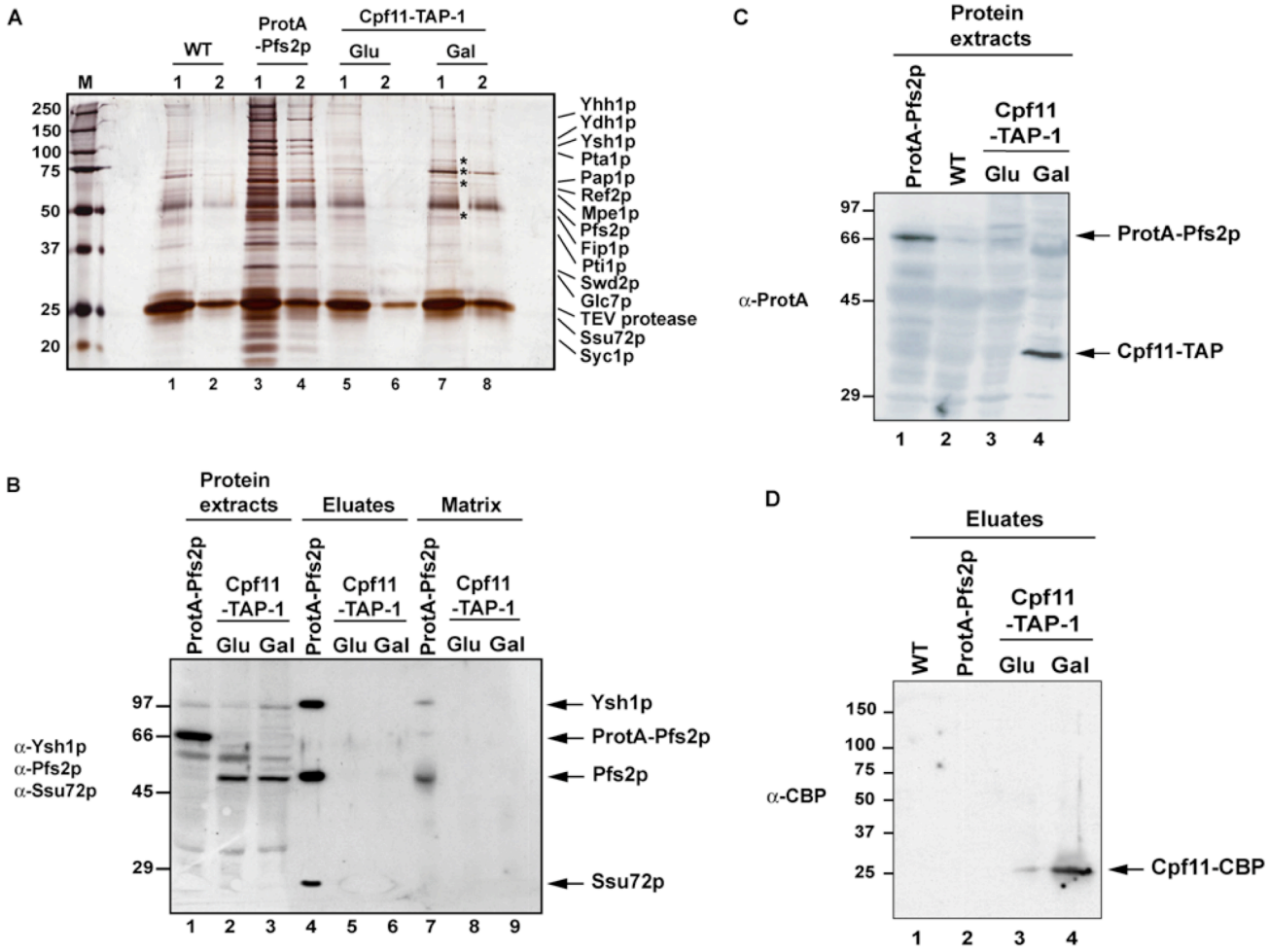

Figure 3.4. TAP-tagged purification of Cpf11-TAP does not pull down CPF.
(A) Purification products obtained from protein extracts prepared from strains indicated on top (WT: wild type). Elution fractions 1 and 2 (as indicated on top) were analyzed on a 10% denaturing gel and stained with silver. Cpf11-TAP-1 strain was grown either in glucose (lane 5-6) or galactose containing medium (lane 7-8). Proteins detected are indicated on the right of the panel. Proteins that coelute with Cpf11-TAP are indicated by an asterisk (*). The molecular weight marker is indicated on the left (in kDa).
(B) Immunodetection of Ysh1p, Pfs2p and Ssu72p CPF subunits by their specific antibodies in protein extracts (lanes 1-3), eluates (lanes 4-6) and IgG matrices (lanes 7-9) obtained from strains indicated on the top of the panel. The ProtA-Pfs2p strain was grown in the presence of glucose (lanes 1, 4 and 7). The Cpf11-TAP-1 strain was grown either with glucose (lanes 2, 5 and 8) or with galactose (lanes 3, 6 and 9), as indicated on the top of the panel. Proteins immunodetected are indicated on the right of the panel, the antibodies used and the molecular weight marker on the left (in kDa).
(C) Immunodetection of ProtA-fusions by the anti-ProtA antibody in protein extracts prepared from the strains indicated on the top of the panel (WT: wild type). Cells were grown in the presence of glucose (lanes 1-3) or galactose (lane 4). Proteins immunodetected are indicated on the right of the panel, the antibody used and the molecular weight marker on the left (in kDa).
(D) Immunodetection of Cpf11-CBP in the eluates obtained from the strains indicated on top of the panel (WT: wild type). The antibody used recognizes the remaining portion (CBP) of the TAP tag after cleavage by the TEV protease. The antibody used and the molecular weight marker are indicated on the left (in kDa).

3.3 Discussion

In the last few years, MS analyses of CPF have identified new proteins associated with the previously known subunits (Dichtl et al., 2002b; Gavin et al., 2002; Nedea et al., 2003). Interestingly, a MS study conducted in our laboratory (Dichtl et al., 2002b) was the only one that identified the gene product of the *YDL094c* gene, named Cpf11p, as a putative new CPF subunit. We analyzed Cpf11p in order to characterize its putative involvement in pre-mRNA 3' end formation, based on the hypothesis that it could be a new CPF component. Therefore we first had to confirm that Cpf11p is a real CPF component as its copurification with CPF has been reported only once, despite the numerous MS analyses performed by other groups.

Cpf11p has no homologues in other species

Astonishingly, Cpf11p is not conserved among eukaryotes, as sequence similarity searches in databases did not lead to the identification of any Cpf11p homologue and not even to some significant degree of homology with proteins from any other species than *S. cerevisiae*. This finding strongly contrasts with the fact that most CPF components have mammalian homologues, or at least present structural similarities with CPSF components that fulfil similar tasks (Barabino et al., 2000; Ohnacker et al., 2000; Zhao et al., 1999a). In fact, the only protein in the databases that clearly showed similarity when performing a BLAST search is encoded by another *S. cerevisiae* gene, the *YDL096c* gene. It is located on the same chromosome (IV) as *CPF11*, directly downstream of it on the same DNA strand and encodes a putative protein that shares 38% identity with Cpf11p. These observations suggest that these two genes might have arisen from a common ancestral gene, possibly by duplication and subsequent sequence divergence during evolution. It is now well established that *S. cerevisiae* has experienced an ancient genome duplication, which generated many protein duplicates (Kellis et al., 2004; Wolfe and Shields, 1997). After genome duplication, one copy of most pairs of genes was lost, but some pairs of duplicated genes have been preserved. *YDL096c* and *CPF11* could well be such a preserved duplicated pair of genes, but it is also conceivable that they arose by domain swapping, i.e. a part of a protein may have been duplicated and fused to another

protein. Duplicated genes can diverge for generation of new gene functions (neofunctionalization) or for subdividing multiple functions (subfunctionalization) through complementary degeneration but they can also be functionally redundant in yeasts (Guan et al., 2007; van Hoof, 2005). Hence, assuming that both gene products could be functionally redundant or functionally complementary, information on the *YDL096c* gene function might help us to predict in which biological process could *CPF11* be involved. Unfortunately, except the fact that *YDL096c*, as *CPF11* is, a non-essential gene, its function is not clearly known. Recently, results from a large-scale screen of viable yeast deletions have suggested that *YDL096c*, also now referred to as *OPI6*, may be involved in the pathway that regulates the transcription of the phospholipidic biosynthetic structural genes (Hancock et al., 2006). Indeed, the *YDL096c* mutant phenotype corresponds to an Opi$^-$ phenotype, an abbreviation for mutant that overproduces and excretes inositol (Opi) into the growth medium in the absence of inositol and choline (Greenberg et al., 1982). However, this possible function of *YDL096c* remains to be investigated further and has no direct relevance for our study of Cpf11p as a putative CPF component.

CPF11 is not involved in pre-mRNA 3' end processing in vitro

As a non-essential gene, *CPF11* was unlikely to be required for the crucial 3' end formation reactions, but its requirement at other levels of the CPF activity, such as in modulation of mRNA polyadenylation or in stabilization of the CPF machinery, was still conceivable. However, we have shown that depletion of *CPF11* is not detrimental to 3' end processing activities *in vitro* since protein extracts prepared from a depleted strain (*cpf11*Δ) are as active in cleavage and polyadenylation as the wild-type extract, even at low and high temperatures (Figure 3.1). This suggests that Cpf11p may not be involved in pre-mRNA 3' end formation and thereby may also not be part of CPF. Indeed, mutations in proteins involved in pre-mRNA 3' end processing generally result in severe altered phenotypes in cleavage and/or polyadenylation activities when tested *in vitro* (Dichtl et al., 2004; Minvielle-Sebastia et al., 1994; Preker et al., 1997). However, there are a few exceptions as some CPF subunits have also been shown to be dispensable for proper 3' end processing activities *in vitro*. For example, null mutation of the non-essential *REF2* gene does not significantly affect wild-type RNA substrates processing *in vitro* (Dheur et al., 2003; Russnak et al., 1995). Similarly,

Pti1p-depleted extracts, or protein extracts prepared from temperature-sensitive *pti1* mutants, do not exhibit defects in pre-mRNA 3' end processing activities (Dheur et al., 2003; W. Hübner, unpublished data). Nevertheless, both Ref2p and Pti1p undeniably belong to CPF and have even been shown to interact with each other and also with other CPF subunits (Cheng et al., 2004; Nedea et al., 2003; Skaar and Greenleaf, 2002). Moreover, although their essential functions are mostly related to 3' end formation of snoRNAs (Dheur et al., 2003; Nedea et al., 2003), Ref2p and Pti1p have also been proposed to be involved in 3' end formation of pre-mRNAs as a negative regulator of the poly(A) synthesis and as a regulator of the cleavage site selection of some transcripts, respectively (Mangus et al., 2004; Skaar and Greenleaf, 2002). Thus, the observation that the depletion of Cpf11p does not affect CPF activity *in vitro* is not sufficient to argue that Cpf11p is not an integral component of CPF. Consequently, the role of Cpf11p in mRNA maturation needed to be further investigated by analyzing its putative association with the 3' end processing complex.

To test whether Cpf11p stably interacts with CPF, we constructed a yeast strain that expresses a TAP-tagged version of *CPF11*. Surprisingly, when verifying that the tagged protein was correctly expressed in the modified strain, we found Cpf11-TAP to be regulated in a sugar-dependent manner (Figure 3.3). Indeed, our data indicate that Cpf11-TAP was expressed at a basal level in the presence of glucose and that the expression increased when galactose was the unique source of carbon in the medium. The expression level related to glucose was too low to detect the fusion protein in total cell extracts by immunodetection methods. In contrast, Cpf11p was immunodetected in the eluates, indicating that its expression was not completely switched off by glucose. The regulation of *CPF11*, as a putative CPF component, is striking as none of the previously characterized CPF subunits is regulated in such a sugar-dependent manner. Indeed, as being the main factor involved in pre-mRNA 3' end processing, which is an essential process for the cell, CPF needs to remain functional independently of environmental changes, such as changes of the carbon source. This raises the question of the physiological relevance of the presence of a protein regulated in a sugar-dependent manner, such as of Cpf11p, into the CPF complex and thus challenges its putative permanent association with CPF.

Cpf11p does not interact with CPF components

The fact that the purification of TAP-tagged Cpf11p does not yield the known CPF subunits strongly suggests that Cpf11p is not stably associated with the complex. It remains however possible that the TAP tag situated at the C-terminus of Cpf11p prevents some important interactions with other CPF subunits. Such an effect of the position of the tag has already been observed with Ssu72p since its C-terminal TAP tagged version did not show a stable association with 3' end factor (Gavin et al., 2002), whereas the N-terminal TAP tagged version did (Dichtl et al., 2002b). Therefore, we tried to fuse the TAP tag to the N-terminus of the protein to investigate this hypothesis. Despite of several attempts, no strain expressing a N-terminal TAP tagged version of Cpf11p could be obtained. However, in accordance with our results, MS analyses preformed by other groups in order to identify new polypeptides associated with CPF have never found Cpf11p in close relation with CPF (Gavin et al., 2002; Nedea et al., 2003). These MS studies have identified six new subunits of CPF. Interestingly, five among these novel proteins, namely Ref2p, Pti1p, Swd2p, Glc7p and Ssu72p were also identified by Dichtl et al. (2002b). But, in contrast to Dichtl et al. (2002b) who identified Cpf11p as the sixth novel protein associated with CPF, the other analyses did not identify Cpf11p but instead another protein of similar size (~20 kDa), Syc1p, which is highly conserved among *Saccharomyces* species (Gavin et al., 2002; He et al., 2003; Nedea et al., 2003; Zhelkovsky et al., 2006). Syc1p is a non-essential protein that shows a high degree of similarity with the C-terminal region of Ysh1p (38% identity and 52% similarity over 188 amino acids). It has been proposed to work as a negative regulator of the 3' end processing by possibly competing with sites that interact with the Ysh1p C-terminus (Zhelkovsky et al., 2006). Like for the other five new putative CPF-associated factors identified, the purification of TAP-tagged Syc1p yielded the polypeptides present in CPF, indicating that it is a genuine component of CPF (Nedea et al., 2003). Likewise, if Cpf11p would be an integral subunit of CPF, purification of the TAP-tagged version would have yielded the polypeptides associated with CPF. Our data show that this is not the case (Figure 3.4). Furthermore, additional MS analysis we carried out in order to identify other putative new components of CPF, have confirmed the presence of Ref2p, Pti1p, Swd2p, Glc7p, Ssu72p and Syc1p proteins in association with the complex (see 4.2.2

of Chapter 4). Strikingly, Cpf11p was never identified despite the numerous attempts we performed, arguing once more against the presence of this protein in CPF.

It is also possible that Cpf11p is able to contact CPF only transiently under certain conditions. This would explain why it could have contaminated the CPF preparation of Dichtl et al. (2002b) and therefore has been wrongly assigned as a CPF component.

The CPF protein profile of protA-tagged Pfs2p (Figure 3.4A lanes 3-4) coincides with the complex composition obtained with other protA-tagged CPF factors, such as with TAP-tagged Yth1p (Barabino et al., 2000; Ohnacker et al., 2000) or with TAP-tagged Yhh1p, Ydh1p, Pta1p, Ref2p, Mpe1p and Pti1p (Gavin et al., 2002; Nedea et al., 2003; Vo et al., 2001). The purification of CPF via the ProtA-tagged version of the Pfs2p subunit is one of the most efficient ways to obtain copurification of all the known CPF subunits in a reproducible manner. In comparison, it is possible that the TAP-tagged purification of Cpf11p is less efficient. In this case, the amount of proteins recovered from the purification could be too low for identification on a gel stained with silver, although silverstaining method has a detection limit as low as 10 ng of proteins. In addition, we used an immunodetection method to detect potential 3' end processing factors in the eluates obtained from the purification of Cpf11-TAP. The use of specific antibodies is a very sensitive method that allows a detection of lower amounts of proteins (less than 1 ng) than any gel staining method. But, in accordance with the results of the silver stained gel (Figure 3.4), the use of antibodies against several CPF subunits has clearly revealed that none of the tested 3' end processing factors was present in the eluates obtained from the purification of Cpf11-TAP (Figure 3.4B). In contrast, the CPF subunits we tested were all immunodetected in the eluates obtained from ProtA-Pfs2p extracts. Hence, the fact that we did not detect CPF subunits in association to Cpf11-TAP should not be due to a low sensitivity of the detection methods used.

Cpf11p antibody as a missing tool

We could not analyze the expression of Cpf11-TAP in comparison to the wild type expression of Cpf11p, nor directly trace Cpf11p over the whole purification procedure, since there is no specific anti-Cpf11p antibody available. We generated two types of recombinant proteins, GST-Cpf11p and H6-Cpf11p, but the amount of

soluble proteins that we obtained was never sufficient to allow the production of an antibody (data not shown). The hydrophobic feature of the protein could be responsible for the low solubility of recombinant Cpf11p. As a consequence of the lack of a specific antibody for Cpf11p, we have used anti-ProtA and anti-CBP antibodies to follow the Cpf11-TAP fusion protein over the purification procedure.

A specific anti-Cpf11p antibody would also be needed to verify that Cpf11p was indeed truly present in the original fraction of Dichtl et al. (2002b). Although it is highly speculative, it is possible that Cpf11p was identified as a false positive (also referred to as false identification) of the MS analysis.

Investigating protein-protein interactions of Cpf11p

Other procedures that are widely used to test for protein-protein interactions, like GST pull down experiments or two-hybrid screens, were also used in our laboratory to search for putative Cpf11p interacting partners. Surprisingly, results from GST pull down experiments have found two different potential interacting partners for Cpf11p among the CPF components. Both GST-tagged Mpe1p (PhD Thesis of Martin Sadowski, 2002) and Ydh1p (Kyburz et al., 2003) were shown to interact with *in vitro* translated Cpf11p. Although these data seem to be in contradiction with our results, the results of both experiments, GST pull down and Cpf11-TAP purification cannot be directly compared, as their relevance for putative *in vivo* association between Cpf11p and CPF is not the same. Indeed, two proteins known as interacting partners will very often give a positive result in GST pull down, whereas a positive signal obtained when testing two candidates does not necessarily mean that these two proteins truly interact under physiological conditions. However, the GST pull down approach is very useful to investigate the interaction network created by the different subunits in the context of a multiprotein complex, such as CPF. A conditional binding of Cpf11p to CPF in response to an appropriate signal cannot be excluded, but it would not make Cpf11p an integral subunit of CPF *per se*. We also tried, in collaboration with Hybrigenics S.A. in Paris, a two-hybrid screen approach with Cpf11p as bait, which did not reveal any potential protein-protein interactions.

In summary, the combined results of our work indicate that Cpf11p is not a genuine subunit of CPF. Further work could however be done to investigate whether Cpf11p can be transiently associated to CPF. We have shown that Cpf11-TAP

copurified with four polypeptides, suggesting that Cpf11p displays specific interactions *in vivo.* Thus, it would be interesting to identify these interacting partners, for example by MS, to further characterize the function of Cpf11p. The analysis of double mutants of *CPF11* and *OPI6* could also be informative on the biological role of Cpf11p.

3.4 Experimental procedures

Yeast strains, plasmids and media

Yeast strains were grown on YPD/YPG (1% yeast extract, 2% bacto-peptone, 2% glucose/galactose), on SD-TRP medium (synthetic medium plus 2% glucose without tryptophan) or on YPD supplemented with 200 μg/ml of kanamycine, as indicated in the text or figure legends.

Yeast strains used in this study and their genotypes are the following: YWK192 (*MATα; his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0*), BY4742 (*MATα; his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0; ydl094c::KanMX4*), BMA41 (*MATa/α; ura3-1/ura3-1; leu2-3,112/leu2-3,112; trp1Δ/trp1Δ; his3-11,15/his3-11,15; ade2-1/ade2-1; can1-100/can1-100*), BMA41α (*MATα; ura3-1; leu2-3,112; trp1Δ; his3-11,15; ade2-1; can1-100*), BMA41a (*MATa; ura3-1; leu2-3,112; trp1Δ; his3-11,15; ade2-1; can1-100*), YBP101 (*MATa; ura3-1; leu2-3,112; trp1Δ; his3-11,15; ade2-1; can1-100;* Cpf11 C-ter. TAP-11, *K.l. TRP1*), YBP102 (*MATα; ura3-1; leu2-3,112; trp1Δ; his3-11,15; ade2-1; can1-100;* Cpf11 C-ter. TAP-7, *K.l. TRP1*), YBP103 (*MATα; ura3-1; leu2-3,112; trp1Δ; his3-11,15; ade2-1; can1-100;* Cpf11 C-ter. TAP-12, *K.l. TRP1*), MO20 (*ura3-1; trp1Δ; ade2-1; leu2-3,112; his3-11,15; pfs2::TRP1* pNOP-PFS2 [ProtA-PFS2-LEU2-CEN]).

The pBS1479 plasmid (Puig et al., 2001) was used to amplify by PCR the C-terminal TAP tagging cassette.

Preparation of yeast cell free extracts

This method is adapted from the liquid nitrogen method of Ansari and Schwer (Ansari and Schwer, 1995). 5 liter cultures were grown in YPD (1% yeast extract, 2% bacto-

peptone, 2% glucose) to an OD_{600} of 2-3. The following procedure was done at 4°C. The cells were harvested by centrifugation at 5000 rpm for 5 min in a Sorvall SLA-3000 rotor and washed once with 200 ml of ice-cold water and once with 200 ml of ice-cold AGK buffer [10 mM Tris-HCl pH 7.9, 200 mM KAc, 10% glycerol and 1 mM dithiothreitol (DTT)]. Then, the cells were resuspended in 20 ml AGK buffer and frozen as drops in liquid nitrogen. The frozen cells were stored at -70°C until they were homogenized. For homogenization, the frozen yeast cells were ground with a pestle in a mortar that was previously cooled with liquid nitrogen to avoid thawing. Grinding was continued until a fine powder was obtained. The frozen powder was transferred to a fresh tube and thawed under cold running water. Immediately after thawing, protease inhibitors were added to the following concentrations: 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin and 0.7 μg/ml pepstatin. The suspension of broken cells was gently mixed for 30 min on ice and was then centrifuged at 10000 rpm for 30 min in a Sorvall SS-34 rotor. The supernatant was removed and centrifuged at 41000 rpm in a Kontron TST 41.14 rotor for 2 h. After removing lipids layer, the pale yellow aqueous phase was carefully saved and solid ammonium sulfate was added to reach 40% saturation. Proteins were precipitated overnight with gentle stirring and the precipitated material was pelleted by centrifugation at 10000 rpm in a Sorvall SS-34 rotor during 30 min. The pellet was finally resuspended in 1 ml of buffer E-50 (20 mM Tris-HCl pH 7.9, 0.2 mM EDTA, 50 mM Kac, 0.02% NP-40, 10% glycerol, 0.5 mM DTT, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin and 0.7 μg/ml pepstatin) and dialyzed three times 2 h against 2 l of buffer E-50.

In vitro cleavage and polyadenylation assays

^{32}P-labelled CYC1 RNA was prepared as run-off transcripts and gel purified as described previously (Minvielle-Sebastia et al., 1994; Preker et al., 1995).

A standard *in vitro* processing reaction was carried out in a 25 μl reaction volume containing up to 5 μl of protein extract. For polyadenylation, the reaction mixture contained 2% polyethylene glycol, 75 mM potassium glutamate, 0.01% NP-40, 2 mM ATP, 1.8 mM magnesium acetate, 20 mM creatine phosphate, 1 mM DTT, 0.2 mg/ml creatine kinase and 0.2 units of RNAguard (Amersham Biosciences). The

volume was brought to 25 μl upon addition of buffer E-50 (20 mM Tris-HCl pH 7.9, 0.2 mM EDTA, 50 mM Kac, 0.02% NP-40, 10% glycerol) when it was necessary. When only the cleavage reaction was assayed, EDTA was used instead of magnesium acetate to prevent poly(A) addition, and degradation of the downstream cleavage fragment (Chen and Moore, 1992). The reactions were incubated 1h at the temperatures indicated in the figure legends. They were then stopped by addition of 75 μl of a stop mix (100 mM Tris-HCl pH7.9, 12.6 mM EDTA, 150 mM NaCl, 1% SDS, 0.2 μg/μl of proteinase K and 0.05 μg/μl glycogen) and incubated for 1h at 42°C. The RNAs were recovered by ethanol precipitation, dried and resuspended in 6 μl of formamide buffer (95% formamide, 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF). Samples were heated at 95°C for 5 min and loaded on a denaturing 6% polyacrylamide gel. The gel was exposed overnight to an X-ray film at -70°C.

Droplet test

Yeast strains were grown overnight on YPD or YPG medium and ten-fold serial dilutions were prepared from an initial OD_{600} of 0.1. 5µl of each dilution was then spotted on YPD or YPG agar-medium and incubated at 30°C and 37°C for two days, or at 15°C for five days.

Western blot analysis

For immunoblots, protein samples were separated on either 10 or 12% SDS polyacrylamide gels (Laemmli, 1970). After transfer of proteins to nitrocellulose membranes, the blots were treated according to standard procedures with antibodies diluted in TBS-tween buffer: anti-ProtA antibody (from Sigma) was diluted 1:4000; anti-CBP antibody (from Open Biosystems) was diluted 1:2000; specific antibodies for 3' end processing factors were diluted 1:2000. Peroxidase-conjugated swine anti-rabbit immunoglobulins (DAKO) were used as secondary antibody and chemiluminescence (ECL kit, Amersham) was used for detection.

Tandem affinity purification (TAP)

The TAP tag procedure involves the fusion of the TAP tag to the protein of interest. The TAP tag consists of two IgG binding domains of *Staphylococcus aureus* ProtA and a calmodulin binding peptide (CBP), which are separated by a tobacco etch virus (TEV) protease cleavage site (Figure 3.5 and 3.7A). The fusion protein and interacting partners are recovered from cell protein extracts by only two affinity purification steps (Figure 3.7B).

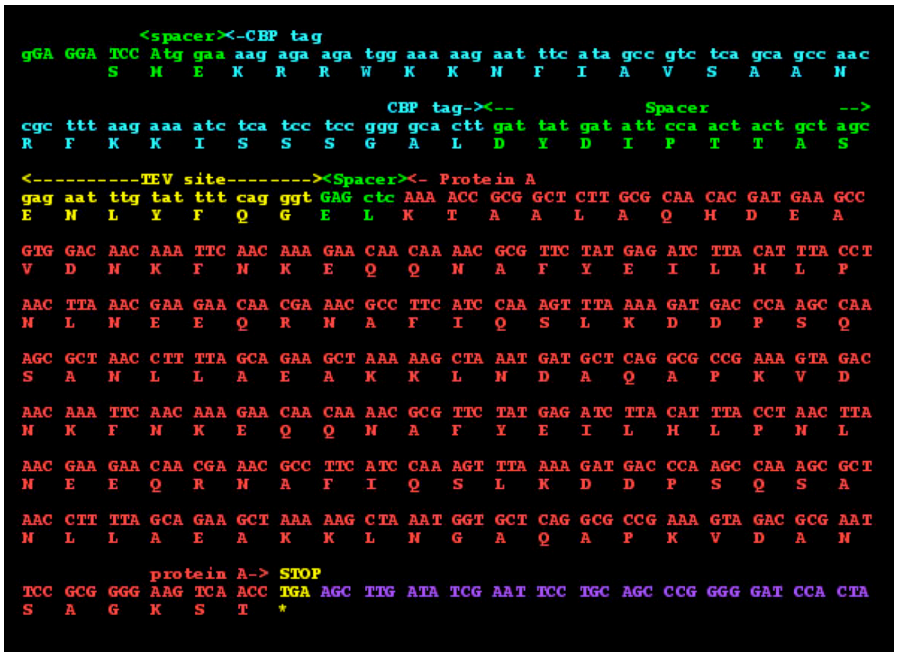

Figure 3.5. TAP tag sequence.
The TAP tag consists of a calmodulin binding peptide (CBP, sequence in blue), a TEV cleavage site (TEV, in yellow) and two ProtA domains (Protein A, in red) in respect to the 5' to 3' orientation. The TEV cleavage site is separated from CBP and ProtA by spacer sequences that maintain the translation frame (in green).

Construction of TAP-tagged Cpf11p

TAP-tagged strains were constructed as described (Puig et al., 2001; Rigaut et al., 1999). Briefly, a pair of gene-specific oligonucleotide primers was synthesized, each of which had been designed to share complementary sequences to the TAP tag cassette at the 3' end and contain 45 base pairs (bp) of homology with the *CPF11* gene to allow in-frame fusion of the TAP tag at the C-terminal coding region of the gene. Primers have the following sequences: Oligo1-TAP (5'-GACATTTTT TTTATTCATGATGTTTTTTTTGTCTATTTTCGTTTTTCCATGGAAAAGAGAA

G-3') and Oligo2-TAP (5'-CGCCACTAGTACAGTTGAAGAAATCACTATAGA AGGGGACGGTCCTACGACTCACTATAGGG-3'). The gene-specific cassette containing the C-terminally positioned TAP tag was then generated by PCR using the plasmid pBS1479 as template (Figure 3.7C), which contains the *TRP1* gene and permits selection of transformed strains in tryptophan-free media. The haploid parent yeast strains BMA41a and BMA41α were transformed with the PCR products according to the lithium acetate method (Gietz et al., 1995; Ito et al., 1983), and strains were selected in SD-TRP medium. By this way, PCR fragments containing the TAP tag sequence can be integrated directly into the genome by taking advantage of the high efficiency of homologous recombination in yeast. Insertion of the cassette was verified by genomic PCR of samples from individual colonies with two pairs of primers designed to produce products of approximately 350 bp with primers ChOl-1 (5'-GGACAATATAGTCAGCTTCA-3') and ChOl-2 (5'-GTCATCTTTTAAACT TTGGA-3'), and 450 bp with primers ChOl-3 (5'-GGTGAA CTGCAGTTGAAATA-3') and ChOl-4 (5'-AAACGTGAATCTCAACCT GC-3'). In each pair of control primers, one oligonucleotide is specific for the insert (the TAP cassette) and the other one is specific for a yeast genomic DNA sequence located near the insertion site of the TAP cassette. The first pair of control primers (ChOl-1 and ChOl-2) and the second one (ChOl-3 and ChOl-4) are used to verify the insertion of the 5' end and of the 3' end of the TAP cassette, respectively (Figure 3.6). If the TAP tag is inserted at the correct locus into the host genome, the use of the first pair of primers will generate a PCR product of 350 bp, while the use of the second pair of primers will generate a PCR product of 450 bp.

Expression of Cpf11-TAP was verified by Western blotting using a rabbit anti-ProtA primary antibody, and a anti-rabbit horseradish peroxidase-coupled secondary antibody.

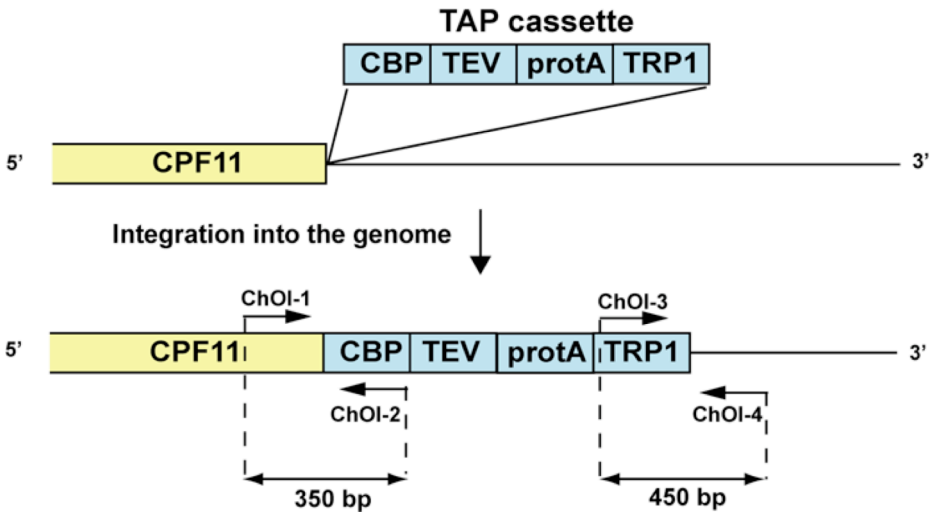

Figure 3.6. Insertion of the TAP tag cassette downstream of *CPF11*.
The correct integration of the TAP tag cassette was verified by PCR using two pairs of specific primers for the TAP tag and the target locus. Oligonucleotide ChOl-1 is homologous to a *CPF11* sequence located 50bp upstream of the stop codon. Oligonucleotide ChOl-2 is specific for the 5' end of the TAP sequence, whereas oligonucleotide ChOl-3 is specific for the 3' end of the TAP cassette. Oligonucleotide ChOl-4 has homology with a genomic region located 270bp downstream of *CPF11*.

Affinity purification of CPF and TAP tag purification of Cpf11-TAP

The CPF complex was affinity purified from an extract of the yeast strain MO20 that expresses a ProtA-tagged version of Pfs2p (Ohnacker et al., 2000). The tag included two IgG-binding domains of the *Staphylococcus aureus* protein A followed by the recognition sequence of the site-specific TEV protease (Senger et al., 1998).

Batch absorption to IgG-sepharose was done to isolate ProtA-Pfs2p or Cpf11-TAP and associated proteins from protein extracts prepared from yeast strains that express these fusions. All incubation steps described below up to the proteolytic cleavage were done at 4°C with gentle mixing. A 1 ml aliquot of IgG-sepharose (Amersham Biosciences), 4 ml of the extract, 10% glycerol and 10 ml of buffer D-120 [20 mM Tris-HCl pH 7.9, 120 mM NaCl, 0.02% NP-40, 0.2 mM EDTA, 1 mM dithiothreitol (DTT)] including 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin and 0.7 μg/ml pepstatin were mixed in a 15 ml Falcon tube and incubated for 2 h on a rotating arm. After centrifugation, the supernatant was removed and the matrix was extensively washed five times for 20 min with 15 ml of D-120 buffer. After washing, the beads were incubated twice for 10 min with 15 ml of buffer TEV (50 mM Tris-HCl pH 7.9, 10% glycerol, 50 mM KCl, 0.5 mM EDTA, 0.02% NP-40, 1 mM DTT, 0.4 μg/ml leupeptin). For elution of the bound material, the beads

were incubated with 1 ml of buffer Tev containing 400 U of recombinant TEV protease (Invitrogen) at 16°C for 90 min. The site-specific proteolytic cleavage released Pfs2p and associated proteins from the protein A tag and thereby from the IgG-sepharose matrix. After centrifugation, the supernatant was collected in a fresh tube and the beads were washed once with 500 μl of Tev buffer.

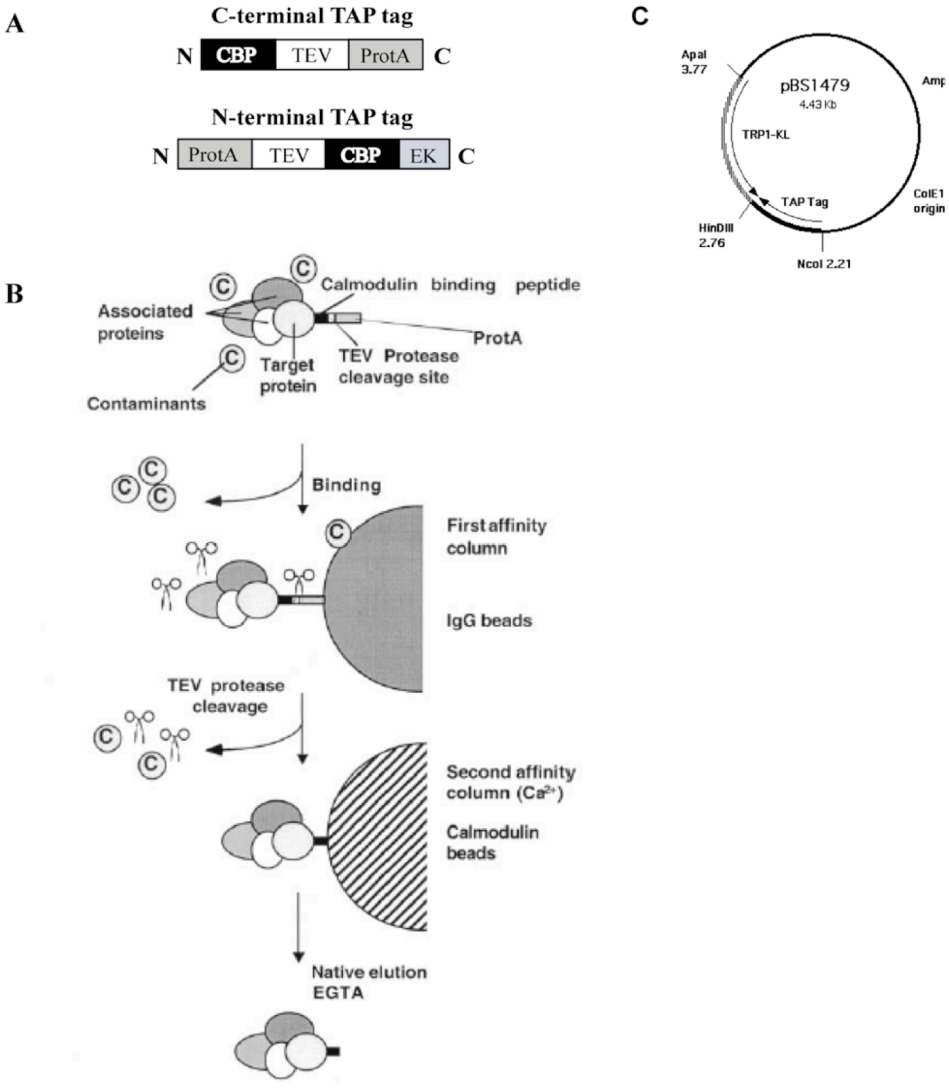

Figure 3.7. TAP tag strategy (from Puig et al., 2001).
(A) Schematic representation of the structure of the C- and N-terminal TAP tags. The domains constituting the TAP tag are indicated. CBP: calmodulin binding protein, TEV: TEV protease cleavage site, ProtA: protein A domain.
(B) Overview of the TAP procedure. The TAP-tagged protein of interest is incubated with IgG matrix, and then, the bound material is eluted by proteolytic cleavage with TEV protease. Eluates are further incubated with calmodulin beads to remove contaminants and the TEV protease. The final elution is carried out in the presence of EGTA.
(C) Restriction map of the pBS1479 plasmid that carries the C-terminal TAP cassette. The cassette contains a tryptophan marker (TRP1) for selection of the transformants.

Chapter 4: Three-dimensional structure of the yeast cleavage and polyadenylation factor (CPF)

4.1 Introduction

Pre-mRNA 3' end processing is a two-step reaction in which each step is catalyzed by a different enzyme. The first step consists of the site-specific endonucleolytic cleavage of the pre-mRNA, catalyzed by the endonuclease Ysh1p, a subunit of CPF (Mandel et al., 2006; Ryan et al., 2004). Two cleavage products are obtained, an upstream cleavage fragment carrying the genetic information, and a downstream cleavage product, which is still attached to the transcribing RNAP II and is then rapidly degraded. In the second step, poly(A) polymerase (Pap1p) adds adenosine residues to the 3' end hydroxyl group of the upstream cleavage product, generating a poly(A) tail of about 70 nucleotides in yeast and 250 nucleotides in mammals (Colgan and Manley, 1997; Keller and Minvielle-Sebastia, 1997; Zhao et al., 1999a).

3' end formation involves a surprisingly large set of protein factors (Colgan and Manley, 1997; Keller and Minvielle-Sebastia, 1997; Zhao et al., 1999a). Four protein factors have been shown to participate in pre-mRNA 3' end processing in yeast: CPF (Dichtl et al., 2002b; Gavin et al., 2002; Nedea et al., 2003; Ohnacker et al., 2000), CF IA (Chen and Moore, 1992), CF IB/Nab4p/Hrp1p (Kessler et al., 1997; Kessler et al., 1996) and Pab1p (Kessler et al., 1997; Kessler et al., 1996; Minvielle-Sebastia et al., 1997). The majority of yeast 3' end processing factors displays high sequence conservation with metazoan counterparts, suggesting a similar mode of action in pre-mRNA 3' end maturation even if the general organization of these factors is different (Keller and Minvielle-Sebastia, 1997; Manley and Takagaki, 1996; Shatkin and Manley, 2000). CPF and CF IA contain homologues of mammalian CPSF and CstF factors, respectively (see Chapter 1, figure 1.5).

Interestingly, both enzymes that catalyze pre-mRNA 3' end processing reactions in yeast are genuine subunits of CPF, the major 3' end processing factor. CPF is a multiprotein complex that consists of a total of 15 polypeptides and is essential for both cleavage and polyadenylation reactions (Ohnacker et al., 2000). It is composed of the following polypeptides: Yhh1p/Cft1p (Preker et al., 1997; Stumpf and Domdey, 1996), Ydh1p/Cft2p (Preker et al., 1997; Zhao et al., 1997),

Ysh1p/Brr5p (Chanfreau et al., 1996; Jenny et al., 1996; Preker et al., 1997), Pta1p (Preker et al., 1997), Pap1p (Lingner et al., 1991b; Preker et al., 1997), Mpe1p (Ohnacker et al., 2000; Preker et al., 1997; Vo et al., 2001), Pfs2p (Ohnacker et al., 2000), Fip1p (Preker et al., 1995), Yth1p (Barabino et al., 1997) and Swd2p, Glc7p, Ssu72p, Pti1p, Ref2p and Syc1p (Dichtl et al., 2002b; Gavin et al., 2002; Nedea et al., 2003). In total, more than 20 proteins participate in this processing step, CPF acting in cooperation with CF IA (which consists of 4 subunits), Nab4p and Pab1p. During the last decade, extensive biochemical studies have been carried out on the proteins involved in 3' end formation in order to characterize their role in this process. Why so many polypeptides are needed for this seemingly simple two-step reaction remains presently unclear. Although the function of numerous components of the 3' end processing machinery has been characterized, further investigations are needed to provide better understanding of the complexity of the protein composition. Several observations have already helped to tackle this issue. One clue is the involvement of some CPF subunits in other cellular processes. For instance, Swd2p also associates with the SET1 multiprotein complex that methylates histone 3 lysine 4 to regulate gene expression (Dichtl et al., 2004; Miller et al., 2001; Nagy et al., 2002; Roguev et al., 2001). Other examples are provided by Glc7p, which is also involved in mRNA export and diverse other cellular processes (Gilbert and Guthrie, 2004; Walsh et al., 2002), and by another phosphatase, Ssu72p, which acts on the phosphorylated serine 5 of the CTD. In addition, the physical and functional coupling between the 3' end processing machinery and other protein assemblies involved in the gene expression pathway (e.g. transcription) also contributes to this complexity. Such a coupling has been suggested to maximize the efficiency and specificity of each step in gene expression (Maniatis and Reed, 2002). Furthermore, several CPF subunits are RNA-binding proteins and contribute to this complexity by contacting the pre-mRNA, such as Yhh1p (Dichtl et al., 2002a), Ydh1p (Dichtl and Keller, 2001), Mpe1p (Vo et al., 2001) and Yth1p (Barabino et al., 1997). Finally, the observation that CPF regulates its own activity by cycles of phosphorylation and dephosphorylation that require the action of the phosphatase Glc7p suggests the presence of an autoregulatory network within CPF (He and Moore, 2005). Hence, pre-mRNA 3' end processing appears to be a more complicated process than what was originally anticipated.

One of the possible approaches to better understand the function of proteins and protein complexes is to reveal their structure and 3D organization. Because CPF

is a nonabundant complex in yeast cells, crystallizing the whole CPF is presently not feasible. So far, only the crystal structures of Ydh1p (Mandel et al., 2006) and Pap1p alone or in complex with ATP have been solved (Bard et al., 2000; Mandel et al., 2006). In mammals, the structures of PAP (Martin et al., 2000) and CPSF-73 (Mandel et al., 2006), the homologue of Ysh1p, were also solved. Another method of looking at large protein complexes and requiring less protein amounts is electron microscopy (EM). EM is one of the fastest and most direct techniques for elucidating the structures of biological macromolecules. 2D electron crystallography is for instance well established to study structures of membrane proteins like receptors, channels, ions pumps and many others. Detailed 3D structures of individual macromolecules can also be obtained by the "single-particle" approach without the need for crystallization, provided that they are sufficiently large (>200 kDa). 3D electron microscopy has contributed a wealth of information on structural organization and on the architecture of large macromolecular assemblies in the last ten years. Single-particle EM is a powerful technique to generate 3D reconstructions of any object imaged in a transmission electron microscope. It gathers many images of the same particle (molecule or molecular complex) in various orientations and uses computational image processing to reconstruct the 3D structure. Since no crystals are needed, the specimen preparation is straightforward, and since the random orientations of the particles are exploited without tilting the specimen, the data collection is also straightforward. Thus, the emphasis of the studies can be focused entirely on the biology of the specimen and its structure-function relationships. Accordingly, the bacterial ribosome structure was first solved at ~20 Å resolution in its empty state (Stark et al., 1995), and later, the structure was elucidated in the pre-translocational and in the post-translocational configurations in association with mRNA and tRNAs (Frank et al., 1995; Stark et al., 1997).

Our aim was to obtain a 3D model of CPF in order to get a more detailed understanding of the 3' end formation mechanism and the role of each of the CPF subunits. We have adapted an efficient method that allows the isolation of highly purified native CPF to analyze the structure of the complex by single-particle EM, a technique that requires fewer amounts of proteins than X-ray crystallography. We have used a yeast strain that expresses a ProtA-tagged version of the Pfs2p subunit of CPF to efficiently purify the 3' end processing factor. Here, we present for the first

time a three-dimensional (3D) model of the CPF complex with a resolution of ~25 Å. This structure was obtained by negatively stained electron microscopy preparations of CPF, angular reconstitution, and random conical tilt (RCT). Our model reveals that CPF has a complex asymmetric architecture in which an outer protein wall surrounds a large inner cavity. We have also used scanning transmission electron microscopy (STEM) to determine the molecular mass of CPF, which we have found to be close to one megadalton. Finally, we discuss the possibility that the inside cavity represents a reaction chamber in which 3' end processing reactions could occur.

4.2 Results

4.2.1 Affinity purification method to prepare CPF complexes for EM analysis

To purify the CPF complex, we have adapted an affinity purification method commonly used in our laboratory, using the yeast strain MO20 that expresses a ProtA-tagged version of Psf2p (Ohnacker et al., 2000). In this strain, 98% of the *PFS2* coding region was replaced by the insertion of a *TRP1* deletion cassette. Since *PFS2* is an essential gene, a yeast expression vector encoding a double ProtA-tagged version of Pfs2p, controlled by the *NOP1* promoter (pNOPATA1L-*PFS2*), was inserted to maintain Pfs2p expression in the cell. The open reading frame of *PFS2* was fused in frame to the ProtA tag that consists of two ProtA domains and a tobacco etch virus (TEV) protease cleavage site fused in frame to the start codon of *PFS2* (Figure 4.1).

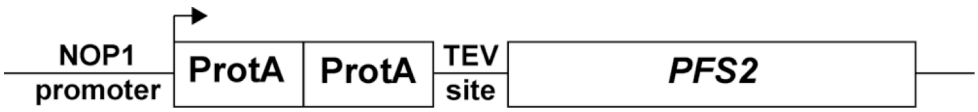

Figure 4.1. Schematic representation of the gene construct coding for ProtA-TEV cleavage site-Pfs2p fusion protein.

The construction of this strain allowed fast purification of functional CPF in only one purification step by incubating protein extract containing ProtA-Pfs2p with

an IgG sepharose matrix and by subsequent release of Pfs2p by TEV cleavage (see below). However, even if EM requires less material than any other 3D structure determining method, significant concentrations of the sample are needed for EM analysis. Obtaining a specimen suitable for EM was not as simple as just scaling up the purification procedure and required not only very concentrated but also highly pure material. Although fully functional, CPF yielded by the single affinity purification step on IgG matrix is only partially pure. Furthermore, as determined by extensive tests we made and as many other protein complexes, the purification of CPF was best achieved in the presence of glycerol and detergents such as NP-40 (data not shown). However, these reagents are incompatible with certain EM techniques since they lead to severe loss of contrast (presence of glycerol for native cryo-EM preparations) or negative stain artefacts (in case of NP-40). Thus, we had to improve and modify the purification method in order to obtain highly concentrated and pure CPF suitable for further EM studies.

We have tested a wide variety of purification conditions, including different methods and composition of buffers, before the establishment of a suitable purification procedure for CPF. For example, we have found that CPF could not be purified with Hepes-containing buffers (Figure 4.2). Purification of ProtA-Pfs2p on IgG sepharose matrix yielded the components of CPF when Tris-containing buffers were used (Figure 4.2A), whereas no proteins were shown to bind the IgG matrix when Hepes was used instead of Tris (Figure 4.2B). The same results were obtained when purification was performed with other ProtA-tagged subunits of CPF, as ProtA-Fip1p (Figure 4.2B).

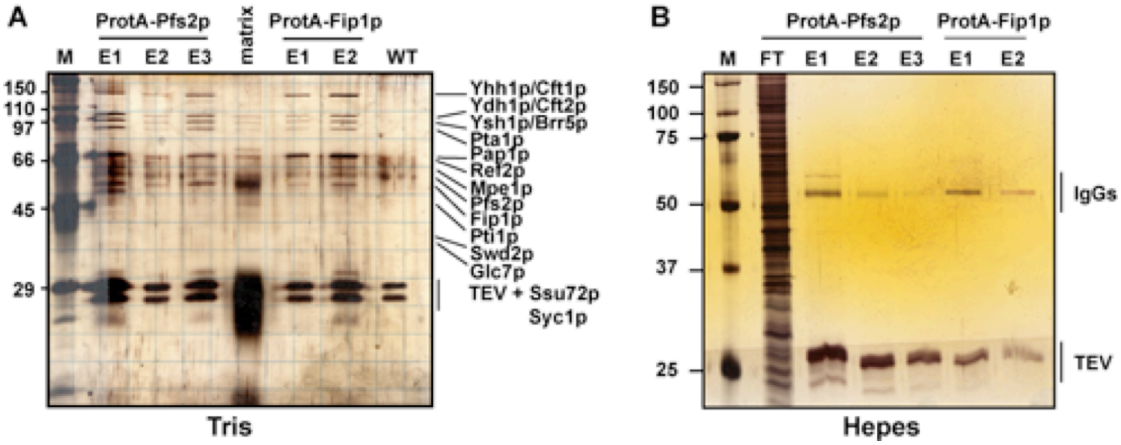

Figure 4.2. Tris-containing buffers are required to purify CPF.
(A) Protein extracts containing ProtA-Pfs2p, ProtA-Fip1p or wild type protein extract (WT) were incubated with IgG sepharose matrix in the presence of Tris-containing buffers. Eluates (E1-E3) were analyzed by SDS PAGE followed by silverstaining. CPF components detected and TEV protease are indicated on the right (see also figure 4.5B for protein composition of CPF) and the molecular weight marker on the left (M, in kDa).
(B) Same experiment as in (A) but carried out with Hepes-containing buffers. FT: flow-through. IgGs: Immunoglobulins G dissociated from the matrix. TEV: TEV protease.

In order to obtain clean CPF, we also tried to purify the complex first on IgG matrix and subsequently by a FLAG affinity chromatography step. To do this, we inserted a FLAG tag between the TEV cleavage site and the Pfs2p sequence of the ProtA-Pfs2p construct. Due to its small size (less than 1 kDa), the FLAG tag is unlikely to disturb protein interactions mediated by Psf2p, which are important for stabilizing the complex (Ohnacker et al., 2000). As described for ProtA-Pfs2p, the new ProtA-FLAG-Pfs2p construct yielded CPF components after purification over an IgG matrix (Figure 4.3A). The IgG eluates obtained were pooled and then subjected to FLAG tag chromatography. As revealed by SDS polyacrylamide gel electrophoresis (SDS PAGE) analysis followed by silverstaining of the FLAG eluates, this procedure did not result in CPF purification (Figure 4.3B). Only few bands that could correspond to some CPF subunits were weakly detected on the gel. Nevertheless, the whole complex was not obtained. It is possible that the FLAG tag was hidden by a structural element of the CPF complex due to its small size, preventing binding to the anti-FLAG affinity gel. A FLAG tagged version of Tad2p, a subunit of the tRNA-specific adenosine 34 deaminase, was used as positive control for the FLAG tag purification (kindly provided by M. Schaub).

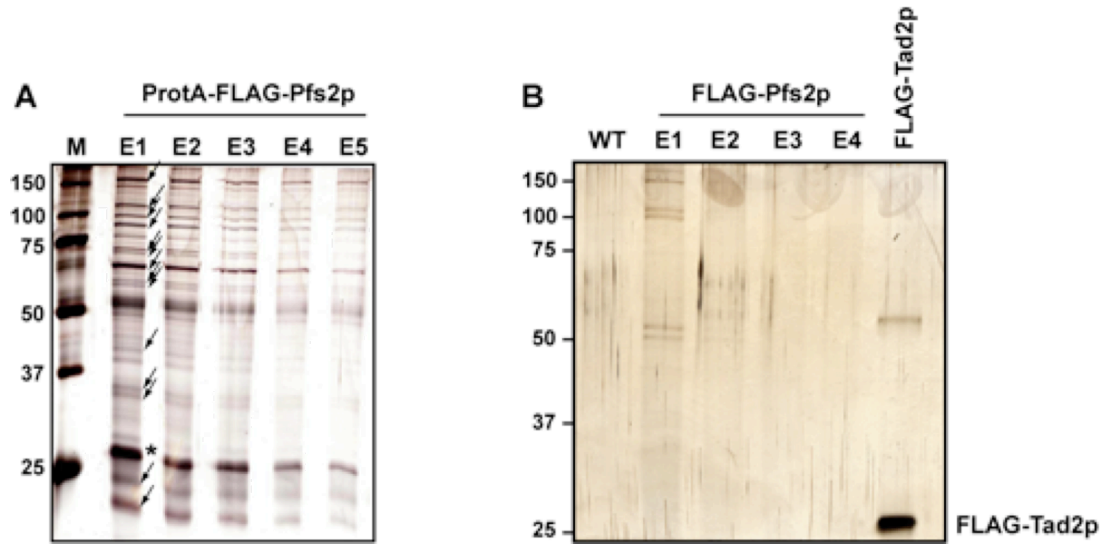

Figure 4.3. FLAG tag purification of ProtA-FLAG-Pfs2p did not yield CPF.
Protein extract containing the ProtA-FLAG-Pfs2p construct was incubated with IgG matrix (A). Proteins eluted were analyzed by SDS PAGE and silverstaining. Arrows indicate subunits of CPF (see figure 4.5B for protein composition of CPF) and the asterisk indicates TEV protease used for the cleavage of the ProtA tag during elution. IgG eluates (E1-E5), wild type protein extract (WT) and an extract containing FLAG-Tad2p were then subjected to FLAG tag chromatography (B). E1-E4: FLAG eluates. The molecular weight marker is indicated on the left (M, in kDa).

After many tests, we have finally established a double affinity purification method that allows the isolation of highly pure CPF providing sufficient amounts for EM.

In the first affinity chromatography step protein extracts were incubated with IgG sepharose matrix, which specifically binds to the ProtA epitope of ProtA-Pfs2p (Figure 4.4). In this step, the matrix retained ProtA-Pfs2p and all associated proteins. After extensive washes, the bound material was eluted from the matrix by site-specific proteolytic cleavage with the TEV protease, which has been shown to allow efficient release under native conditions (Senger et al., 1998). The TEV protease specifically recognizes the seven amino acid sequence ENLYFQG, which is not encoded by the yeast genome, and cleaves between glutamine (Q) and glycine residues (G) (Carrington and Dougherty, 1988). This way Pfs2p and its interacting partners are released from the matrix, hence leading to the purification of CPF.

The efficiency of the IgG purification was monitored by Western blot with specific antibodies against Ysh1p and Pfs2p, two subunits of CPF (Figure 4.5A). Both

polypeptides were immunodetected in the protein cell extracts and in the eluates. A shift in size was observed for the Pfs2p signal in ProtA-Pfs2p extract due to the presence of the ProtA tag (Figure 4.5A lanes 1-2). No signal was observed in the unbound and wash fractions (Figure 4.5A, lanes 3, 4 and 9), indicating that the ProtA-Pfs2p fusion and the associated proteins were efficiently precipitated by the IgG sepharose matrix (lanes 3-4). In contrast, both Ysh1p and Pfs2p CPF subunits (henceforth free of the ProtA tag) were well immunodetected in the eluate fractions, whereas no signal could be observed from the IgG beads after elution, indicating that CPF was fully recovered (lanes 5-9). Taken together, these immunodetection data show that the first affinity purification step for CPF was very efficient. The advantage of this method over conventional purification procedures is that it allows purification of a tagged protein directly from a yeast total protein extract under native conditions.

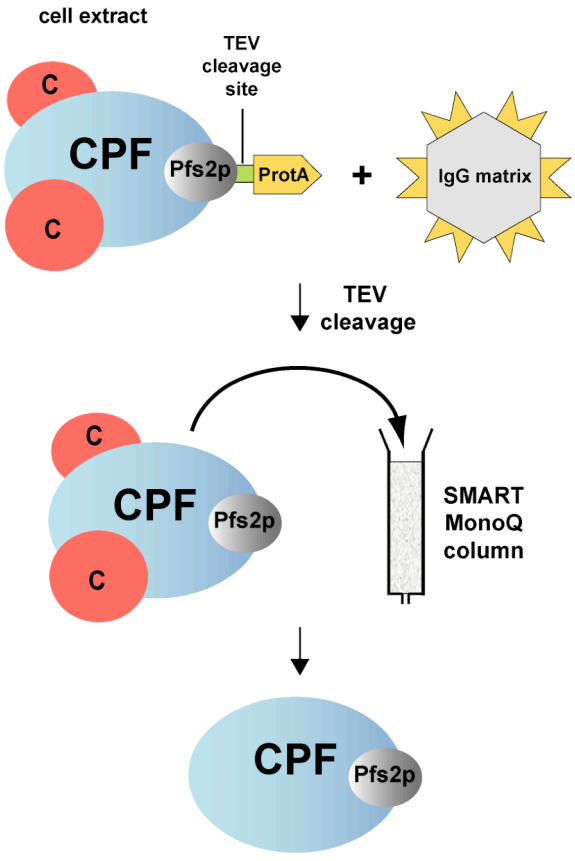

Figure 4.4. Overview of the CPF purification method.
In the first affinity purification step, ProtA-Pfs2p extracts are incubated with IgG sepharose matrix that retains ProtA. The bound material is then eluted by proteolytic cleavage with TEV protease and the eluate is further subjected to a second chromatography step over the anion-exchange MonoQ column. C: protein contaminants bound to CPF.

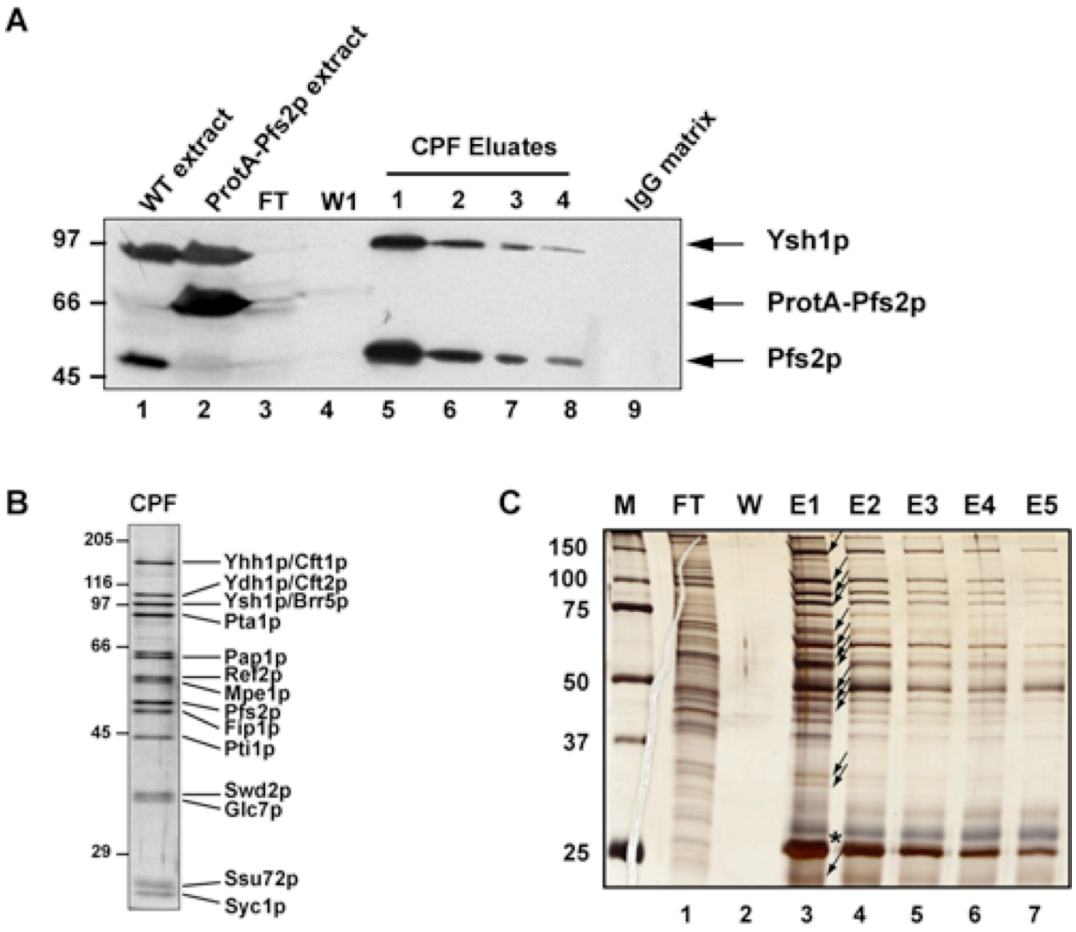

Figure 4.5. Affinity purification of proteins bound to ProtA-Pfs2p yields the CPF complex.
(A) Western blotting analysis of fractions obtained by IgG sepharose purification of extracts from the ProtA-Pfs2p strain. The presence of Ysh1p, Pfs2p and ProtA-Pfs2p was tested in protein extracts of wild type (WT) and ProtA-Pfs2p strains, and at different steps of the IgG purification procedure, as indicated on the top of the panel. (FT: flow-through; W1: first wash; IgG matrix: IgG matrix after elution). Specific antibodies were used to detect Ysh1p and Pfs2p. Proteins immunodetected are indicated on the right of the panel, and the molecular weight marker on the left (in kDa).
(B) Protein composition of pure CPF revealed by protein gel stained with silver (adapted from Dichtl et al., 2002b). The molecular weight marker is indicated on the left (in kDa).
(C) CPF obtained after IgG sepharose purification of ProtA-Pfs2p from protein extracts was analyzed by SDS-PAGE on a 10% gel and stained with silver. 5 µl of the flowthrough (FT), the first wash step (W) and the eluates (E1-E5) fractions were loaded on the gel. Arrows indicate CPF subunits and the asterisk (*) the TEV protease. The molecular weight marker is indicated on the left (M, in kDa).

However, at this step, CPF is not completely pure, as revealed by a silver staining analysis of the IgG eluates after SDS PAGE: figure 4.5C shows a pattern of bands similar to the one previously reported for CPF (Figure 4.5B; Dichtl et al.,

2002b; Gavin et al., 2002; Nedea et al., 2003), but additional bands that correspond to interacting protein contaminants were also present. These contaminants include the TEV protease used for the elution, CF IA (detected by Western blot, data not shown) and other proteins. Indeed, CPF and CF IA intimately interact with one another (Kyburz et al., 2003; Nedea et al., 2003; Ohnacker et al., 2000; Preker et al., 1995; Preker et al., 1997).

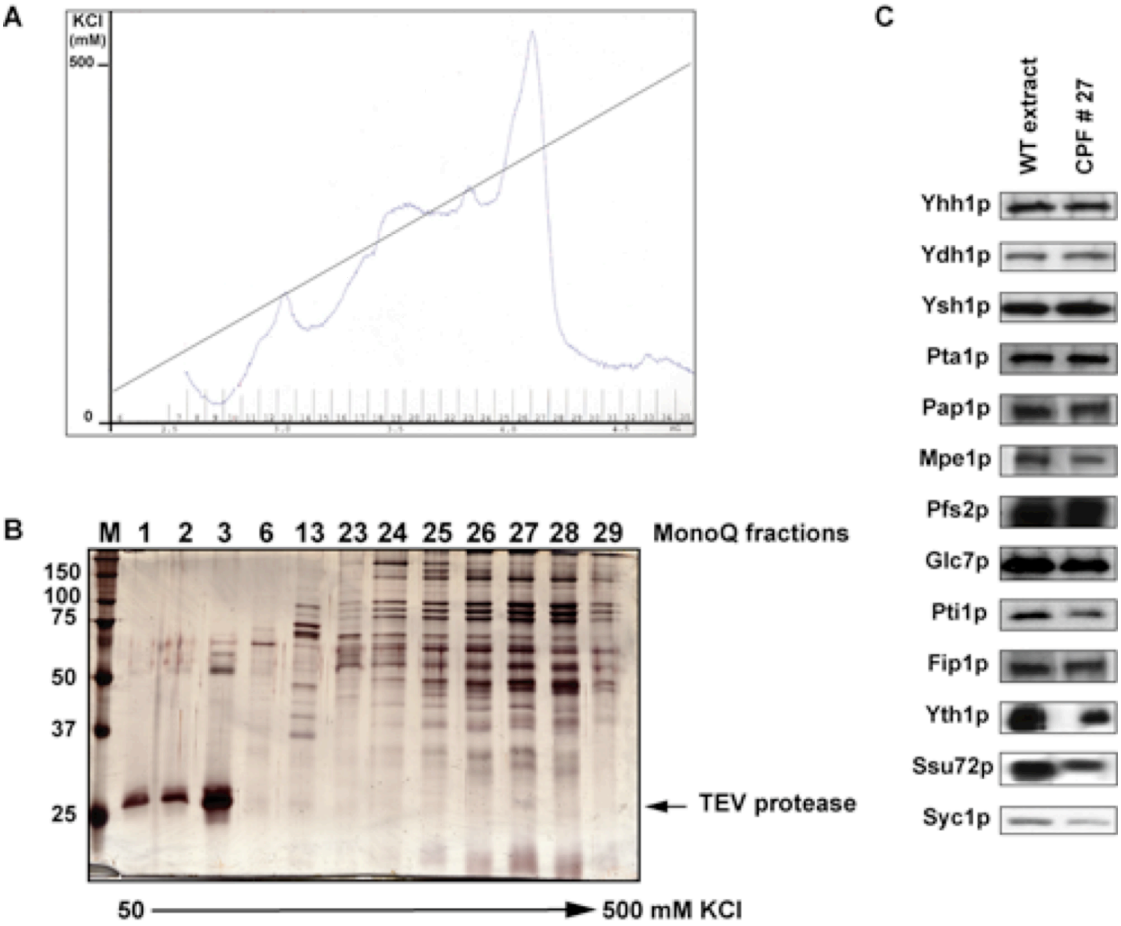

Figure 4.6. Removal of the CPF interacting contaminants by anion-exchange chromatography on MonoQ column.
(A) Elution profile of proteins (blue curve) obtained from a SMART MonoQ column to which CPF obtained after IgG sepharose purification was applied. Proteins were eluted with a linear KCl gradient (black line), as indicated on the left of the panel.
(B) 5 µl of the fractions obtained after elution were separated by SDS-PAGE on a 10% gel and analyzed by silver staining. The number of the fractions, the KCl gradient and the molecular weight marker (M, in kDa) are indicated on top, below and on the left of the panel, respectively. For protein composition of CPF, see also figure 4.5B.
(C) Immunobloting analysis of the polypeptides present in the CPF peak fraction number 27 in comparison with a wild type (WT) protein extract. The subunits of CPF that were detected with specific antibodies are indicated on the left.

In order to separate these contaminants and to concentrate CPF, eluates obtained from this first purification step were pooled and further purified by SMART MonoQ anion exchange chromatography (Figures 4.6 and 4.4). Elution of the bound material was carried out by applying a linear KCl gradient. Different types of gradients were tested, as for instance KCl step gradients to concentrate CPF complexes in only one single fraction. Although using such a step gradient resulted in the concentration of CPF in one fraction, we also observed that protein contaminants were not removed as expected but accumulated in the CPF sample (data not shown). Therefore, we used a linear KCl gradient of 50-500 mM and found that the CPF complex was released at approximately 320 mM (Figure 4.6A and B). The fractions obtained from the MonoQ column were analyzed by SDS PAGE followed by silverstaining, which revealed that all the bands corresponding to the polypeptides of CPF were found to coelute in the peak fractions 27 and 28 (Figure 4.6B). The presence of the known subunits of CPF was verified in the peak fraction 27 by Western blotting analysis using specific antibodies for each subunit, with the exception of Ref2p and Swd2p for which no antibodies were available (Figure 4.6C). In contrast, the TEV protease was separated from CPF and eluted in the early fractions of the KCl gradient. (Figure 4.6B, fractions 1-3). Furthermore, the application of a linear gradient of salt (Figure 4.6A) allowed the progressive removal of other protein contaminants that were eluted at lower concentrations of KCl than for CPF release (Figure 4.6B, fractions 1-23), yielding a highly pure protein complex. Thus, performing the purification in two steps allowed us to use detergent-containing buffers in the first step, and appropriate buffers for EM in the second step, without damaging the sample or lowering the purification yield.

4.2.2 Mass spectrometry analysis of CPF

Before analyzing the purified CPF complexes by EM, we confirmed their protein composition and 3' end processing activities.

Several MS analyses conducted in our laboratory and by other groups in the past have identified the six novel proteins Ref2p, Pti1p, Swd2p, Glc7p, Ssu72p and Syc1p among the previously known CPF subunits Yhh1p/Cft1p, Ydh1/Cft2p, Ysh1p/Brr5p, Pta1p, Pap1p, Mpe1p, Pfs2p, Fip1p and Yth1p (Dichtl et al., 2002a;

Dichtl et al., 2002b; Gavin et al., 2002; Nedea et al., 2003). In our laboratory, Dichtl et al. (2002b) also identified Ref2p, Pti1p, Swd2p, Glc7p, Ssu72p, but in contrast to the other MS studies, they did not identify Syc1p (Gavin et al., 2002; Nedea et al., 2003). Instead, they have found another protein of similar size named Cpf11p.

In order to identify the composition of the purified CPF, we analyzed proteins of the MonoQ peak fractions number 27 and 28 (Figure 4.6B) by peptide fingerprinting with MALDI-MS analysis and nanoelectrospray tandem mass spectrometry (Figure 4.7). We have identified all previously reported subunits of CPF, among which also Syc1p. Moreover, Cpf11p was not detected. This is in agreement with our observations described in this thesis in Chapter 3. Furthermore, no additional proteins were detected.

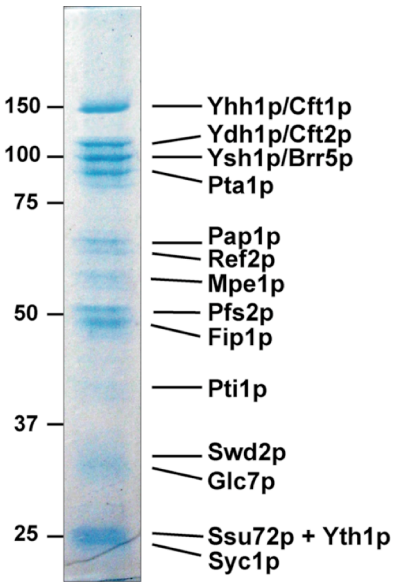

Figure 4.7. Polypeptide composition of CPF analyzed by mass spectrometry. Polypeptides present in the MonoQ peak fraction were separated on a 10% SDS gel, stained with colloidal blue and identified by MALDI-TOF-MS or nanoelectrospray tandem mass spectrometry. Protein names are indicated on the right and the molecular weight marker on the left (in kDa).

4.2.3 Purified CPF is active in 3' end cleavage and polyadenylation *in vitro*

Once conditions for efficient purification and recovery of CPF were established, and before proceeding to EM analysis, the purified complex was tested for its ability to cleave and polyadenylate synthetic radioactively labeled pre-mRNAs. 3' end processing activities can be reconstituted *in vitro* using protein extracts or purified 3' end processing factors with a synthetic RNA substrate that was *in vitro* transcribed and radioactively labeled by incorporation of ^{32}P α-UTP. For the cleavage assay, we

used the *CYC1* precursor that corresponds to the 3' untranslated region (3' UTR) of the iso-1-cytochrome c gene, which contains all the *cis*-acting sequences elements required for proper 3' end formation (Butler and Platt, 1988). For the *in vitro* polyadenylation assay, we used the pre-cleaved *CYC1* (pre-*CYC1)* precursor that corresponds to the upstream cleavage product of *CYC1*. Site-specific *in vitro* cleavage requires the presence of CPF, CF IA and Nab4p, whereas specific polyadenylation is obtained when CPF, CF IA, Nab4p and Pab1p are present. The cleavage assay was performed with the peak and selected side fractions from the MonoQ purification of CPF in the presence of purified CF IA and recombinant Nab4p (Figure 4.8A and B). CF IA was purified on IgG sepharose using protein extracts prepared from a yeast strain expressing a ProtA-tagged version of the CF IA subunit RNA15p (see experimental procedures and figure 4.8C). Among all the fractions tested, only fractions 25-29 showed cleavage activity, as revealed by appearance of both 5' and 3' cleavage products in the presence of purified CF IA and recombinant Nab4p (Figure 4.8A lanes 4-8). Cryptic cleavage products were also observed. Cryptic cleavage occurs due to the absence of Nab4p or when it is present at low concentrations (Minvielle-Sebastia et al., 1998). Thus, in the assay shown in figure 4.8A, Nab4p was not present in sufficient concentrations to suppress cryptic cleavage.

Next, the same fractions were tested for polyadenylation activity. As expected, in the absence of other factors, purified CPF displayed characteristic non-specific polyadenylation activity (Figure 4.8B lanes 5-8) (Ohnacker et al., 2000; Vo et al., 2001). The length of the poly(A) tails generated by Pap1p in the complex was longer than in the wild type due to the absence of Pab1p in the reaction mixture. The same result was obtained for CPF purified only on IgG sepharose matrix. Only when CF IA, Nab4p and Pab1p were added, specific polyadenylation was obtained in fractions 27 and 28 as compared with polyadenylation generated by wild type yeast extract (Figure 4.8B, lanes 12, 13 and 16). Shorter poly(A) tails were observed with the side fractions 26 and 29 (lanes 11 and 14) and fractions 24 and 25 did not display polyadenylation activity at all (lanes 9 and 10). Taken together, these results show that the peak in activity correlates very well with the peak in the protein elution profile presented in figure 4.6B. Hence, our purification procedure allows efficient isolation of fully active CPF that is suitable for further analysis by EM.

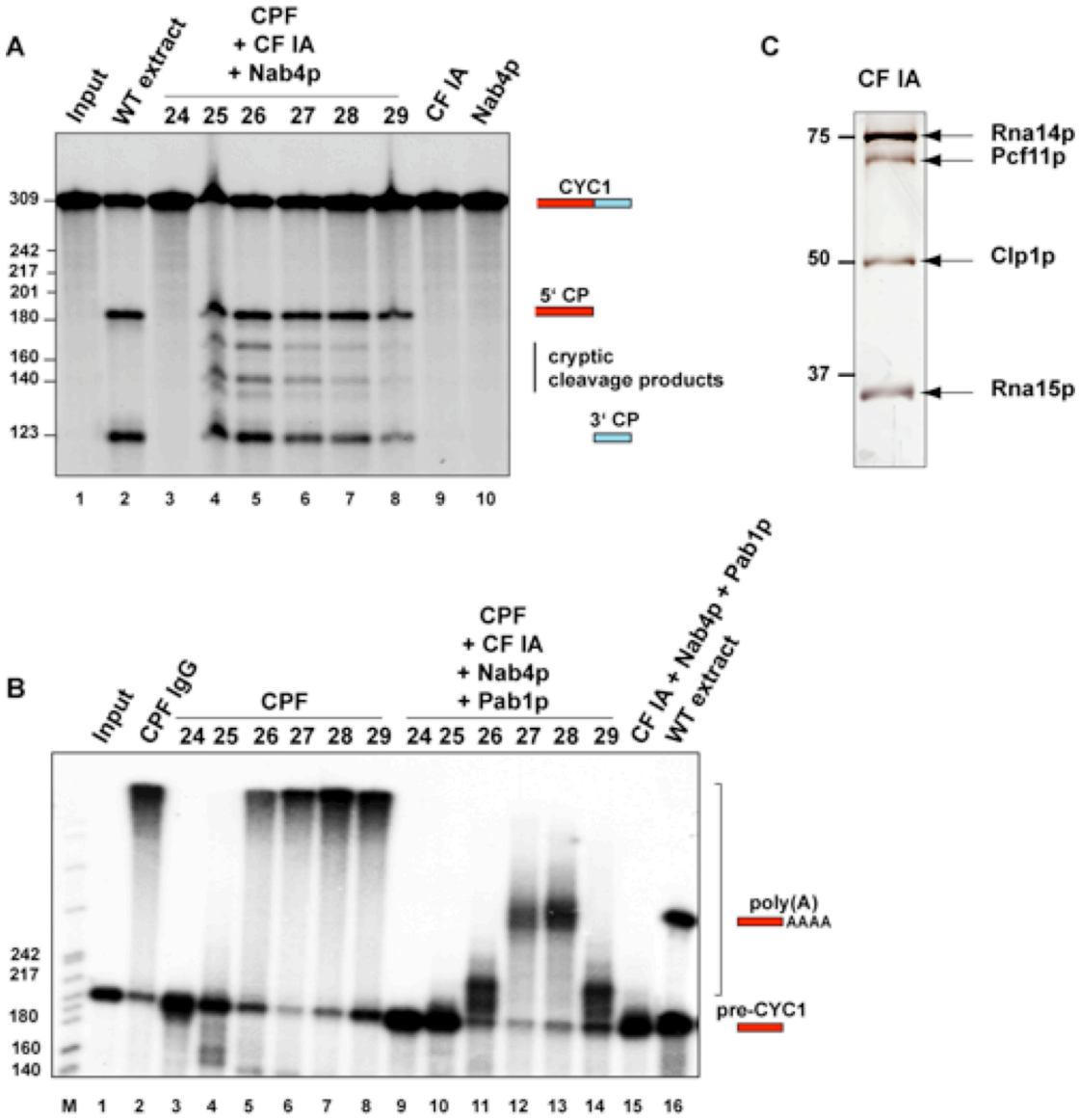

Figure 4.8. Purified CPF reconstitutes specific *in vitro* 3' end processing activities.
(A) *In vitro* cleavage was assayed on the *CYC1* substrate with 30 µg of wild type protein extract (WT, lane 2), or 1 µl of CPF fractions 24-29 purified from ProtA-Pfs2p extracts, in the presence of 1 µl of purified CF IA and 10 ng of recombinant Nab4p, as indicated on the top. The input lane (1) represents the mock-treated reaction. The migration positions of *CYC1*, the 5' and 3' cleavage products (CP) and the cryptic cleavage products are indicated on the right. The position and size of the marker bands are indicated on the left (in number of nt).
(B) *In vitro* polyadenylation was assayed on the *pre-CYC1* substrate with 2 µl of wild type protein extract (WT, lane 16), or 1 µl of CPF purified from ProtA-Pfs2p extracts, in the absence (lanes 3-8) or in the presence of 1 µl of purified CF IA, 10 ng of recombinant Nab4p and 100 ng of Pab1p (lanes 9-14), as indicated on the top. The eluate obtained from the IgG sepharose purification step was assayed alone (CPF IgG, lane 2). The input lane (1) represents the mock-treated reaction. The migration positions of *pre-CYC1* and the polyadenylated products [poly(A)] are indicated on the right. The position and size of the marker bands are indicated on the left (in number of nt).
(C) Native CF IA was purified essentially as described for CPF, but from protein extracts containing a ProtA-tagged version of the subunit Rna15p.

4.2.4 Principle of single-particle electron microscopy

The principle of single-particle EM is described in figure 4.9A. Several thousand molecular images are needed to calculate a 3D structure. Two-dimensional projection images of individual macromolecules are obtained by imaging the sample, which has previously been spread on a grid, in an electron microscope. Several thousands of individual images are extracted from the digitized micrographs and a 3D structure of the macromolecule is calculated by exploiting the random orientation of the particles. The microscope is almost an ideal projection device that provides many projection images of molecules that are randomly oriented on the EM grid. The high amount of noise present in electron micrographs of biological samples requires extensive computer aided image processing. After digitizing the negatives, computational image processing is used to improve the poor signal-to-noise ratio, to determine the angular relationship between the different randomly oriented molecules, and to calculate a 3D reconstruction (Figure 4.9B). To avoid averaging of images that represent different views, multivariate statistical analysis (Frank and van Heel, 1982) and classification is carried out and results in groups of images that are averaged into so called class averages or characteristic views (Harauz et al., 1987). The group members are supposed to be perfectly aligned and should all be representative of the same 2D projection image. The angular orientation of these "class averages" is then determined (Van Heel, 1987) and a 3D reconstruction is finally calculated. In several refinement steps reprojections from the calculated 3D model are computed and used for an iterative improvement of the alignment and projection angle parameters. After several refinement steps a stable 3D structure can finally be obtained.

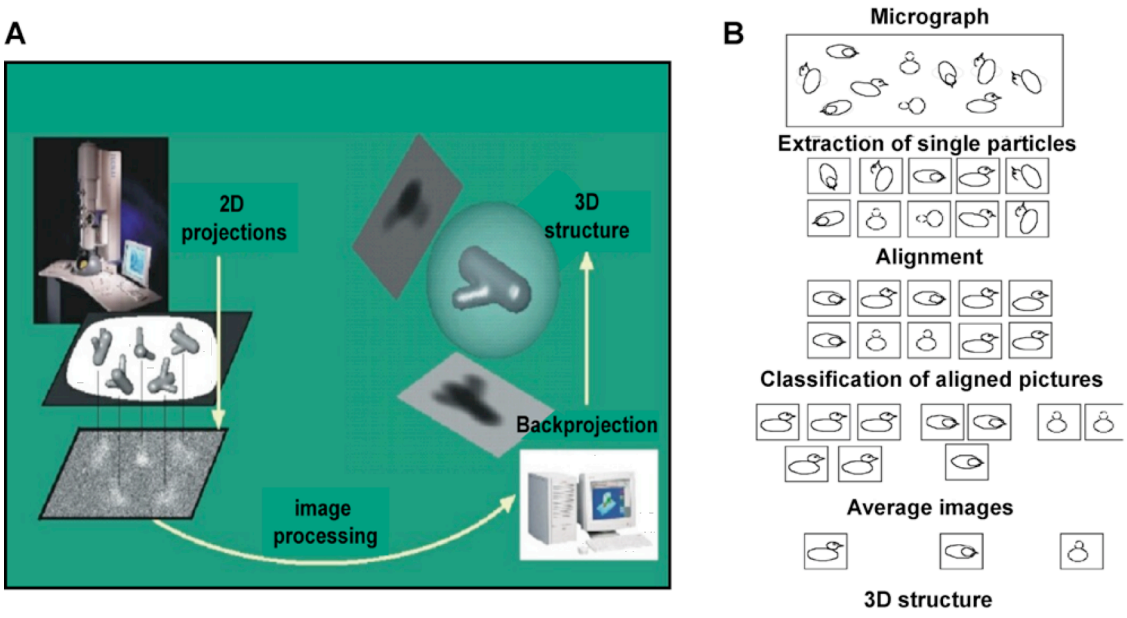

Figure 4.9. Structure determination of single particles by angular reconstitution.
(A) The electron microscope provides 2D projections of individual macromolecules that are randomly oriented on the EM grid. Many images are then extracted from the digitized micrographs and are processed with computer programs allowing an initial 3D model to be calculated. A backprojection step is used to finally refine the 3D structure.
(B) Overview of image processing used to calculate a 3D model. Individual images that correspond to different views of the macromolecular complex (examplified by a duck) lying on the EM grid are extracted from the micrograph, are aligned and classified into groups in which they share similarities and represent the same 2D projection of the complex. Members of each group are then merged to give rise to class averages. The angular orientation of these class averages is determined to finally calculate the 3D reconstruction.

4.2.5 Stabilization of CPF assembly for electron microscopy analysis

4.2.5.1 CPF components dissociate in contact with EM grids

The quality of the purified CPF sample was verified by loading the complex preparation on an EM grid and staining with heavy metals salts, such as uranyl acetate or uranyl formate, to enhance the contrast. We have observed by negative staining studies that CPF is unstable over time. We additionally found that freezing the complex for storage caused damage to it, thus considerably reducing the quality of the specimen for EM purposes. Consequently, the purified material was applied to EM grids immediately after the elution from the MonoQ column. A typical micrograph obtained from CPF negatively stained with uranyl acetate is presented in figure 4.10. The grid shows a homogeneous and monodisperse population of roughly spherical

particles that look similar in size and shape. The CPF complex has a predicted mass of about 944 kDa, which corresponds to the sum of the masses of each individual subunit assuming a 1:1 stoichiometry for all components. As described in the literature, one megadalton (MDa) complexes have a size in the range of 20-25 nm. An example is the anaphase-promoting complex/cyclosome (APC/C), an ubiquitin ligase that has essential functions in mitosis, meiosis, and G1 phase of the cell cycle (Dube et al., 2005). This protein complex of at least 12 subunits has a mass of ~1.4 MDa, and shows particles with a size ranging from 20 nm to 25 nm (Dube et al., 2005; Passmore et al., 2005). Surprisingly, CPF particles in these preparations appeared to be smaller than expected and their size was not more than 10 nm (Figure 4.10). This suggested that the components of CPF disassembled on the grid. Indeed, large protein complexes can have a tendency to fall apart when adsorbed to EM grids, owing to the harsh treatment during staining with heavy metals. However, dissociation of the subunits could also indicate instability of the CPF complex. Therefore, we subsequently tested whether the subunits of CPF are stably associated within the complex.

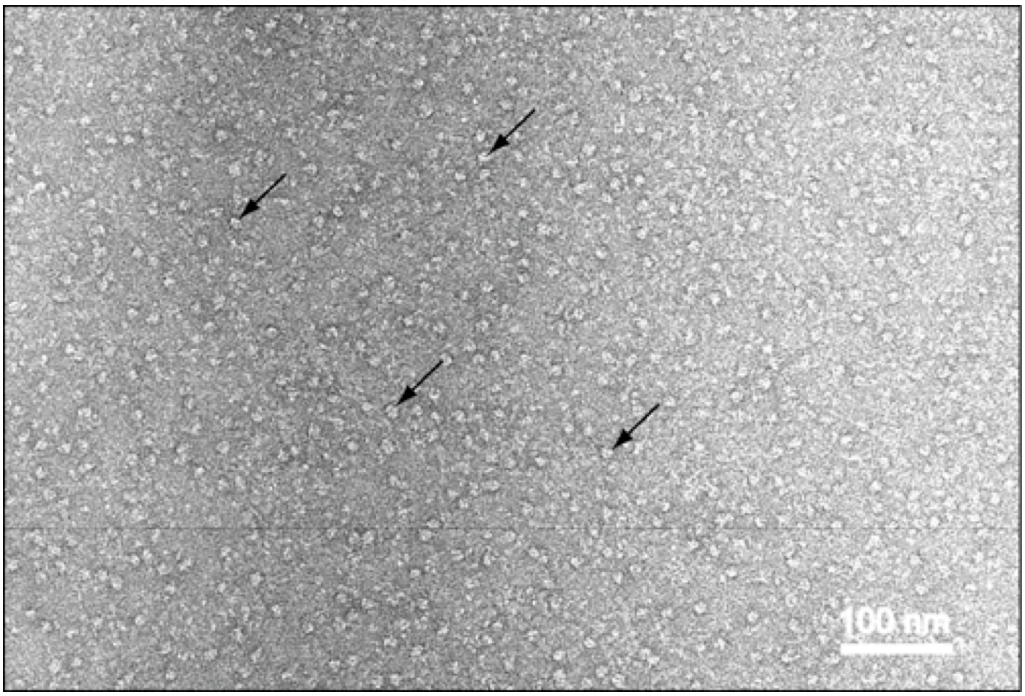

Figure 4.10. EM field of disassembled CPF complexes.
Visualization of CPF by negative stain. The micrograph shows a population of particles with a maximun size of ~10 nm, most likely corresponding to parts of broken complexes. Arrows point to examples of single particles.

4.2.5.2 CPF subunits are stably associated within the complex

We purified CPF under different conditions to test the stability of the interaction network of the components. CPF was purified over the two affinity chromatography steps described above, but was washed with buffers of increasing ionic strength (120, 150, 300 and 400 mM KCl) during the first purification step on IgG sepharose matrix to verify whether some subunits are more loosely associated to the complex than others. Washing with buffers that contained up to 400 mM of KCl did not affect CPF stability since the patterns of bands obtained after SDS PAGE analysis of CPF eluates was not different than the one obtained under standard conditions (120 mM of KCl; Figure 4.11A). Moreover, purification of CPF in the presence of buffers with different pH (from pH 6.8 to pH 7.9) also did not affect CPF subunit composition, as revealed by protein gel analysis in figure 4.11B. These results indicate that the components of CPF were stably associated within the complex. Thus, the dissociation of the CPF subunits observed by EM is not due to the dissociation of ionic bonds between the components but is rather a consequence of the deposit on the EM grid.

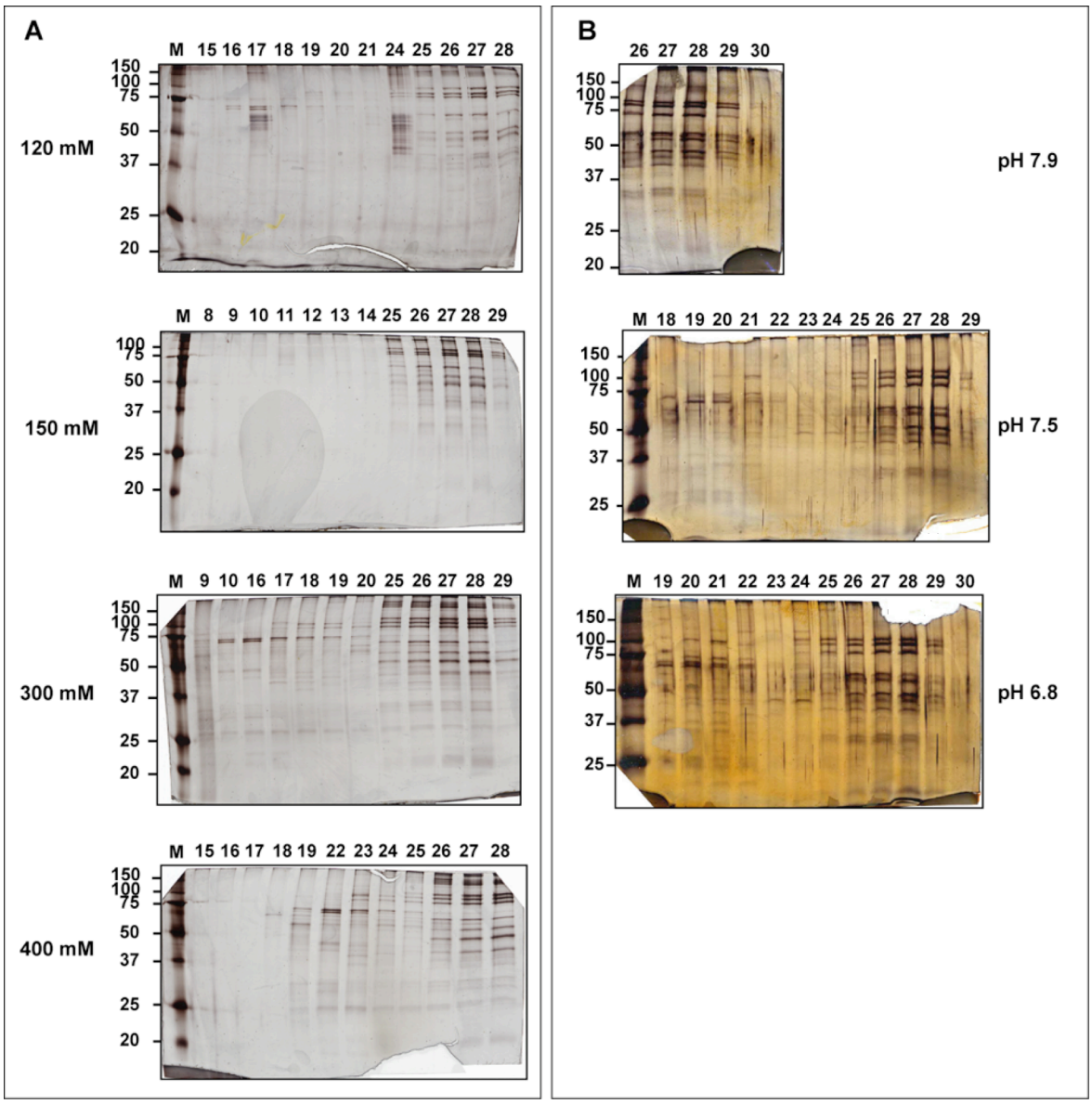

Figure 4.11. CPF components are stably associated within the complex.
(A) Fractions of CPF purified with buffers of increasing ionic strength were analyzed on 10% SDS-PAGE followed by silverstaining. 120 mM of KCl represents the standard conditions for CPF purification. The number of the CPF fractions eluted from the MonoQ column is indicated on the top, the KCl concentration used in washing buffers and the molecular weight markers (M, in kDa) on the left of the panels.
(B) Purifications of CPF in the presence of buffers of different pH values. For each pH tested, subunit composition of CPF was analyzed on a 10% gel stained with silver. A buffer at pH 7.9 represents the standard conditions. Fractions of CPF eluted from the MonoQ column are indicated on the top, pH conditions on the right and the molecular weight markers (M, in kDa) on the left of the panels.

4.2.5.3 Cross-linking with glutaraldehyde stabilizes CPF on EM grids

In order to prevent the separation of the CPF subunits, we first tried to reinforce the stability of the CPF structure by incubating the complex with synthetic RNAs and the other 3' end processing factors CF IA and Nab4p. We incubated different amounts of protein factors with *CYC1* RNAs, but could not find conditions to get stabilized complexes (data not shown). Therefore, we chemically stabilized the interactions within the complex by cross-linking the components with glutaraldehyde. This technique involves the formation of covalent bonds between two interacting proteins by linking primary amines of amino acid residues. We tested different concentrations of glutaraldehyde on purified CPF to determine the optimal concentration. The reagent concentration should be sufficient to allow cross-link of all the CPF components but should not lead to the formation of aggregates. Purified CPF was incubated with different concentrations of glutaraldehyde for 15 min on ice and then loaded on an SDS gel to visualize the effect of glutaraldehyde treatment (Figure 4.12A). The bands that correspond to the subunits of CPF progressively disappeared in the presence of increasing amounts of glutaraldehyde, and a slowly migrating band appeared in the upper part of the gel (compare lanes 2-6 with lane 1). The upper band corresponds to the cross-linked product and has a molecular mass that is too high to be resolved on the gel. Concentrations of 0.01% and 0.05% of glutaraldehyde were not sufficient to cross-link all the CPF components (lane 2 and 3). Traces of individual proteins were still visible with 0.1% of cross-linker but they disappeared with 0.2% (lanes 4 and 5). We therefore used 0.2% of glutaraldehyde to cross-link CPF for EM analysis and, as shown in figure 4.12B, a field of larger particles was now observed on the EM grids compared to the grids loaded with untreated CPF. Although some particles resemble others and seem to be similar in size, the micrograph rather shows a population that is not monodisperse due to the presence of incomplete cross-linked products and aggregates (Figure 4.12C). Such a field of heterogeneous molecules considerably complicated single-particle selection for image processing. We carefully extracted more than 1000 images of appropriate size from the micrographs obtained with cross-linked CPF, and tried to classify them into different groups of particles sharing similar features. But, owing to the high heterogeneity in size and shape of the sample, it was not possible to obtain class averages from the gallery of particles that was made. We therefore tried to further

improve the homogeneity of cross-linked CPF preparation, by adding a third step in the purification procedure of CPF.

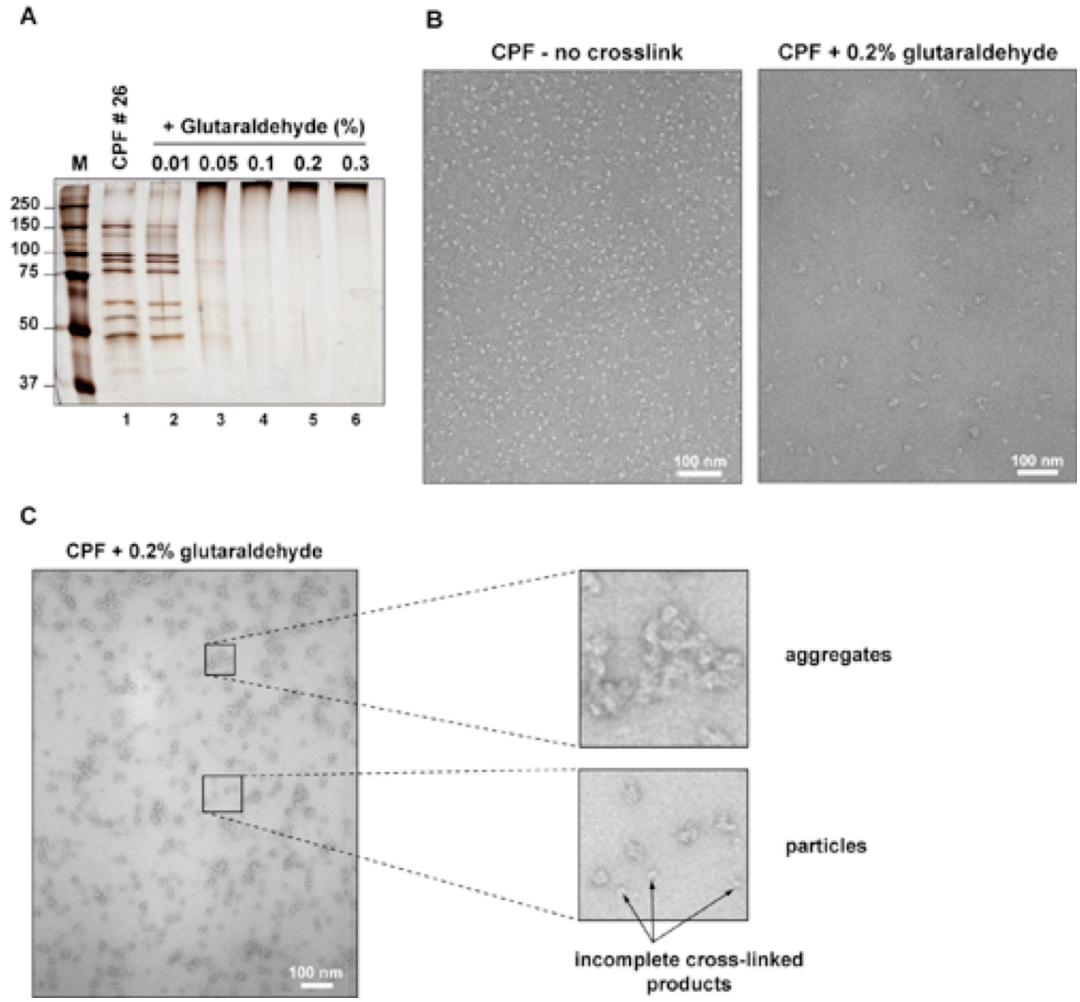

Figure 4.12. Glutaraldehyde cross-link stabilizes the interactions between CPF subunits.
(A) Purified CPF (fraction number 26 eluted from the MonoQ, figure 4.6B) was incubated with different amount of glutaraldehyde as indicated on the top of the panel, and analyzed by SDS-PAGE on a 8% gel stained with silver. The molecular weight marker (M) is indicated on the left of the panel (in kDa).
(B) Comparison of EM raw images of CPF not cross-linked (on the left) and CPF cross-linked with 0.2% of glutaraldehyde (on the right).
(C) Region of the micrograph that contains cross-linked CPF showing the presence of aggregates and incomplete cross-linked products among particles of similar size.

4.2.5.4 Combining density and cross-linker gradients to obtain homogenous CPF

In order to improve the homogeneity of cross-linked CPF populations, we added another step to the double affinity purification procedure of CPF. This final purification step consists of combining a glycerol gradient and a glutaraldehyde gradient in the same tube. This way, in one centrifugation step, CPF is subjected at

the same time to a density and a fixation gradient. This method was called "GraFix" for gradient and fixation (Holger Stark, unpublished data). During centrifugation, the complex travels across the gradients and encounters an increasing amount of fixative reagent. Thus, there is a higher probability of stabilizing individual macromolecules instead of fixing aggregates. Moreover, performing both fixation and centrifugation at the same time helps to dissolve aggregates since complexes are exposed to slightly higher pressure during centrifugation. Thus, there is no need for fixation by glutaraldehyde after the final elution step, which can lead to various problems such as aggregation of protein complexes (Holger Stark, personal communication). CPF eluted from the MonoQ column was applied to a combined linear gradient of 10-30% glycerol and 0-0.15% glutaraldehyde. After centrifugation, the gradient was fractionated from the bottom of the tube into ~180 µl fractions with a peristaltic pump. Aliquots were immediately loaded on EM grids to look for the cleanest and most concentrated fraction, and were negatively stained by the double carbon-sandwich method (see experimental procedures). The electron microscopic images of fractions obtained after GraFix revealed a homogenous, monodisperse population of complexes very similar in size (Figure 4.13). No aggregates and particles of small size were detected, indicating a high efficiency of this method in obtaining a homogenous population of cross-linked CPF.

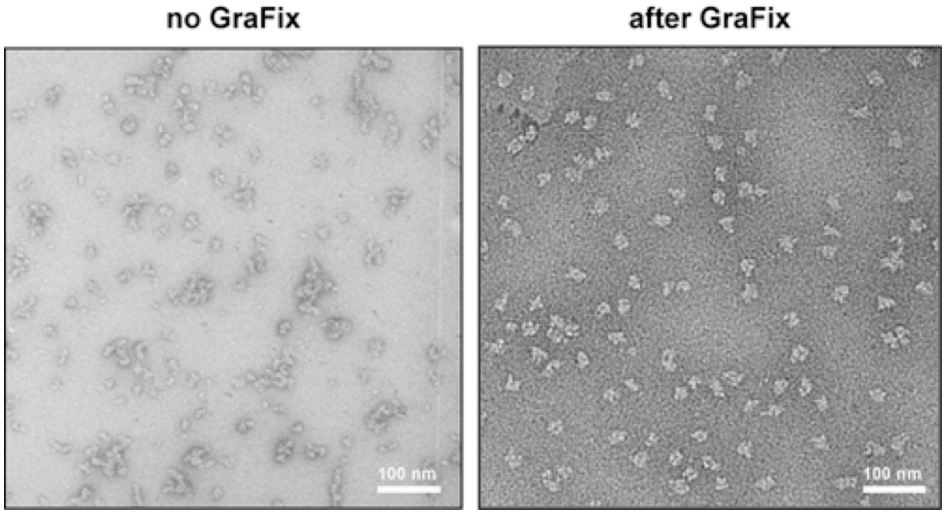

Figure 4.13. GraFix centrifugation improved the homogeneity of cross-linked CPF.
Electron micrographs of negatively stained CPF cross-linked with glutaraldehyde without GraFix centrifugation (left panel), and with GraFix centrifugation showing a more monodisperse and more homogeneous population of particles (right panel, fraction 15 of figure 4.14).

Fractions were also analyzed by gel electrophoresis followed by silverstaining to verify their protein composition. For this purpose, a gradient centrifugation without glutaraldehyde was run in order to enable the resolution of proteins on the gel and further activity tests. The absence of glutaraldehyde had no effect on the migration of the complexes during the centrifugation and very pure CPF was obtained since all the known polypeptides of CPF were detected without any contaminant (Figure 4.14). The presence of the subunits for which the bands were not visible because they stain poorly with silver (Yth1p; Preker et al., 1997) or because the signal was too weak (Pti1p, Glc7p, Ssu72p and Syc1p), was verified by Western blotting (data not shown). Hence, GraFix centrifugation as final purification step greatly improved the quality of purified CPF for EM analysis.

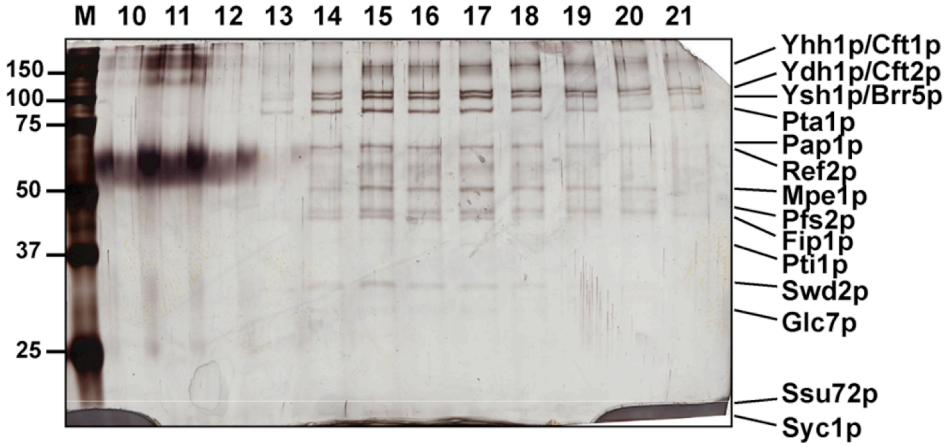

Figure 4.14. GraFix centrifugation yielded highly pure CPF.
Fractions obtained after fractionation of the GraFix gradient were separated by SDS-PAGE on a 10% gel and analyzed by silverstaining. The gel revealed a pattern of bands that is characteristic for the subunit composition of CPF. The numbers of the tested fractions are indicated on the top of the panel, and the molecular weight marker on the left (M, in kDa).

4.2.6 CPF purified after GraFix displays pre-mRNA 3' end processing activities

In order to verify that CPF purified via the GraFix procedure was still functional, we tested the complex for its ability to direct *in vitro* cleavage and polyadenylation of synthetic pre-mRNAs in a reconstituted system with purified 3' end processing factors. For this, we used the control fractions without glutaraldehyde, as cross-linking of CPF results in inactivation of its enzymatic activities (data not shown). We

first monitored cleavage of the *CYC1* pre-mRNA with purified CF IA and recombinant Nab4p (Figure 4.15A). CPF fractions alone were inactive to cleave the pre-mRNA precursor, in contrast to the wild type extract (compare lanes 3-12 with lane 2). However, the cleavage reaction was reconstituted when CPF was combined with purified CF IA and recombinant Nab4p, as shown by the appearance of both 5' and 3' cleavage products (lanes 13-22). Cleavage activity was present in several fractions (lanes 13-22) with a peak activity at fraction 17 for which most of the precursor was processed (lane 17).

In addition to cleavage, polyadenylation on the *pre-CYC1* precursor was also reconstituted in the presence of purified CF IA, recombinant Nab4p and Pab1p (Figure 4.15B). As already described, CPF displays non-regulated polyadenylation activity in the absence of the other 3' end processing factors (lanes 3-12) (Ohnacker et al., 2000; Vo et al., 2001). The addition of the other 3' end processing factors to CPF restored specific polyadenylation as shown by the size of the poly(A) tails generated, comparable to those obtained with wild type extract (compare lanes 16-19 to lane 2). Thus, CPF preparation after the three purification steps is fully competent for pre-mRNA 3' end processing.

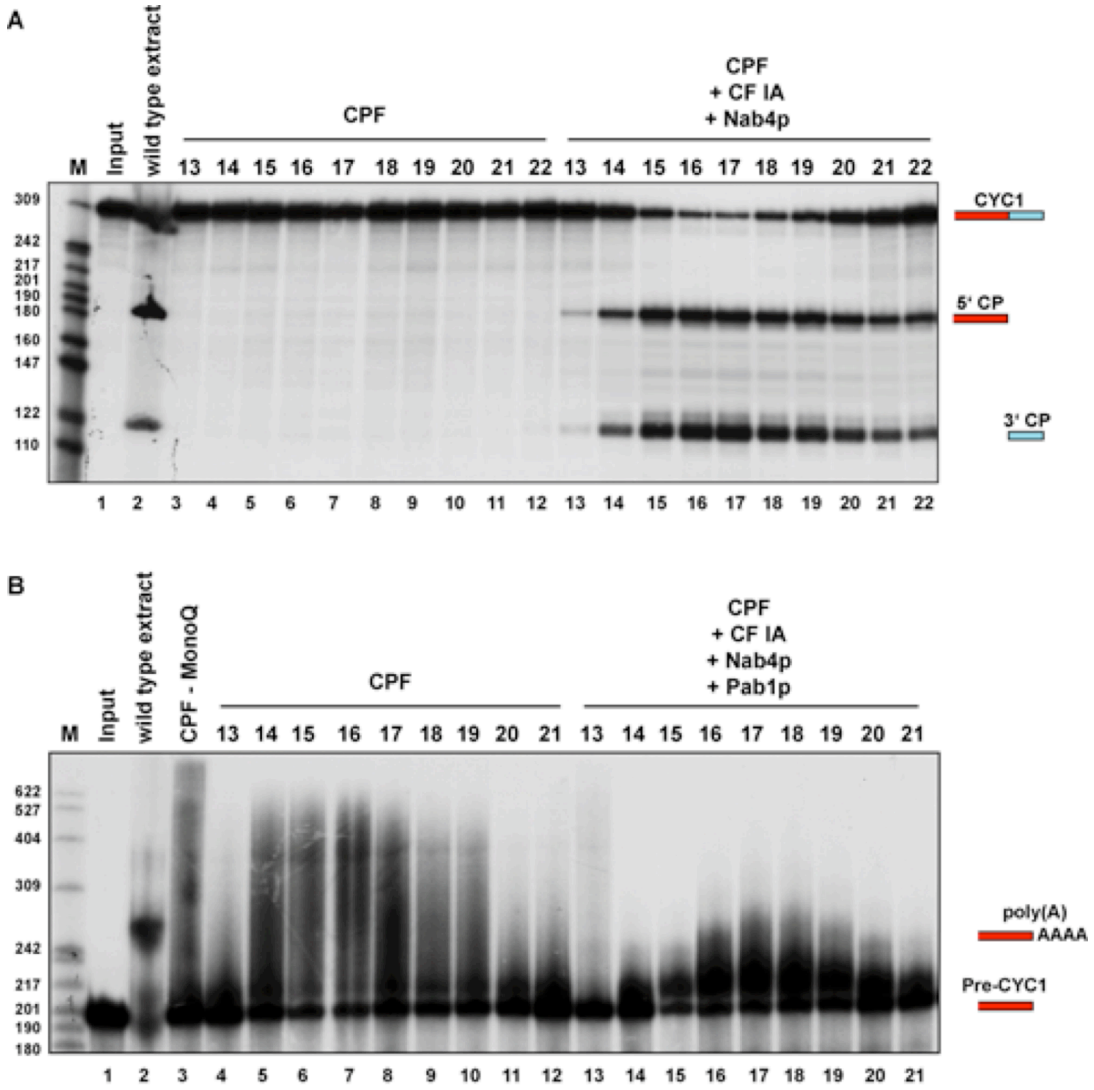

Figure 4.15. CPF purified via the GraFix procedure specifically cleaves and polyadenylates synthetic RNA substrates *in vitro.*

(A) 3 µl of the fractions obtained from GraFix centrifugation were tested for reconstitution of specific *in vitro* cleavage activity on the *CYC1* RNA precursor, in the presence (lanes 13-22) or in the absence (lanes 3-12) of 1 µl of purified CF IA and 80 ng of recombinant Nab4p. 3 µl of wild type protein extract (30 µg of proteins) were tested in parallel as positive control (lane 2). The input represents the mock-treated sample. The positions of the RNA substrate (*CYC1*) and of the 5' and 3' cleavage products (5' and 3' CP) are indicated on the right. The molecular weight marker is indicated on the left (M, in nt).

(B) 3 µl of the fractions obtained from GraFix centrifugation were also tested for reconstitution of specific *in vitro* polyadenylation activity on the *pre-CYC1* RNA precursor, in the presence (lanes 13-21) or in the absence (lanes 4-12) of 1 µl of purified CF IA, 80 ng of recombinant Nab4p and 100 ng of Pab1p. 3 µl of wild type protein extract were tested in parallel as positive control (lane 2). The input represents the mock-treated sample. The positions of the RNA substrate (Pre-*CYC1*) and of the polyadenylated product [poly(A)] are indicated on the right. The molecular weight marker is indicated on the left (M, in nt).

4.2.7 Molecular mass measurement of CPF by scanning transmission electron microscopy (STEM)

In scanning transmission electron microscope (STEM**)** the electron beam is focused to a spot a few Angstroms in diameter and raster scanned over the sample (Muller and Engel, 2001). The electrons pass through the sample and are collected by a series of detectors. Those scattered without loss of energy, elastic scattering, diverge at large angles and are collected by an annular dark-field detector capable of single electron counting. Each pixel of the resulting digital image is, therefore, a quantitative measurement from which in a calibrated image the mass of the protein particle imaged can be calculated. STEM provides "quantitative microscopy", since the digital STEM image not only provides just a picture, but also data that are used in mass analysis. This ability to provide simultaneously both images and quantitative data is of great value for the investigation of protein structures. One of its most successful applications is the mass determination of individual molecules or assemblies of molecules from a range of 100 kDa to several MDa. The importance of the STEM for mass determination resides in its ability to analyze biological structures regardless of their shape and at the same time to provide an image from which the quality and homogeneity of the specimen may be independently assessed. Today, only two STEMs are routinely used in biological research, one at the Maurice E. Müller Institute at the Biozentrum of Basel, Switzerland and the other one at the Brookhaven National Laboratory, Upton, USA.

Experiments	Conditions	Grid treatment	Final scaling	Mass	(n_{peak})	(total n)
1	no fixation	freeze-dried	preliminary	453±158	(641)	(754)
				606±139	(68)	(501)
				516±98	(20)	(78)

Table 4.1. Summary of the STEM mass analyses of CPF.
The table presents the conditions of each experiment (no fixation: CPF not cross-linked; glut., in test tube: glutaraldehyde fixation of CPF directly after MonoQ elution; glut., gradient: glutaraldehyde fixation of CPF by GraFix), the grid treatment, the masses obtained (in kDa; the main peak is in bold when several peaks are yielded), the number of particles corresponding to each single peak (n_{peak}) and the total number of particles used in every experiment (total n).

To obtain accurate molecular mass values for CPF, we analyzed by STEM preparations of glycerol-free, freeze-dried and air-dried unstained CPF isolated by GraFix, and freeze-dried CPF cross-linked with glutaraldehyde directly after the elution from the MonoQ column (Figure 4.16). Table 4.1 shows a summary of seven independent STEM mass analyses of CPF. The non cross-linked CPF had a mass of 453 kDa (Table 4.1 experiment 1), a value that was lower than the expected mass (~940 kDa by adding up the sizes of all subunits). We have observed that uncross-linked CPF is not stable on EM grids (Figure 4.10), thus this mass likely corresponds to stable CPF sub complexes. Different analysis of CPF cross-linked directly after elution from the MonoQ column gave concordant results with a main peak around 973 kDa as the average value of the main peaks obtained in experiments 2, 4 and 5. Experiment 3 was not taken into account due to the large difference obtained in comparison to the other experiments. A typical dark field image of the representative experiment 5 is shown in figure 4.16A together with the mass histogram that

represents the final evaluation. The digital image shows a population of particles that is not homogenous in size and in shape, due to the presence of aggregates and non-complete cross-linked products. Particles of similar size were however carefully selected for mass measurement. The histogram shows a major peak at 976 ± 206 kDa among minor peaks (around 500, 750 and 1400 kDa). The presence of multiple peaks and the large standard deviation of the measurement reflect protein heterogeneity observed in the sample. With cleaner CPF obtained after gradient centrifugation, the population of particles observed was now homogenous and monodisperse, although more diluted due to the gradient fractionation (Figure 4.16B and C). No small particles or aggregates that disturb the measurement were present in the samples and both freeze-dried and air-dried unstained CPF gave similar results. A total of 360 particles from the air-dried grids yielded a single peak at 956 ± 126 kDa (Table 4.1 and figure 4.16B). Additionally, 373 particles from the freeze-dried grids also yielded a unique peak centered at 1006 ± 119 kDa (Table 4.1 and figure 4.16C). The measured mass for the particles on the air-dried grids is slightly lower, but the difference is within the standard deviation of the measurements and in the range of the variations observed in the earlier experiments. Nevertheless, although unlikely, the loss of weakly cross-linked components upon air-drying cannot be entirely ruled out. Interestingly, the agreement between these data and the previous experiments is very good, since merging the results of experiments 6 and 7 gives a single peak at 982 ± 126 kDa (Figure 4.16D), whereas merging the results of experiments 2, 4 and 5 gives a major peak at 976 kDa (see above). These data indicate that the functional CPF has a size close to one megadalton, consistent with the predicted mass obtained by the sum of the masses of the individual subunit. This suggests a one to one relative stoichiometry of subunits within the complex.

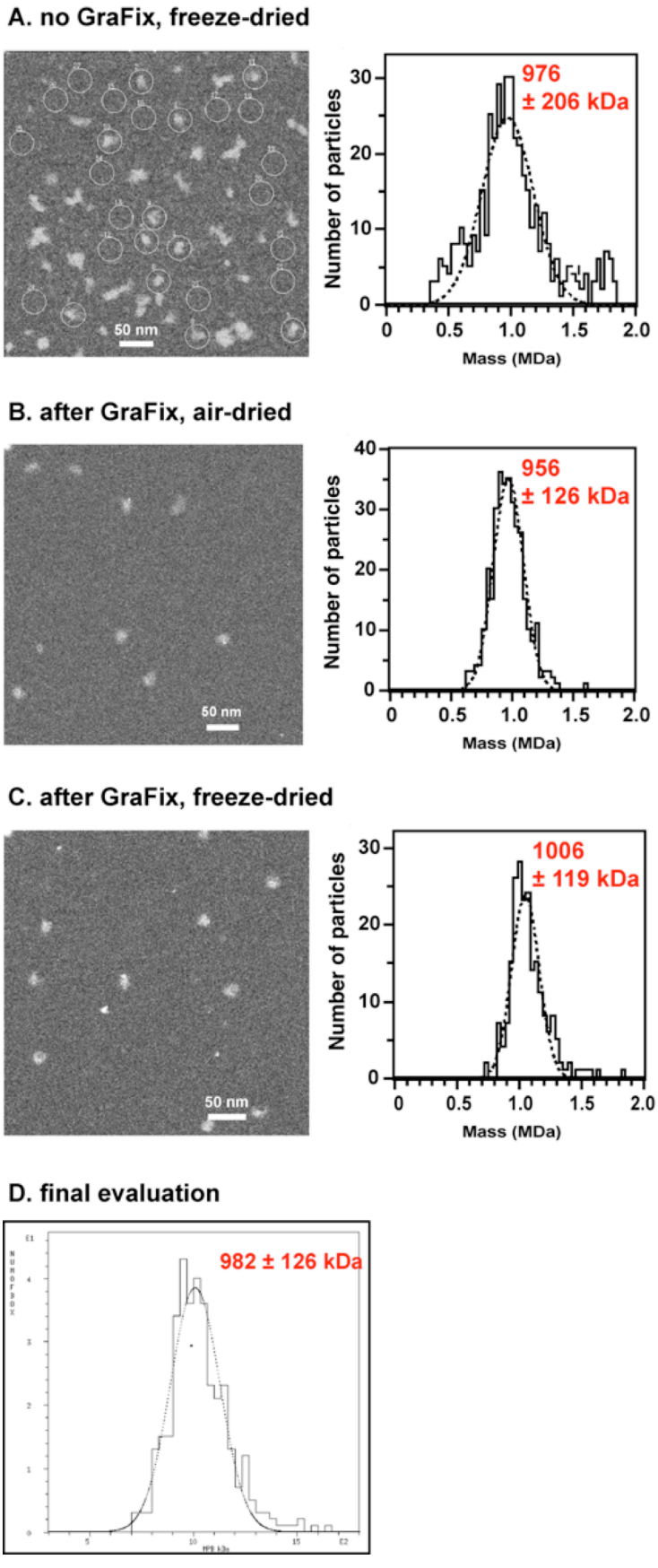

Figure 4.16. STEM mass measurement of cross-linked CPF complexes.
(A) Image of unstained freeze-dried CPF not purified by GraFix (left panel) and the corresponding histogram showing the mass distribution obtained from 319 CPF selected particles (right panel). The Gaussian curve indicates one main peak at 976 ± 206 kDa.
(B) Image of unstained air-dried CPF purified by GraFix (left panel) and the corresponding histogram showing the mass distribution obtained from 360 CPF selected particles (right panel). The Gaussian curve indicates one unique peak at 956 ± 126 kDa.
(C) Image of unstained freeze-dried CPF purified by GraFix (left panel) and the corresponding histogram showing the mass distribution obtained from 373 CPF selected particles (right panel). The Gaussian curve indicates one unique peak at 1006 ± 119 kDa.
(D) Histogram of the final mass evaluation of CPF obtained by merging the results of (B) and (C). After correction for beam-induced mass-loss, a unique peak with a mass of 982 ± 126 kDa was obtained.

4.2.8 Single-particle electron microscopy analysis of negatively stained CPF

4.2.8.1 Three-dimensional structure of CPF determined by angular reconstitution and random conical tilt

To obtain insights into the 3D structure of CPF, we have analyzed CPF fractions purified by the previously described three-step method by negative stain EM. EM grids were prepared by the double carbon foil sandwich technique (Golas et al., 2003) and were stained with uranyl formate, revealing particles with a maximum diameter of approximately 20 nm (Figure 4.17A). The particles appeared to be similar in size, confirming that intact CPF complexes were isolated, whereas the shape of the particles was rather heterogeneous, suggesting an intricate overall structure of CPF.

To determine an initial 3D structure of CPF, we used angular reconstitution (Van Heel, 1987) and RCT (Radermacher et al., 1987). In EM preparations, macromolecules can adopt many stable positions on the grid depending on their shape. The basic goal for solving the 3D reconstruction is to obtain many views of the object in different orientations and then to determine the relative positions and angular orientations of the different projections of the object. Angular reconstitution is a zero-tilt reconstruction method that exploits the random orientations of the molecules on the EM grids (Van Heel, 1987). This approach is based on finding one common line projection between 2D projection images of the same 3D structure (central section theorem). In addition, noise reduction by MSA, classification and averaging of the raw EM projections, which is used to obtain characteristic views of a molecule, is an integral part of this approach. Indeed, raw electron microscopic images of individual particles are usually very noisy and sometimes molecules are hardly visible. Therefore, averaging procedures are required to improve the signal-to-noise ratio. First, 2D projection images of randomly oriented individual particles are aligned to provide an inventory of the different views present on the micrographs. Images that resemble each other are aligned together with the help of correlation procedures. Multivariate statistical classification procedures are then used to identify similar images and to sort them into groups in which all the members share similarities and are representative of the same 2D projection image. Then, all the images of a same group are averaged together, thereby increasing the signal-to-noise

ratio. Finally, the relative angular orientations of the class averages obtained are calculated and from these data a 3D reconstruction can be determined.

In the random conical tilt method, the different views are obtained by recording tilted image pairs of the specimen, providing the relative angle information needed to combine the views in three dimensions. The specimen is therefore imaged twice, first at a high tilt angle and then untilted. Tilting methods require more experimental effort, but provide the handedness of the structure. In contrast, angular reconstitution methods are simpler experimentally, but require more computational work. Finally, once a first 3D map is available from either approach, the structure is refined by projection matching, which means that reprojections computed from the 3D structure are used iteratively at an increasingly smaller angular grid to improve the alignment and orientation parameter.

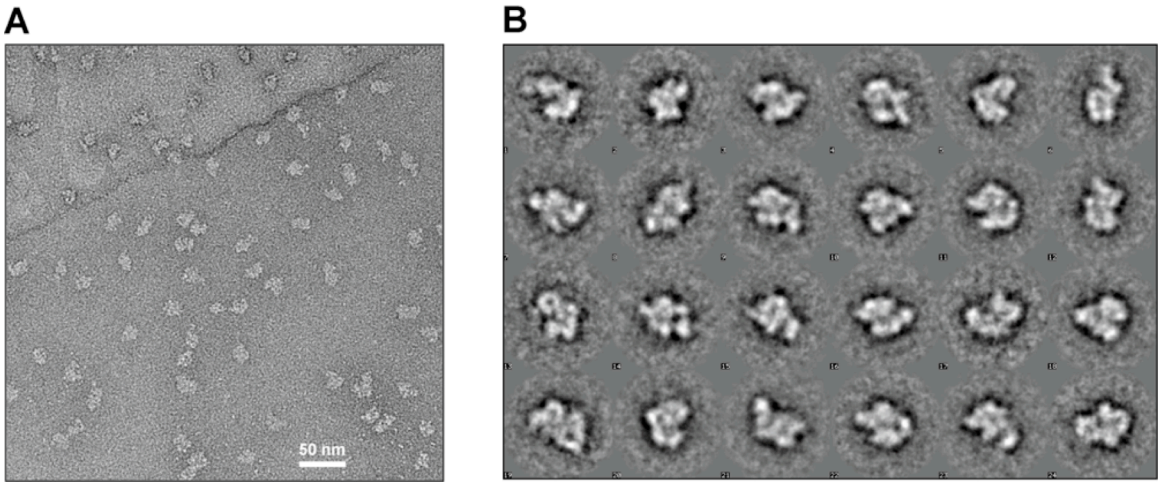

Figure 4.17. Electron microscopy of CPF.
(A) Electron micrograph of negatively stained CPF. Particles were sandwiched between two carbon films in 2% uranyl formate. The EM field shows a monodisperse population of particles with a size of ~20 nm.
(B) Gallery of class averages of CPF obtained by alignment, multivariate statistical analysis and classification of single particles.

About 6000 molecular images of randomly oriented CPF particles from digitized micrographs were collected and furthermore subjected to image processing procedures described above to generate an initial 3D model. A total of 400 class averages, each consisting of 10 to 20 class members, were used to compute the initial 3D structure of the CPF complex. A gallery displaying some of these class averages of CPF is presented in figure 4.17B. These images show some interesting

characteristic views, like the slightly elongated shape of CPF, and also the presence of negative stain in the central part of the structure, suggesting the presence of at least one cavity or deep groove in the complex. Figure 4.18 displays another gallery of additional ten class averages (upper panel) that are depicted together with their corresponding reprojections (lower panel), showing excellent agreement at all projection angles. Moreover, consistent with the gallery presented in figure 4.17B, the same fine structural details were observed. The class average 1 seems to show a depression from the outside to the center of the structure, which could correspond to a large groove at the surface of the structure. Other prominent structural features are revealed by the presence of negative stains in the central part of the structure, as shown by projection views 2, 4 and 5 of figure 4.18: the negative stains in 2 and 4 are quite large and may indicate the presence of a central cavity in the complex. In some projections more than one negative stain dots are observed on one face of the structure (class 5). This is consistent with the potential presence of a central cavity, but it could also suggest the presence of several smaller cavities or grooves in which metal ions accumulated.

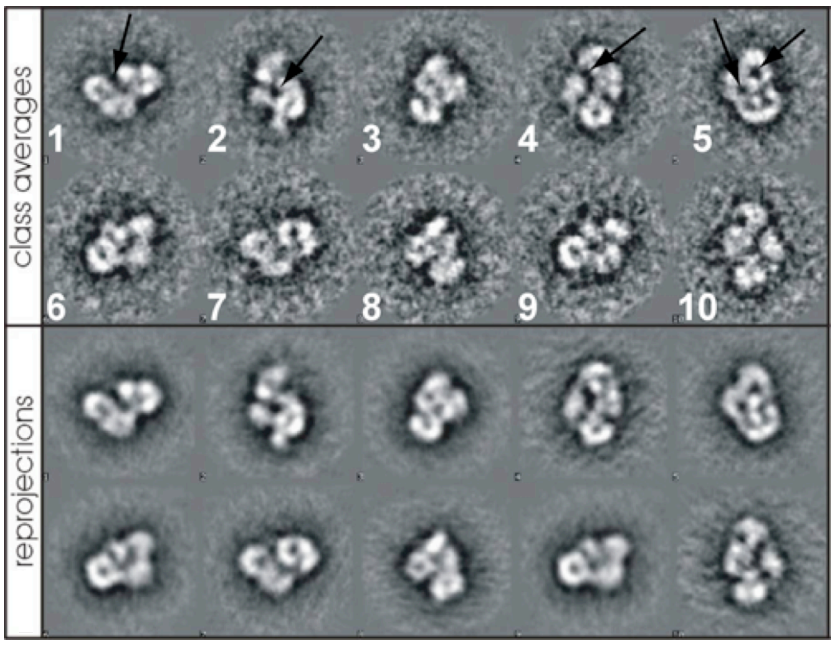

Figure 4.18. Gallery of class averages of CPF and their corresponding reprojections.
The class averages (two rows on top) determined from the 2D projections images used to calculate the 3D structure of CPF and the corresponding reprojections (two rows below) show very good agreement, demonstrating the validity of the 3D structure. Negative stain spots in the center of the structure are indicated by arrows.

4.2.8.2 General architecture of CPF

The method described above allowed us to determine the 3D structure of CPF to a resolution close to 25 Å (Figure 4.19). As the shape of CPF is rather asymmetric, complexes can adopt many possible orientations on the EM grid, consistent with the relative heterogeneity in shape observed on the micrographs. The stereo view shown in figure 4.19 is referred to the front view and describes the most prominent structural features of the complex. The inner contour (white and opaque) well represents the inner surface of the molecules. A gallery of different CPF views that explore the structure is depicted in figure 4.20. The general architecture of CPF revealed a quite elongated, globular appearance with an outer protein wall that surrounds a strikingly large central cavity, and shows several openings of different sizes (Figures 4.19 and 4.20). Thus, the overall assembly appears quite fragile. Interestingly, the presence of the central cavity is entirely consistent with the interpretation of the class averages that have shown negative stain in the central region of the molecule (Figures 4.17B and 4.18).

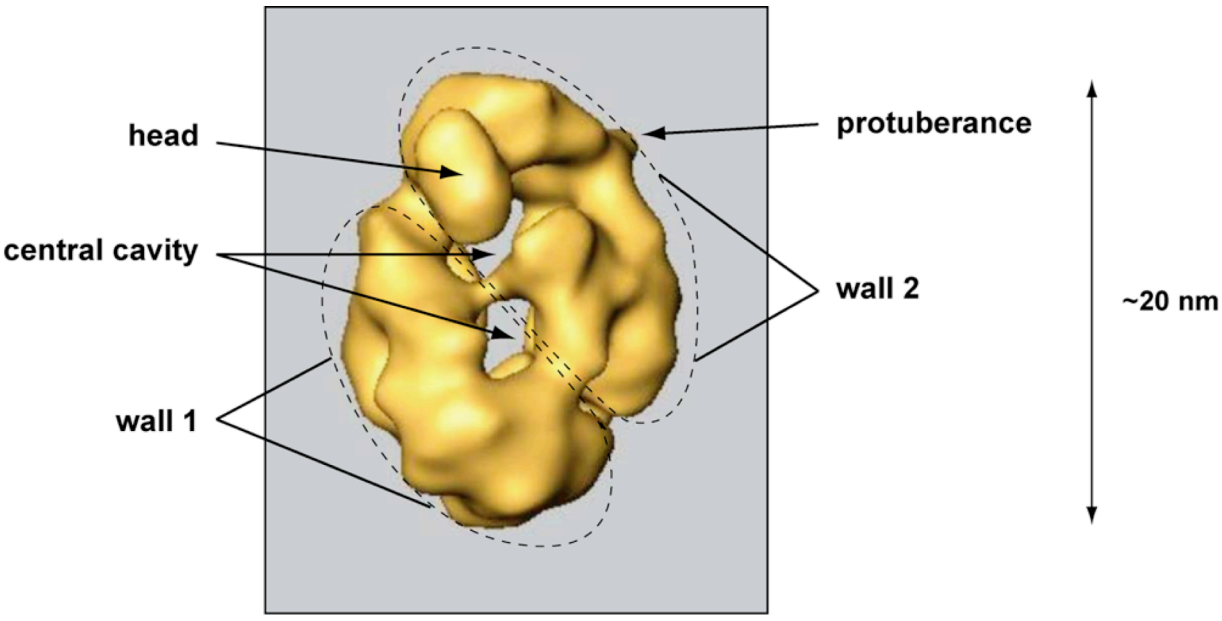

Figure 4.19. Three-dimensional model of native CPF at a resolution of ~25 Å obtained by single particle EM.
Front view of the CPF reconstruction showing the most prominent features (head, central cavity, wall 1 and 2, protuberance) of the structure. The maximun dimension of the complex is ~20 nm as indicated on the right.

The protein wall can be divided in two major parts that roughly look like two overlapping “half-shells” separated by the inner cavity (Figures 4.19, black dotted lines, and figure 4.20) and connected by thin density elements located around the cavity. One moiety is constituted by the bottom part of the particle that is a compact density element with a channel appearing on one side (Figure 4.20, bottom and lateral 2 views), and is extended upward by a roughly annular-like structure (Figure 4.20, side 2 view) from which emerges a thick protuberance denoted “bulk” (Figure 4.20, back view). The other main part of the protein wall is a massive claw-like structural element that includes the top region covering the structure (Figure 4.20, top view) and which is best seen from the side 1 view (Figure 4.20). In addition, the upper region of CPF contains two protruding elements, one big head-like element that we called “head” (Figure 4.19) and a smaller protuberance on the other side of the structure (both can be seen on the side 1 view of the figure 4.20) denoted “protuberance”. Furthermore, the front view and the lateral 1 view revealed the presence of two deep grooves, that could accommodate single-stranded pre-mRNA (Figure 4.21B and C), although this speculation is not supported by any data. The diameter of these grooves can be estimated at 20-40 Å, which would be sufficient to accommodate a single stranded RNA molecule (about 8-12 Å wide, depending on the nucleotides). The blue and white lines on figures 4.21B and C represent two examples for putative paths for pre-mRNA, assuming that the grooves are large enough to accommodate RNA molecules (see also discussion). Beginning from the upper left (Figure 4.20, front view), a long and rather straight groove is surrounded by the upper part of the wall 1, and the head on the other side (Figure 4.20, lateral 1 view), and continues on the surface of the molecule toward the lower right (Figure 4.20, front view). Branching off from this long groove toward the upper right is a narrow, smaller and curved groove that is capped by the top region of the molecule (figure 4.20, front and side 1 views). Furthermore, this smaller groove is nearly completely surrounded by a clamp-like structure located on the left of the protuberance in the top region of wall 1 (Figure 4.20, side 1 view).

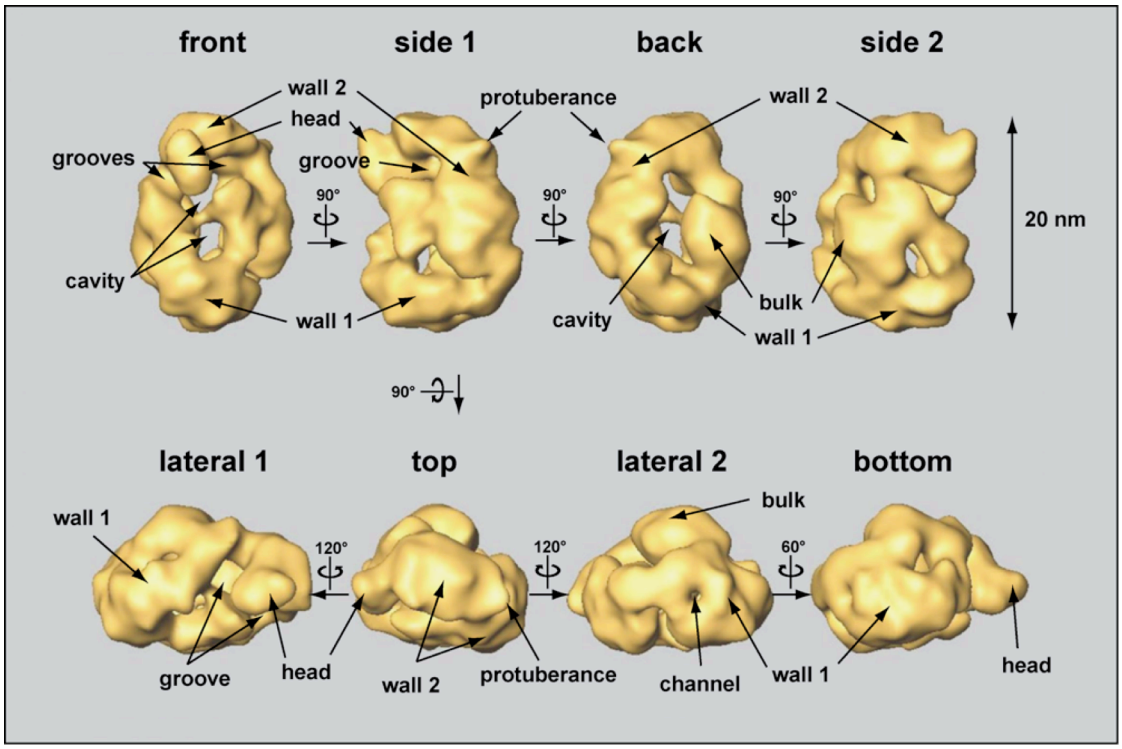

Figure 4.20. Gallery of views of the 3D CPF complex reconstruction.
Surface views of CPF showing the main structural features of the complex: the central cavity surrounded by the two moieties of the protein wall (wall 1 and 2), the head, the protuberance, the bulk, the grooves and channels. The side 1, back and side 2 views are obtained by tilting the front view stepwise along the vertical axis by 90° (first row). The top view was obtained by tilting the side 1 view along the horizontal axis by 90°. The lateral (lateral 1 and 2) views were obtained by tilting the top view along the vertical axis by 120° (second row), whereas tilting the lateral 2 view along the vertical axis gave the bottom view. The maximum size of the particle is indicated on the right.

The presence of several openings and holes in the wall structure provides contacts with the exterior to the central cavity. At least five large openings are visible, two from the front view, one from the side 1 view separating the wall moieties, the biggest one from the back view, and a fifth one in the annular-like structure from the side 2 view (Figure 4.20). Interestingly, they seem large enough to fit an RNA molecule, as they are few nanometers wide. In addition, two channels of smaller size are also present. One is easily visible by turning the molecule to the lateral 2 view, which shows a ring-like element at the bottom extremity (Figure 4.20, lateral 2 view). The ring-like structure is extended toward the left by a mass density that encloses a depression (Figure 4.21A). The entry of the depression is formed by a cleft that progressively narrows, ending with another channel of about the same size, which we

called “engulfed channel” due to its deeper location (Figure 4.21A). All of these openings represent possible paths for the pre-mRNA and suggest that the inner cavity could be the site of one or both enzymatic 3’ end processing reactions. Although speculative at this stage, this hypothesis is supported by the finding that many inner cavities described in other macromolecular machines are the site of enzymatic reactions, e.g. the exosome (Lorentzen and Conti, 2006), the 26S proteasome (Groll et al., 1997; Lowe et al., 1995) and chaperone complexes such as GroEL and TRIC (Bukau and Horwich, 1998; Leroux and Hartl, 2000; Martin and Hartl, 1997). Furthermore, the inner cavity is completely free of any density element, thus providing enough space for the pre-mRNA to be correctly accommodated.

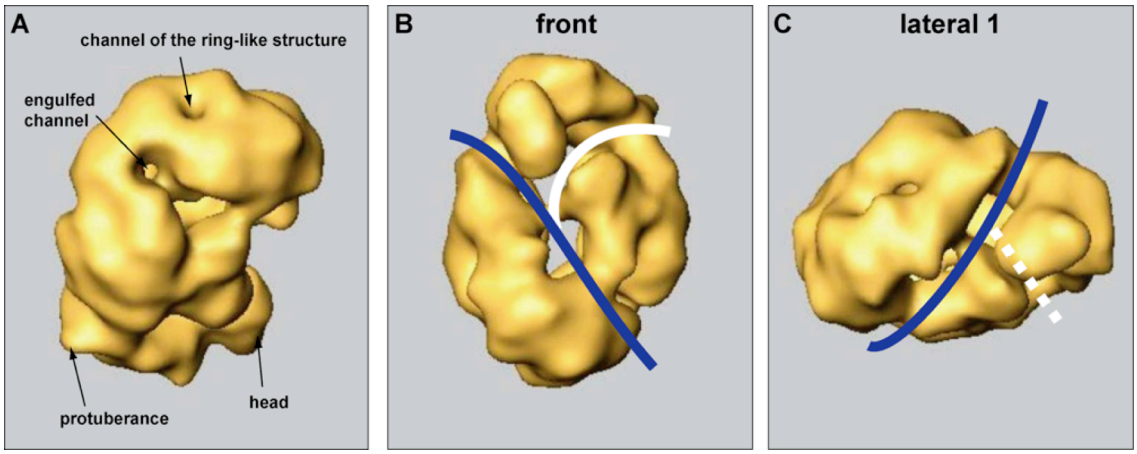

Figure 4.21. Grooves and channels at the surface of CPF are putative paths for pre-mRNA substrates.
(A) View of CPF showing two channels that may accommodate RNA molecules or perhaps nucleotides.
(B) and (C) Front and lateral 1 views showing the presence of two branching grooves at the surface of the complex. The blue and white lines represent potential path for the pre-mRNA in the larger groove and in the narrower groove, respectively.

In conclusion, the negative staining studies performed in this work have yielded a low-resolution 3D reconstitution of CPF. This gross 3D model is nevertheless sufficient to describe the most prominent features of the 3’ end processing complex. On the other hand, a higher resolution 3D structure would allow us to do subunit localization and atomic structure fitting in order to understand the structural organization of the complex. As a first goal, localizing the Ysh1p and Pap1p enzymes that catalyze 3’ end processing activities into the 3D structure would allow us to understand how CPF and the RNA substrate associate to form an active

complex. Furthermore, the interpretation of the CPF structure can be used together with biochemical data to better understand the molecular mechanism of the complex in the context of pre-mRNA 3' end processing.

4.3 Discussion

CPF is the main protein factor of the 3' end processing machinery in yeast and catalyzes the essential cleavage and polyadenylation of pre-mRNAs. As major part of this machinery, CPF has a key role in 3' end formation. Indeed, it contains both enzymes that catalyze the 3' end processing reactions, the site-specific endonuclease Ysh1p (Mandel et al., 2006) and the poly(A) polymerase Pap1p (Lingner et al., 1991a; Lingner et al., 1991b; Preker et al., 1997). Therefore, obtaining insight into the structure of CPF is crucial to decipher the molecular mechanism. So far, the crystal structures of Pap1p and Ydh1p, two of the fifteen subunits of CPF, were the only structural informations on the complex available for the scientific community (Bard et al., 2000; Mandel et al., 2006). Here, we have purified native, functional CPF by using a two-step affinity purification method followed by a combined density and fixation gradient centrifugation and have analyzed the structure of this complex by single-particle EM. Using STEM experiments we have calculated an overall size of almost one megadalton for the CPF complex. Furthermore, we report for the first time the 3D structure of CPF at a resolution of ~25 Å, which we obtained by angular reconstitution and random conical tilt analysis of negatively stained preparations of CPF (Figure 4.19).

Using electron microscopy to solve the 3D structure of CPF

Low-abundance, multisubunit complexes like CPF are difficult to study by traditional structural and analytical techniques such as X-ray crystallography and nuclear magnetic resonance (NMR). Despite both techniques can provide atomic model, they are limited to tackle very large multiprotein complexes close to or greater than one megadalton. In contrast, EM offers advantages for determining the structures of large biological machines. Indeed, specimen can be analyzed in many forms and shapes, i.e., two- or three-dimensional crystals, and single-particles with or without symmetry. Furthermore, protein assemblies can undergo conformational changes to

carry out a specific function (Alberts, 1998; Alberts and Miake-Lye, 1992). Thus, native structures of proteins can be different from those in the crystallized form. In such cases, EM studies can provide structures of multiprotein complexes in their native conformations or interacting states that are relevant to their cellular functions. In addition, sample preparation for EM is quick and simple compared to NMR or crystallography, which require higher amounts of material than those required for EM by a factor of at least 100 (usually 0,1 to 1 ng of protein is sufficient for EM). Finally, large protein complexes are often unstable during purification and crystal growth. At present, a number of protein complexes have been preferentially studied by EM because of these reasons: recent examples are the anaphase-promoting complex/cyclosome (Dube et al., 2005; Passmore et al., 2005), the spliceosomal components (Sander et al., 2006; Stark and Luhrmann, 2006), or the yeast RNA polymerase III (Fernandez-Tornero et al., 2007).

So far, no high-resolution structure is available for any of the multiprotein factors involved in pre-mRNA 3' end processing, although crystal structures of 4 individual components out of the 21 known have been obtained: until today structures were published for Pap1p (Bard et al., 2000), Pcf11p (Meinhart and Cramer, 2004), Ydh1p (Mandel et al., 2006) and Clp1p (Noble et al., 2007). Hence, this work contributes to the knowledge of structure and mechanism of the pre-mRNA 3' end processing machinery.

Purification of native CPF and preparation for EM

One crucial step in investigating structures of protein complexes is sample purification. The quality and resolution of the final 3D reconstruction strongly depends on the purification procedure. Here, we have used an affinity purification method followed by a density gradient centrifugation to isolate native CPF by tagging the Pfs2p subunit with a ProtA tag. As for most macromolecular complexes, purification conditions are critical for CPF recovery and require the presence of glycerol and NP-40. In the absence of glycerol, protein complexes may dissociate or aggregate. We have extensively searched for buffer conditions that allow the isolation of highly pure, concentrated and functional CPF. Although the absence of glycerol did not dramatically reduce the purification efficiency in terms of protein amounts, purifying the complex in the absence of NP-40 strongly reduced the rate of protein

recovery after the first affinity chromatography step over IgG matrix. Although EM of CPF molecules in the presence of low concentrations of glycerol (below 10%) was feasible, preparations of CPF in the presence of NP-40 (negative stain artefacts) resulted in severe loss of contrast, as it was already shown for other proteins. Interestingly, in our purification procedure, both glycerol and NP-40 could be included in the buffers during the first affinity chromatography step, since they could later be removed in the subsequent purification steps without affecting the efficient isolation of CPF.

Most purified multiprotein complexes are damaged when they are frozen for storage over long periods, hence reducing the quality of the sample for EM studies. Likewise, the sample for EM analysis should be prepared as quickly as possible after the final elution in order to prevent degradation inherent to stability problems that can affect most of the protein assemblies. Therefore, after purification of CPF, the material was immediately loaded on EM grids for single-particle EM studies. In spite of the high stability of protein interactions within CPF, first observations of EM grids revealed dissociation of the components of the complex. To counteract this problem, we first tried to reinforce the stability of the CPF structure by incubating the complex with synthetic RNAs and the other 3' end processing factors CF IA and Nab4p (data not shown). The assumption was that the whole pre-mRNA 3' end processing machinery, thus generated by the assembling of all the factors involved in the process, would be more stable as functional machinery. However, incubating CPF with the *CYC1* substrate, which we used for *in vitro* cleavage assays and which contains all the *cis*-acting elements required for 3' end formation, did not overcome the problem (data not shown). When Nab4p and CF IA were added to the CPF sample, the particles on EM grids did not increase in size, possibly because the added factors and the *CYC1* RNA were unable to stabilize the whole machinery. We therefore artificially stabilized the CPF structure by cross-linking the subunits with glutaraldehyde, which is a widely used method to avoid disassembling of molecular complexes.

The CPF complex has a mass of almost one megaDalton

The size of the functional CPF complex has not been investigated so far. To determine the overall mass of the complex, we performed STEM mass analysis, which revealed the intact CPF complex to have a global mass of 982 ± 126 kDa

(Figure 4.16D). This value is in good agreement with the predicted mass calculated by the sum of the masses of all individual components of ~940 kDa. This suggests that each subunit is present only once per complex. Inspection of the band pattern of CPF subunits on an SDS gel stained with silver also suggest a 1:1 stoichiometry. However, we cannot exclude that few components are not present in stoichiometric amounts. The mass difference observed between the mass calculated by STEM and the predicted mass could also be imputed to the glutaraldehyde fixation treatment of CPF. Indeed, such a cross-link adds the mass of aldehyde groups reacting with accessible arginine and lysine residues of the complex. The extent of the addition of mass due to glutaraldehyde cross-link is highly variable and depends on the specimen that is studied: as examples, no mass increase was observed for the tobacco mosaic virus (TMV), but 19% mass increase was measured for the fd phage (Muller et al., 1992). The size of CPF calculated by STEM also correlates well with gelfiltration data obtained with yeast protein extracts. When wild type protein extract was subjected to size exclusion chromatography on a calibrated Superose 6 column (Pharmacia), Western blotting analysis showed that CPF eluted over several fractions with a peak corresponding to a molecular weight of roughly one megaDalton (data not shown).

Structure of the CPF complex: where is the site of pre-mRNA 3' end processing reactions?

We present in this work an initial 3D reconstruction of CPF at a resolution of 25 Å determined by single-particle EM and the location of some of its major structural elements. It is the first example of a 3D structure obtained for a multiprotein complex involved in pre-mRNA 3' end processing. Negative staining EM of CPF preparations has revealed an overall shape rather globular and slightly elongated with a maximum diameter of ~ 20 nm (Figure 4.19).

The most prominent structural feature of CPF is the presence of a strikingly large central cavity defined by a surrounding protein wall. This characteristic element of the structure was already suggested by the observation of negative stain in the center of some class average images (Figures 4.17B and 4.18). The inner cage is connected to the outside of the structure by several openings and channels of various sizes in the protein wall (Figures 4.19 and 4.20). Hence, the whole assembly seems fragile with no core or body density that would confer hardiness to the structure,

probably explaining why the complex disassembles on EM grids. Five wide holes and two channels with roughly the same diameter are revealed by the structure. However, the number of openings may vary with a higher resolution model, which could also reveal additional structural features. It is nevertheless remarkable to find such a large space inside a multiprotein assembly and so many possibilities for the inner cavity to communicate with the exterior of the structure. Although the function of this cavity is not known, its relative importance in terms of size and location in the structure tempts us to speculate that it may be the site where 3' end processing reactions take place. Other multiprotein complexes, such as the 26S proteasome, chaperone complexes and the exosome use similar cage mechanisms. The proteolytic active site of the 26S proteasome, which degrades proteins, is enclosed in a cylindrical arrangement of subunits (Groll et al., 1997; Lowe et al., 1995), an architecture that is similar to the barrel-like conformation of the exosome, which degrades mRNAs (Lorentzen and Conti, 2006). In addition, chaperonins such as GroEL and CCT are also known to use inner cavities to fold the polypeptides they contain (Leroux and Hartl, 2000; Martin and Hartl, 1997). Likewise, the cavity within CPF, assuming it is spacious enough, could represent a reaction chamber in which pre-mRNAs would be cleaved and polyadenylated by the complex. Because polyadenylation occurs immediately after cleavage of the RNA substrate, both enzymes would be predicted to be close to one another within the structure. The inner cavity thus could provide an excellent configuration to bring together both enzymes and RNA that have a direct role in the reactions. Other subunits could regulate access to the cavity by interacting with protein factors required for accurate 3' end formation, such as Nab4p and CF IA. How the pre-mRNA enters the cavity to undergo cleavage would have to be determined. A traditional view of cleavage by CPF represents the complex sitting on the linear single-stranded RNA. However, supposing that CPF uses a cage mechanism, this conception would appear too simple. It could be necessary for the RNA substrate to fold somehow, in order for the cleavage site to reach the inner cavity. Furthermore, it is not clear at present whether the openings in the protein wall of the structure would be wide enough to accommodate a folded RNA. This question could be however addressed with a higher resolution model.

Another possibility would be that the cleavage reaction occurs at the surface of CPF, and that the upstream cleavage product enters the cavity by a hole to undergo polyadenylation. In accordance with this scenario, the two large grooves at the surface

of CPF might serve as rails to place the RNA in the appropriate position (Figure 4.21B and C). As for the openings, it is difficult to calculate the size of these grooves at the present low resolution of the structure. However, grooves seem to be wide enough to accommodate a single-stranded RNA molecule. It has been shown that many other protein machineries utilize grooves to accommodate RNAs. Examples are RNA polymerases in general, such as the E. coli RNA polI (Ollis et al., 1985) and the yeast RNAP II (Darst et al., 1991). Thus, the cleaved RNA could then reach the inner cavity by passing through one of the holes that are localized on the bottom of the grooves. A role for the channels (Figure 4.21A) in RNA binding could also be possible if they are wide enough to accommodate it. However, they are relatively narrow in comparison to the other openings, and so another possibility is that they could serve as a passage for small molecules such as substrate nucleotides like ATP.

Another potential involvement of the holes could be in the accommodation of other proteins. CPF intimately interacts with CF IA to cooperatively process pre-mRNAs at their 3' end. On the other hand, several CPF subunits associate with the CTD to couple 3' end formation and transcription. It is thus possible that some protein-protein interactions are mediated by the holes to stabilize association.

At the present, these hypotheses are highly speculative and many other scenarios are conceivable. In the next studies it will be important to determine the position of the different subunits within the CPF complex to better understand its mechanism of action. One would also have to measure precisely the diameter of the openings and of the grooves, which is difficult at the present low resolution of our model, to verify whether they can accommodate RNA molecules. Together with subunit localization, it should be then possible to understand how the pre-mRNA contacts CPF.

Protein localization and atomic structure fitting

The purpose of studying an entire protein complex is to understand the relationship between its parts as this affects the function of the complex. So far, none of the domains of our CPF structure can be assigned to known components of CPF. Further interpretation of CPF function and mechanism requires higher resolution structures and subunit localization. In the present work, a resolution of ~25 Å was obtained, but this resolution will most likely be improved by ongoing cryo-EM studies, which

usually provide higher resolution models. With a better resolution, immuno or genetic labeling and structure fitting should allow the localization of the components of the complex by further EM studies. Antibodies directed against specific subunits can be very informative to analyze the position of these subunits, as the antibodies are large enough to be seen in EM. However, antibodies can also damage the complex. In such a case, immunolabeling experiments would fail like it has for example been shown for the localization of some components of the prespliceosomal A complex (Behzadnia et al., 2007). Once a certain resolution has been obtained, it is also possible to localize a complex component directly by fitting its known atomic resolution structure into the 3D model of the complex. Although cryo-EM structures can yield significant information on the fold of a subunit of a protein complex, it is still advantageous to fit the crystal structure of individual components to a map of the complex. It may reveal subunit location and orientation, but also subunit contacts. Using this method, localization of both catalytically active subunits of CPF should be feasible. Although the atomic structure of Pap1p is available (Bard et al., 2000), the structure of Ysh1p is not solved yet. Instead of using antibodies to localize Ysh1p, one approach would be to fuse the enzyme with a tag, such as a 20 kDa SNAP tag (from Covalys) that provides additional density to the protein. This approach would prevent possible damages of the complex by using antibodies or failure in subunit localization if the epitope is not accessible. It is then possible to localize the additional mass contributed by the fusion protein in the 3D structure or to add an even larger label to the SNAP tag. For those labeled components with known high-resolution X-ray or NMR structure, fitting can be performed, avoiding systematic tagging of each subunit: successful examples are described by Dube et al. (2005), with the docking of a WD40 domain into the APC/C structure allowing the localization of the coactivator Cdh1, and by Golas et al. (2003) with the fitting of the RNA-recognition motif (RRM) domains of the p14 and SF3b49 proteins of the splicing factor SF3b.

Finally, we cannot exclude the possibility of conformational changes in the CPF complex upon RNA binding or during the processing reactions. Catalyzing a chemical reaction often requires multiprotein assemblies to be dynamic and additional EM studies could help to resolve this issue. Indeed, structural rearrangements can be investigated with reasonable precision by cryo-EM studies (Golas et al., 2005). Hence, the future work could help us to understand better the mode of action of CPF and the whole protein machinery in 3’ end formation at a molecular level.

4.4 Experimental procedures

Strains used in this study

The following *S. cerevisiae* strains were used in this study: W303-1B (*MATa; ade2-1; leu2-3,112; ura3-1; trp1-1; his3-11,15*); MO20 (*ura3-1; trp1Δ; ade2-1; leu2-3,112; his3-11,15; pfs2::TRP1* pNOP-PFS2 [ProtA-PFS2-LEU2-CEN]); YMS 162 (*ura3-1; trp1-1; ade2-1; leu2-3,112; his3-11,15; rna15::TRP1* pNOP-RNA15 [ProtA-RNA15-LEU2-CEN]); MO7 (*ura3-1; trp1Δ; ade2-1; leu2-3,112; his3-11,15; pfs2::TRP1* pFL38-PFS2-URA3); YBP7 (*ura3-1; trp1Δ; ade2-1; leu2-3,112; his3-11,15; pfs2::TRP1* pNOP-FLAG-PFS2 [ProtA-TEV-FLAG-PFS2-LEU2-CEN]).

Construction of a yeast strain expressing ProtA-FLAG-Pfs2p

A 1.47 kb fragment containing the FLAG sequence and the coding region of *PFS2* was PCR amplified from genomic DNA of the W303-1B strain using the following oligonucleotides: 5'-NcoI-flag (5'-CATATGGCCATGGTTGACTACAAGGATGAC GATGACAAGACTAGTATGGACGGGCATAAT-3') and 3'-NcoI-flag (5'-GCTGCAGGTCGACGGATCCCCGGG-3'). The PCR fragment was then digested by *Nco*I and *Sma*I and inserted in the pNOP-PFS2 plasmid (Ohnacker et al., 2000) also digested by the same restriction enzymes. The new plasmid pNOP-FLAG-PFS2, which encodes for a ProtA-TEV-FLAG-Pfs2p construct, was verified by sequencing and used to transform the haploid *PFS2* deletion strain MO7 in which Pfs2p expression is maintained by the presence of the pFL38-PFS2-URA3 plasmid (Ohnacker et al., 2000). Then, the loss of pFL38-PFS2-URA3 was forced on 5-fluoro-orotic acid medium and only cells carrying pNOP-FLAG-PFS2 could grow on leucin-lacking medium. Correct expression of the ProtA-TEV-FLAG-Pfs2p construct was verified by Western blotting analysis on protein extracts prepared from the new strain generated YBP7.

Preparation of yeast cell free extracts

For the preparation of yeast cell free extracts, a modified version of the liquid nitrogen method by Ansari and Schwer (1995) was used. 5 liters cultures were grown

in YPD (1% yeast extract, 2% bacto-peptone, 2% glucose) to an OD_{600} of 2-3. All the following procedure were performed at 4°C. Cells were harvested by centrifugation at 5000 rpm for 5 min in a Sorvall SLA-3000 rotor and washed once with 200 ml of ice-cold water and once with 200 ml of ice-cold AGK buffer [10 mM Tris-HCl pH 7.9, 200 mM KAc, 10% glycerol and 1 mM dithiothreitol (DTT)]. Cells were then resuspended in 20 ml AGK buffer and frozen as drops in liquid nitrogen. The frozen cells were stored at -70°C until they were homogenized. For homogenization, the frozen yeast cells were ground with a pestle in a mortar that was previously cooled with liquid nitrogen to avoid thawing. Grinding was continued until a fine powder was obtained. The frozen powder was transferred to a fresh tube and thawed under cold running water. Immediately after thawing, protease inhibitors were added to the following concentrations: 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin and 0.7 μg/ml pepstatin. The suspension of broken cells was gently mixed for 30 min on ice and was then centrifuged at 10000 rpm for 30 min in a Sorvall SS-34 rotor. The supernatant was removed and centrifuged at 41000 rpm in a Kontron TST 41.14 rotor for 2 h. After removing the lipids layer, the pale yellow aqueous phase was carefully saved and solid ammonium sulfate was added to reach 40% saturation. Proteins were precipitated overnight with gentle stirring and the precipitated material was pelleted by centrifugation at 10000 rpm in a Sorvall SS-34 rotor for 30 min. The pellet was finally resuspended in 1 ml of buffer E-50 (20 mM Tris-HCl pH 7.9, 0.2 mM EDTA, 50 mM Kac, 0.02% NP-40, 10% glycerol, 0.5 mM DTT, 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin and 0.7 μg/ml pepstatin) and dialyzed three times 2 h against 2 l of buffer E-50.

Affinity purification of CPF

The CPF complex was affinity purified from an extract of the yeast strain MO20 that expresses a ProtA-tagged version of Pfs2p (Ohnacker et al., 2000). The tag included two IgG-binding domains of the *Staphylococcus aureus* ProtA followed by the recognition sequence of the site-specific TEV protease (Senger et al., 1998).

Batch absorption to IgG-sepharose was done to isolate ProtA-Pfs2p or Cpf11-TAP and associated proteins from protein extracts prepared from yeast strains that express these constructs. All incubation steps described below up to the proteolytic

cleavage were done at 4°C with gentle mixing. A 1 ml aliquot of IgG-sepharose (Amersham Biosciences), 4 ml of the extract, 10% glycerol and 10 ml of buffer D-120 [20 mM Tris-HCl pH 7.9, 120 mM NaCl, 0.02% NP-40, 0.2 mM EDTA, 1 mM dithiothreitol (DTT)] including 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.4 μg/ml leupeptin and 0.7 μg/ml pepstatin were mixed in a 15 ml Falcon tube and incubated for 2 h on a rotating arm. After centrifugation, the supernatant was removed and the matrix was extensively washed five times for 20 min with 15 ml of D-120 buffer. After washing, the beads were incubated twice for 10 min with 15 ml of buffer TEV (50 mM Tris-HCl pH 7.9, 10% glycerol, 50 mM KCl, 0.5 mM EDTA, 0.02% NP-40, 1 mM DTT, 0.4 μg/ml leupeptin). For elution of the bound material, the beads were incubated with 1 ml of buffer TEV containing 400 U of recombinant TEV protease (Invitrogen) at 16°C for 90 min. The site-specific proteolytic cleavage released Pfs2p and associated proteins from the ProtA tag and thereby from the IgG-sepharose matrix. After centrifugation, the supernatant was collected in a fresh tube and the beads were washed four times with 500 μl of buffer TEV.

Elution fractions were pooled and applied to a SMART MonoQ column (100 μl column volume). The column was developed with a linear KCl gradient (50–500 mM) and 80 μl fractions were collected. CPF peak fractions eluted at approximately 320 mM KCl.

For GraFix centrifugation, 200 μl of the peak fraction and a side fraction were pooled and loaded onto a 4 ml linear (10%-30%) glycerol and (0-0.15%) glutaraldehyde gradient prepared with 50 mM Hepes, 120 mM KCl and 0.5 mM EDTA. After 18 h centrifugation in a Sorvall TH660 rotor at 34000 rpm at 4°C, the gradient was fractionated from the bottom to the top with a peristaltic pump in 5 drops (~180 μl) aliquots. For electrophoretic gel analysis, a gradient centrifugation step was conducted without glutaraldehyde to allow the proteins to be resolved on the gel.

Affinity purification of CF IA

CF IA was natively affinity purified essentially as described for CPF, except that the extract was prepared from a yeast strain that expresses a ProtA-tagged version of the CF IA subunit Rna15p. CF IA eluted from the MonoQ column at approximately 180 mM KCl.

FLAG tag purification

Purification of FLAG tagged Pfs2p and FLAG-Tad2p were performed using anti-FLAG M2 affinity gel following the manufacturer's instructions (Sigma). Briefly, 50 µg of total protein extract from FLAG-Tad2p strain and 3 ml of FLAG-Pfs2p obtained by IgG purification of ProtA-FLAG-Pfs2p were incubated for 2h at 4ºC with 0.5 ml of anti-FLAG M2 affinity gel previously washed with 0.1M glycine pH 3.5 and equilibrated 3 times with 10 ml TEK120 buffer (50 mM Tris-HCl pH 7.9, 120 mM KCl, 0.5 mM EDTA). Unbound proteins were removed by washing 5 times with TEK120 and captured proteins were eluted in 5 fractions with 0.5 ml of TEK120 and 150 ng/µl of 3xFLAG peptide. Eluates were then analyzed by SDS PAGE and sliverstaining.

Western blotting analysis

Proteins were separated on 10% SDS polyacrylamide gels, transferred to nitrocellulose membranes, and incubated according to standard procedures with the antibodies indicated in the legends to the figures. Peroxidase-conjugated swine anti-rabbit immunoglobulins (DAKO) were used as secondary antibody for detection of the primary antibodies.

Identification of polypeptides in CPF by mass spectrometry analysis

Bands were excised from SDS gels, destained with 30% acetonitrile in 0.1 M ammonium bicarbonate, and washed with water. The gel pieces were reswollen with 4 ml of 1 mM Tris-HCl (pH 8.8) containing 0.2 mg of trypsin (Promega) after drying in a speedvac evaporator. The digestion was carried out overnight at room temperature. 7 ml 50% acetonitrile containing 0.3% trifluoroacetic acid (TFA) was added, and the content was vortexed, centrifuged for 2 min, and sonicated for 5 min. 1 ml of the supernatant was mixed with 1 ml of saturated alpha-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. NanoElectrospray Tandem Mass Spectrometry: Unseparated peptide mixtures obtained after tryptic in-gel digestion of still unknown polypeptides were analyzed by nanoelectrospray tandem mass spectrometry (Wilm and Mann, 1996). Tryptic peptides were extracted as described (Fountoulakis and Langen, 1997), desalted, and further concentrated on a pulled capillary containing approximately 100 nl POROS R2 reverse phase material (Perceptive Biosystems, Framingham, MA) and eluted with 1 ml of 60% methanol in 5% formic acid directly into the nanoelectrospray needle as described. Electrospray mass spectra were acquired on an API 365 triple quadrupole mass spectrometer (Sciex, Toronto, Canada) equipped with a nanoES ion source developed by Wilm and Mann (1996). Q1 scans were performed 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 such that the fragment masses could be assigned better than 1 Da.

In vitro pre-mRNA 3' end processing assays

^{32}P-labelled *CYC1* and pre-*CYC1* RNAs were prepared as run-off transcripts and gel purified as described previously (Minvielle-Sebastia et al., 1994; Preker et al., 1995).

A standard *in vitro* processing reaction was carried out in a 25 μl reaction volume containing up to 5 μl of protein extract or purified protein factors. For polyadenylation assay, the reaction mixture contained 2% polyethylene glycol, 75 mM potassium glutamate, 0.01% NP-40, 2 mM ATP, 1.8 mM magnesium acetate, 20 mM creatine phosphate, 1 mM DTT, 0.2 mg/ml creatine kinase and 0.2 units of RNAguard (Amersham Biosciences). The volume was brought to 25 μl upon addition of buffer E-50 (20 mM Tris-HCl pH 7.9, 0.2 mM EDTA, 50 mM Kac, 0.02% NP-40, 10% glycerol) if was necessary. When the cleavage reaction was assayed, EDTA was used instead of magnesium acetate to prevent poly(A) addition, and degradation of the downstream cleavage fragment (Chen and Moore, 1992). The reactions were incubated 1h at the temperatures indicated in the figure legends. They were then stopped by addition of 75 μl of a stop mix (100 mM Tris-HCl pH7.9, 12.6 mM EDTA, 150 mM NaCl, 1% SDS, 0.2 μg/μl of proteinase K and 0.05 μg/μl glycogen)

and incubated for 1h at 42°C. The RNAs were recovered by ethanol precipitation, dried and resuspended in 6 μl of formamide buffer (95% formamide, 20 mM EDTA, 0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF). Samples were heated at 95°C for 5 min and loaded on a denaturing 6% polyacrylamide gel. The gel was exposed overnight to an X-ray film at -70°C.

STEM and image analysis

The glutaraldehyde cross-linked CPF sample was diluted 4 times, and 5 µl aliquots were adsorbed for 60s to glow-discharged STEM films (thin carbon films that span a thick fenestrated carbon layer covering 200-mesh/inch, gold-plated copper grids). The grids were then blotted, washed on 5 drops of quartz double-distilled water, and freeze-dried at -80 °C and 5 x 10^{-8} Torr overnight in the microscope. Tobacco mosaic virus (TMV) particles (kindly provided by R. Diaz Avalos) were used for absolute mass calibration. These particles were similarly adsorbed to separate STEM films, washed on 4 drops of 10 mM ammonium acetate, and air-dried.

A STEM HB-5 vacuum generator interfaced to a modular computer system (Tietz Video and Image Processing System GmbH, Gauting, Germany) was employed. Series of 512 x 512-pixel, dark-field images were recorded from the unstained sample at an acceleration voltage of 80 kV and a nominal magnification of 200,000x. The recording dose ranged from 300 to 900 electrons/nm^2. The digital images were evaluated using the program package IMPSYS (Muller et al., 1992). Accordingly, the projections were selected in circular boxes, and the total scattering of each region was calculated. The background scattering of the carbon support film was then subtracted, and the mass was calculated. The results were scaled according to the mass measured for TMV and corrected for beam-induced mass loss on the basis of the behavior of proteins in a similar mass range (Muller and Engel, 2001). The mass values were displayed in histograms and described by Gauss curves. The overall experimental uncertainty of the results was estimated from the corresponding SE (SE = SD/√n) and the ~5% uncertainty in the calibration of the instrument.

EM of CPF

Grid and negative stain preparations

The CPF complexes were sandwiched in 2% uranyl formate between two layers of thin carbon using holey carbon grids as support. The samples were imaged with a Philips CM200 FEG electron microscope (high tension 160 kV, magnification 154000x, temperature –179 °C) equipped with a Gatan cryoholder. Images were taken under low-dose conditions at various defoci (0.7–1.5 µm) on a 4x4k CCD camera (TVIPS, Gauting, Germany) in 2x binning mode. The Imagic-5 software package (Image Science, Berlin, Germany) was used for image processing of 6000 individual molecular images.

Angular reconstitution

A total of 6000 molecular images of CPF were selected. The raw images were corrected for the contrast transfer function by phase flipping as described (Sander et al., 2003a). Standard image processing procedures, including alignment (Sander et al., 2003b), multivariate statistical analysis (van Heel, 1984), classification, and angular reconstitution (Van Heel, 1987), were used as implemented in the Imagic-5 image processing software (van Heel et al., 1996). The final resolution of the 3D maps was ~25 Å as determined by the Fourier-shell-correlation function using the 5 sigma criterion (or the FSC 0.5 criteria).

Random conical tilt

Tilted image pairs of CPF were recorded at a magnification of 154000x and a defocus of ~2 µm first with a tilt of 45° and afterwards at a tilt of 0°. Images were taken on a 4k × 4k CCD camera using two-fold pixel binning. Particles were selected manually from the untilted and from the corresponding tilted images. Untilted images were aligned by using iteratively refined class averages as references and subsequent MSA for classification into class averages. Angles of the tilted images of the five largest classes were determined according to (Radermacher, 1988). A 3D map was calculated independently for each class, and an average 3D of the five largest classes was obtained after 3D alignment of individual 3D structures by using the 3D alignment tool of the software AmiraDev 2.3 (TGS Europe, Merignac Cedex, France).

5. General discussion

Pre-mRNA 3' end processing is the last maturation step that the primary transcripts synthesized by RNAP II have to undergo during maturation. This processing event consists of the specific endonucleolytic cleavage at the poly(A) site and subsequent poly(A) tail addition at the 3' end of the upstream cleavage product. At first sight, this two-step reaction looks simple, nevertheless, a surprisingly complex protein machinery is required to complete the task. Interestingly, protein factors involved in 3' end formation are highly conserved from yeast to mammals, suggesting that the reaction mechanism is similar.

In *S. cerevisiae*, CPF, CF IA, Nab4p and Pab1p are the protein factors required for specific 3' end processing of a primary transcript. During the last years, every polypeptide involved in the process was characterized by numerous biochemical and genetic studies. This was greatly facilitated by the establishment of *in vitro* assays that allowed researchers to test for 3' end processing phenotypes of mutant strains, and by efficient affinity purification methods that allowed analysis of the protein composition of each factor. CPF is an important factor for pre-mRNA 3' end processing, as this multiprotein complex is required for both cleavage and polyadenylation reactions and contains 75% of the polypeptides involved in 3' end formation. Until recently, CPF was believed to comprise nine subunits. Then, several affinity purification studies combined with MS identified six novel components of CPF (Dichtl et al., 2002b; Gavin et al., 2002; Nedea et al., 2003; Walsh et al., 2002). One of these studies identified Cpf11p, a non-essential protein with unknown function, to interact with CPF (Dichtl et al., 2002b). In this study we show that Cpf11p is not a genuine subunit of CPF and has no role in 3' end processing activities *in vitro*. TAP tag purification of Cpf11p did not yield the components of CPF, whereas TAP-tagging of any other subunit of the complex resulted in CPF purification. Also intriguing is the fact that we did not find any homologue of Cpf11p in other species, whereas most of the proteins involved in pre-mRNA 3' end processing are evolutionary conserved. Furthermore, protein extracts from mutants lacking Cpf11p did not show any defect in cleavage and polyadenylation *in vitro*. Finally, MS analyses we did on highly purified CPF never found Cpf11p to copurify with the known components of the complex. There is however the possibility that

Cpf11p only transiently interacts with CPF, which would explain its copurification with the 3' end processing complex by Dichtl et al. (2002b). As it was shown for the CTD and other processing factors, Cpf11p could contact CPF under certain conditions. To test this hypothesis, further work concerning the identification of Cpf11p function and additional binding tests with CPF components would be necessary. To characterize Cpf11p, one could identify the proteins that specifically interact with TAP tagged Cpf11p. This information could provide indications on the pathway or cellular process Cpf11p is involved in.

Although all the proteins involved in pre-mRNA 3' end processing have been characterized and numerous protein-protein and protein-RNA interactions were identified, the complexity of the 3' end processing machinery is such that no model was proposed so far to explain the assembly with the pre-mRNA and the mechanism of the 3' end processing reactions. One way leading to a better understanding of the molecular mechanisms of the 3' end processing reactions is to analyze the three-dimensional structure of the factors involved in the process. In this study, we attempted to investigate this issue and as a result we report the first 3D structure of CPF at a 25 Å resolution. Interestingly, the structure revealed a large central cavity that could serve as a reaction chamber in which processing reactions could occur. Such cage-like protein complexes with enzymatic activities have already been described, like the 26S proteasome (Groll et al., 1997; Lowe et al., 1995), chaperone complexes (Martin and Hartl, 1997) or the exosome (Hernandez et al., 2006). However, the overall structure of CPF is not sufficient to understand how and where reactions exactly take place. For this, it would be important to localize the different subunits in the structure to understand how they are arranged. We are currently trying to localize Ysh1p and Pap1p, the two enzymes responsible for cleavage and polyadenylation reactions, respectively. The crystal structure of Pap1p can be used to fit the polymerase directly into the CPF structure. For Ysh1p, we use another approach, by fusing the protein to a SNAP tag, which allows localization of Ysh1p by visualization of the additional mass provided by the SNAP tag on the 3D structure. This method is preferred to immunolocalization for two reasons. First, the use of antibodies can damage the complex. Second, if the epitope of Ysh1p is localized inside the structure, it is unlikely to be recognized by the antibody. For subunit localization, it is necessary to use a 3D structure with a high resolution to minimize multiple possible fittings caused by a too low-resolution model. Currently, we are

using cryo-EM to improve the resolution of the 3D model described in this work. In addition to subunit localization, incubation of pure CPF with a synthetic RNA molecule that contains all the polyadenylation signals required for 3' end formation and further observation by EM could reveal how CPF accommodates pre-mRNAs.

Furthermore, it is likely that CPF is subjected to conformational changes during processing of the pre-mRNA, so not only one but probably several 3D models corresponding to different stages of the 3' end formation could be obtained. Indeed, protein complexes that have enzymatic activities need to be dynamic to fulfil the reactions they are assigned to, and single-particle cryo-EM is the method of choice to study conformational changes in multiprotein assemblies.

It would also be very interesting and informative to integrate the 3D structure of CF IA in the model, as CF IA is another important factor involved in pre-mRNA 3' end processing. Consistent with its requirement for specific cleavage and polyadenylation, CF IA contacts the pre-mRNA and displays a network of protein-protein interactions with CPF components (Nedea et al., 2003). Thus, a complete understanding of the molecular mechanism of pre-mRNA 3' end processing probably involves CF IA. Smaller than CPF, CF IA is composed of only four polypeptides and can be natively affinity purified to homogeneity through ProtA tagging of the Rna15p subunit. Thus, solving a 3D structure of CF IA should be feasible. Then, a three-dimensional model of a tripartite reconstituted system comprising CPF, CF IA and a synthetic pre-mRNA could be generated, which would provide information on the molecular mechanism of 3' end formation.

A 3D structure of CPF could also help to understand the basis of the association with RNAP II, as many CPF components were shown to interact with the CTD. Indeed, the analysis of CPF structural features could show how the CTD enhances the efficiency of pre-mRNA cleavage. One possibility could be that the interaction of the CTD with CPF would induce conformational changes that favor enzymatic activity.

Moreover, as mainly *in vitro* studies demonstrated that pre-mRNA 3' end processing and transcription are interconnected, we started to investigate the relevance of this coupling *in vivo*. We tagged proteins expected to be involved in these processes with fluorophores (GFP and some variants) and generated a series of yeast strains in order to perform FRET and FRAP experiments. So far, the fluorescence signals obtained from the strains we have constructed were too low to

use these techniques. Despite the recent expansion in the diversity of available fluorophores, none of those we used was bright enough to provide a signal sufficiently above the autofluorescence of yeast cells. In addition, the small size of yeast cells requires us to systematically construct diploid strains, which is time consuming. Finally, we are currently testing an overexpression system for the tagged versions of the 3' end processing proteins to increase the amount of proteins of interest and thus to improve the fluorescence signals.

Future work could also investigate in more detail the functional relationship of pre-mRNA 3' end processing to other pre-mRNA processing events, such as capping and splicing. For example, splicing and 3' end processing are well conserved from yeast to mammals and were shown to be coupled, as revealed by several studies in mammals (Kyburz et al., 2006; Wassarman and Steitz, 1993). Recent work suggests that coupling also occurs in yeast (A. Kyburz, unpublished data). The resolution of the 3D structure of factors involved in pre-mRNA processing events and the investigation of effects caused by mutant forms of these factors would help to understand the mechanism by which the maturation events are connected.

6. References

Abe, A., Hiraoka, Y. and Fukasawa, T. (1990) Signal sequence for generation of mRNA 3' end in the *Saccharomyces cerevisiae* GAL7 gene. *EMBO J*, **9**, 3691-3697.

Abruzzi, K.C., Lacadie, S. and Rosbash, M. (2004) Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J*, **23**, 2620-2631.

Abuodeh, R., Wei, H. and Yuan, D. (1998) Effect of upstream RNA processing on selection of mu S versus mu M poly(A) sites. *Nucleic Acids Res*, **26**, 5417-5424.

Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol Cell Biol*, **6**, 2932-2943.

Ahn, S.H., Kim, M. and Buratowski, S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell*, **13**, 67-76.

Alberts, B. and Miake-Lye, R. (1992) Unscrambling the puzzle of biological machines: the importance of the details. *Cell*, **68**, 415-420.

Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, **92**, 291-294.

Allen, N.P., Huang, L., Burlingame, A. and Rexach, M. (2001) Proteomic analysis of nucleoporin interacting proteins. *J Biol Chem*, **276**, 29268-29274.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389-3402.

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.

Anderson, J.T., Paddy, M.R. and Swanson, M.S. (1993) PUB1 is a major nuclear and cytoplasmic polyadenylated RNA-binding protein in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **13**, 6102-6113.

Ansari, A. and Hampsey, M. (2005) A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, **19**, 2969-2978.

Ansari, A. and Schwer, B. (1995) SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO J*, **14**, 4001-4009.

Aranda, A., Perez-Ortin, J.E., Moore, C. and del Olmo, M.L. (1998) Transcription termination downstream of the *Saccharomyces cerevisiae* FBP1 [changed from FPB1] poly(A) site does not depend on efficient 3'end processing. *RNA*, **4**, 303-318.

Aranda, A. and Proudfoot, N. (2001) Transcriptional termination factors for RNA polymerase II in yeast. *Mol Cell*, **7**, 1003-1011.

Aravind, L. and Koonin, E.V. (1999) DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. *Nucleic Acids Res*, **27**, 1609-1618.

Ares, M., Jr., Grate, L. and Pauling, M.H. (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA*, **5**, 1138-1139.

Barabino, S.M.L., 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.

Barabino, S.M., Ohnacker, M. and Keller, W. (2000) Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J*, **19**, 3778-3787.

Bard, J., Zhelkovsky, A.M., Helmling, S., Earnest, T.N., Moore, C.L. and Bohm, A. (2000) Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP. *Science*, **289**, 1346-1349.

Barilla, D., Lee, B.A. and Proudfoot, N.J. (2001) Cleavage/polyadenylation factor IA associates with the carboxyl- terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **98**, 445-450.

Baskaran, R., Dahmus, M.E. and Wang, J.Y. (1993) Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A*, **90**, 11167-11171.

Baudin-Baillieu, A., Guillemet, E., Cullin, C. and Lacroute, F. (1997) Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. *Yeast*, **13**, 353-356.

Beelman, C.A. and Parker, R. (1995) Degradation of mRNA in eukaryotes. *Cell*, **81**, 179-183.

Behzadnia, N., Golas, M.M., Hartmuth, K., Sander, B., Kastner, B., Deckert, J., Dube, P., Will, C.L., Urlaub, H., Stark, H. and Luhrmann, R. (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J*, **26**, 1737-48.

Bentley, D. (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol*, **11**, 347-351.

Bentley, D.L. (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol*, **17**, 251-256.

Beyer, A.L., Bouton, A.H. and Miller, O.L., Jr. (1981) Correlation of hnRNP structure and nascent transcript cleavage. *Cell*, **26**, 155-165.

Beyer, K., Dandekar, T. and Keller, W. (1997) RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3'-end processing of pre-mRNA. *J Biol Chem*, **272**, 26769-26779.

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

Bienroth, S., Keller, W. and Wahle, E. (1993) Assembly of a processive messenger RNA polyadenylation complex. *EMBO J*, **12**, 585-594.

Bird, G., Zorio, D.A. and Bentley, D.L. (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol*, **24**, 8963-8969.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W. and Proudfoot, N.J. (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science*, **280**, 298-301.

Boeck, R., Tarun, S., Jr., Rieger, M., Deardorff, J.A., Muller-Auer, S. and Sachs, A.B. (1996) The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *J Biol Chem*, **271**, 432-438.

Borman, A.M., Michel, Y.M. and Kean, K.M. (2000) Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Res*, **28**, 4068-4075.

Brown, C.E., Tarun, S.Z., Jr., Boeck, R. and Sachs, A.B. (1996) PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **16**, 5744-5753.

Brown, K.M. and Gilmartin, G.M. (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell*, **12**, 1467-1476.

Bukau, B. and Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell*, **92**, 351-366.

Buratowski, S. (2003) The CTD code. *Nat Struct Biol*, **10**, 679-680.

Buratowski, S. (2005) Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol*, **17**, 257-261.

Butler, J.S. and Platt, T. (1988) RNA processing generates the mature 3' end of yeast *CYC1* messenger RNA *in vitro*. *Science*, **242**, 1270-1274.

Butler, J.S., Sadhale, P.P. and Platt, T. (1990) RNA processing *in vitro* produces mature 3' ends of a variety of *Saccharomyces cerevisiae* mRNAs. *Mol Cell Biol*, **10**, 2599-2605.

Callebaut, I., Moshous, D., Mornon, J.P. and de Villartay, J.P. (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res*, **30**, 3592-3601.

Calvo, O. and Manley, J.L. (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol Cell*, **7**, 1013-1023.

Calvo, O. and Manley, J.L. (2005) The transcriptional coactivator PC4/Sub1 has multiple functions in RNA polymerase II transcription. *EMBO J*, **24**, 1009-1020.

Caponigro, G. and Parker, R. (1995) Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. *Genes Dev*, **9**, 2421-2432.

Caponigro, G. and Parker, R. (1996) Mechanisms and control of mRNA turnover in *Saccharomyces cerevisiae*. *Microbiol Rev*, **60**, 233-249.

Carrington, J.C. and Dougherty, W.G. (1988) A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. *Proc Natl Acad Sci U S A*, **85**, 3391-3395.

Carty, S.M. and Greenleaf, A.L. (2002) Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA processing. *Mol Cell Proteomics*, **1**, 598-610.

Chanfreau, G., Noble, S.M. and Guthrie, C. (1996) Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science*, **274**, 1511-1514.

Chen, J. and Moore, C. (1992) Separation of factors required for cleavage and polyadenylation of yeast pre-mRNA. *Mol Cell Biol*, **12**, 3470-3481.

Chen, F., MacDonald, C.C. and Wilusz, J. (1995) Cleavage site determinants in the mammalian polyadenylation signal. *Nucleic Acids Res*, **23**, 2614-2620.

Chen, S. and Hyman, L.E. (1998) A specific RNA-protein interaction at yeast polyadenylation efficiency elements. *Nucleic Acids Res*, **26**, 4965-4974.

Chen, Z., Li, Y. and Krug, R.M. (1999) Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery. *EMBO J*, **18**, 2273-2283.

Cheng, H., He, X. and Moore, C. (2004) The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. *Mol Cell Biol*, **24**, 2932-2943.

Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, **11**, 3319-3326.

Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O. and Reinberg, D. (1999) A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev*, **13**, 1540-1552.

Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J. and Buratowski, S. (2001) Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev*, **15**, 3319-3329.

Chou, Z.F., Chen, F. and Wilusz, J. (1994) Sequence and position requirements for uridylate-rich downstream elements of polyadenylation signals. *Nucleic Acids Res*, **22**, 2525-2531.

Christofori, G. and Keller, W. (1988) 3' cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor and a snRNP. *Cell*, **54**, 875-889.

Colgan, D.F., Murthy, K.G., Prives, C. and Manley, J.L. (1996) Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature*, **384**, 282-285.

Colgan, D.F. and Manley, J.L. (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev*, **11**, 2755-2766.

Connelly, S. and Manley, J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev*, **2**, 440-452.

Coppola, J.A., Field, A.S. and Luse, D.S. (1983) Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proc Natl Acad Sci U S A*, **80**, 1251-1255.

Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr. and Dahmus, M.E. (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A*, **82**, 7934-7938.

Corden, J.L. and Patturajan, M. (1997) A CTD function linking transcription to splicing. *Trends Biochem Sci*, **22**, 413-416.

Cramer, P., Srebrow, A., Kadener, S., Werbajh, S., de la Mata, M., Melen, G., Nogues, G. and Kornblihtt, A.R. (2001a) Coordination between transcription and pre-mRNA processing. *FEBS Lett*, **498**, 179-182.

Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001b) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**, 1863-1876.

Curtis, D., Lehmann, R. and Zamore, P.D. (1995) Translational regulation in development. *Cell*, **81**, 171-178.

Dahmus, M.E. (1995) Phosphorylation of the C-terminal domain of RNA polymerase II. *Biochim Biophys Acta*, **1261**, 171-182.

Dahmus, M.E. (1996) Reversible phosphorylation of the C-terminal domain of RNA polymerase II. *J Biol Chem*, **271**, 19009-19012.

Darst, S.A., Edwards, A.M., Kubalek, E.W. and Kornberg, R.D. (1991) Three-dimensional structure of yeast RNA polymerase II at 16 A resolution. *Cell*, **66**, 121-128.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H. and Keller, W. (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J*, **19**, 5895-5904.

Deardorff, J.A. and Sachs, A.B. (1997) Differential effects of aromatic and charged residue substitutions in the RNA binding domains of the yeast poly(A)-binding protein. *J Mol Biol*, **269**, 67-81.

Decker, C.J. and Parker, R. (2002) mRNA decay enzymes: decappers conserved between yeast and mammals. *Proc Natl Acad Sci U S A*, **99**, 12512-12514.

Dehe, P.M., Dichtl, B., Schaft, D., Roguev, A., Pamblanco, M., Lebrun, R., Rodriguez-Gil, A., Mkandawire, M., Landsberg, K., Shevchenko, A., Shevchenko, A., Rosaleny, L.E., Tordera, V., Chavez, S., Stewart, A.F. and Geli, V. (2006) Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem*, **281**, 35404-35412.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W. and Barabino, S.M. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J Biol Chem*, **279**, 35788-35797.

Dheur, S., Vo le, T.A., Voisinet-Hakil, F., Minet, M., Schmitter, J.M., Lacroute, F., Wyers, F. and Minvielle-Sebastia, L. (2003) Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. *EMBO J*, **22**, 2831-2840.

Dichtl, B. and Keller, W. (2001) Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J*, **20**, 3197-3209.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S. and Keller, W. (2002a) Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J*, **21**, 4125-4135.

Dichtl, B., Blank, D., Ohnacker, M., Friedlein, A., Roeder, D., Langen, H. and Keller, W. (2002b) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol Cell*, **10**, 1139-1150.

Dichtl, B., Aasland, R. and Keller, W. (2004) Functions for *S. cerevisiae* Swd2p in 3' end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. *RNA*, **10**, 965-977.

Dominski, Z. and Marzluff, W.F. (1999) Formation of the 3' end of histone mRNA. *Gene*, **239**, 1-14.

Dominski, Z., Yang, X.C., Purdy, M., Wagner, E.J. and Marzluff, W.F. (2005) A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol*, **25**, 1489-1500.

Dube, P., Herzog, F., Gieffers, C., Sander, B., Riedel, D., Muller, S.A., Engel, A., Peters, J.M. and Stark, H. (2005) Localization of the coactivator Cdh1 and the cullin subunit Apc2 in a cryo-electron microscopy model of vertebrate APC/C. *Mol Cell*, **20**, 867-879.

DuBois, R.N., McLane, M.W., Ryder, K., Lau, L.F. and Nathans, D. (1990) A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence. *J Biol Chem*, **265**, 19185-19191.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J. and Ingles, C.J. (2002) Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. *RNA*, **8**, 1102-1111.

Feaver, W.J., Gileadi, O., Li, Y. and Kornberg, R.D. (1991) CTD kinase associated with yeast RNA polymerase II initiation factor b. *Cell*, **67**, 1223-1230.

Feng, Z.H., Wilson, S.E., Peng, Z.Y., Schlender, K.K., Reimann, E.M. and Trumbly, R.J. (1991) The yeast GLC7 gene required for glycogen accumulation encodes a type 1 protein phosphatase. *J Biol Chem*, **266**, 23796-23801.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W. and Schoehn, G. (2007) Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol Cell*, **25**, 813-823.

Fong, N. and Bentley, D.L. (2001) Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev*, **15**, 1783-1795.

Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**, 929-933.

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.

Frank, J. and van Heel, M. (1982) Correspondence analysis of aligned images of biological particles. *J Mol Biol*, **161**, 134-137.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R.K. and Agrawal, R.K. (1995) A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature*, **376**, 441-444.

Furuichi, Y. and Shatkin, A.J. (2000) Viral and cellular mRNA capping: past and prospects. *Adv Virus Res*, **55**, 135-184.

Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C. and Faye, G. (2003) Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. *EMBO J*, **22**, 1588-1598.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.A., Copley, R.R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G. and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737-741.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J.,

Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W. and Johnston, M. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387-391.

Gietz, R.D., Schiestl, R.H., Willems, A.R. and Woods, R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, **11**, 355-360.

Gil, A. and Proudfoot, N.J. (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation. *Cell*, **49**, 399-406.

Gilbert, W. and Guthrie, C. (2004) The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. *Mol Cell*, **13**, 201-212.

Gilmartin, G.M., Fleming, E.S., Oetjen, J. and Graveley, B.R. (1995) CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev*, **9**, 72-83.

Golas, M.M., Sander, B., Will, C.L., Luhrmann, R. and Stark, H. (2003) Molecular architecture of the multiprotein splicing factor SF3b. *Science*, **300**, 980-984.

Golas, M.M., Sander, B., Will, C.L., Luhrmann, R. and Stark, H. (2005) Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, **17**, 869-883.

Gonzalez, C.I., Ruiz-Echevarria, M.J., Vasudevan, S., Henry, M.F. and Peltz, S.W. (2000) The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. *Mol Cell*, **5**, 489-499.

Graber, J.H., Cantor, C.R., Mohr, S.C. and Smith, T.F. (1999a) *In silico* detection of control signals: mRNA 3'-end-processing sequences in diverse species. *Proc Natl Acad Sci U S A*, **96**, 14055-14060.

Graber, J.H., Cantor, C.R., Mohr, S.C. and Smith, T.F. (1999b) Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res*, **27**, 888-894.

Graber, J.H. (2003) Variations in yeast 3'-processing cis-elements correlate with transcript stability. *Trends Genet*, **19**, 473-476.

Greenberg, M.L., Reiner, B. and Henry, S.A. (1982) Regulatory mutations of inositol biosynthesis in yeast: isolation of inositol-excreting mutants. *Genetics*, **100**, 19-33.

Greenblatt, J., Nodwell, J.R. and Mason, S.W. (1993) Transcriptional antitermination. *Nature*, **364**, 401-406.

Groll, M., Ditzel, L., Lowe, J., Stock, D., Bochtler, M., Bartunik, H.D. and Huber, R. (1997) Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature*, **386**, 463-471.

Gross, S. and Moore, C. (2001) Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I. *Proc Natl Acad Sci U S A*, **98**, 6080-6085.

Grunberg-Manago, M. (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annu Rev Genet*, **33**, 193-227.

Gu, M. and Lima, C.D. (2005) Processing the message: structural insights into capping and decapping mRNA. *Curr Opin Struct Biol*, **15**, 99-106.

Guan, Y., Dunham, M.J. and Troyanskaya, O.G. (2007) Functional Analysis of Gene Duplications in *Saccharomyces cerevisiae*. *Genetics*, **175**, 933-943.

Guo, Z., Russo, P., Yun, D.-F., Butler, J.S. and Sherman, F. (1995) Redundant 3' end-forming signals for the yeast *CYC1 mRNA*. *Proceedings of the National Academy of Sciences USA*, **92**, 4211-4214.

Guo, Z. and Sherman, F. (1995) 3'-end-forming signals of yeast mRNA. *Mol Cell Biol*, **15**, 5983-5990.

Guo, Z. and Sherman, F. (1996) 3'-end-forming signals of yeast mRNA. *Trends in Biochemical Sciences*, **21**, 477-481.

Haghighat, A. and Sonenberg, N. (1997) eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J Biol Chem*, **272**, 21677-21680.

Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C. and Cole, C.N. (2002) Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol*, **22**, 6441-6457.

Hancock, L.C., Behta, R.P. and Lopes, J.M. (2006) Genomic analysis of the Opi- phenotype. *Genetics*, **173**, 621-634.

Harauz, G., Stoeffler-Meilicke, M. and van Heel, M. (1987) Characteristic views of prokaryotic 50S ribosomal subunits. *J Mol Evol*, **26**, 347-357.

He, X., Khan, A.U., Cheng, H., Pappas, D.L., Jr., Hampsey, M. and Moore, C.L. (2003) Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes Dev*, **17**, 1030-1042.

He, X. and Moore, C. (2005) Regulation of yeast mRNA 3' end processing by phosphorylation. *Mol Cell*, **19**, 619-629.

Heidmann, S., Schwindewolf, C., Stumpf, G. and Domdey, H. (1994) Flexibility and interchangeability of polyadenylation signals in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **14**, 4633-4642.

Helmling, S., Zhelkovsky, A. and Moore, C.L. (2001) Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. *Mol Cell Biol*, **21**, 2026-2037.

Henry, M., Borland, C.Z., Bossie, M. and Silver, P.A. (1996) Potential RNA binding proteins in *Saccharomyces cerevisiae* identified as suppressors of temperature-sensitive mutations in *NPL3*. *Genetics*, **142**, 103-115.

Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B. and Robinson, C.V. (2006) Subunit architecture of multimeric complexes isolated directly from cells. *EMBO Rep*, **7**, 605-610.

Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. and Jensen, T.H. (2001) Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature*, **413**, 538-542.

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

Hirose, Y., Tacke, R. and Manley, J.L. (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev*, **13**, 1234-1239.

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

Ho, C.K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B. and Shuman, S. (1998) The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, **273**, 9577-9585.

Ho, C.K. and Shuman, S. (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell*, **3**, 405-411.

Hofmann, I., Schnolzer, M., Kaufmann, I. and Franke, W.W. (2002) Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes. *Mol Biol Cell*, **13**, 1665-1676.

Holm, L. and Sander, C. (1995) DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem Sci*, **20**, 345-347.

Huang, Y. and Carmichael, G.G. (1996) Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol*, **16**, 1534-1542.

Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S. and O'Shea, E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686-691.

Hyman, L.E., Seiler, S.H., Whoriskey, J. and Moore, C.L. (1991) Point mutations upstream of the yeast ADH2 poly(A) site significantly reduce the efficiency of 3'-end formation. *Mol Cell Biol*, **11**, 2004-2012.

Iizuka, N., Najita, L., Franzusoff, A. and Sarnow, P. (1994) Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from *Saccharomyces cerevisiae*. *Mol Cell Biol*, **14**, 7322-7330.

Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J Bacteriol*, **153**, 163-168.

Jacobson, A. and Favreau, M. (1983) Possible involvement of poly(A) in protein synthesis. *Nucleic Acids Res*, **11**, 6353-6368.

Jacobson, A. and Peltz, S.W. (1996) Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu Rev Biochem*, **65**, 693-739.

Jansen, G., Wu, C., Schade, B., Thomas, D.Y. and Whiteway, M. (2005) Drag&Drop cloning in yeast. *Gene*, **344**, 43-51.

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.

Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.

Jove, R. and Manley, J.L. (1984) *In vitro* transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites. *J Biol Chem*, **259**, 8513-8521.

Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12**, 5-14.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H. and Keller, W. (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J*, **23**, 616-626.

Keller, W., Bienroth, S., Lang, K.M. and Christofori, G. (1991) Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J*, **10**, 4241-4249.

Keller, W. and Minvielle-Sebastia, L. (1997) A comparison of mammalian and yeast pre-mRNA 3'-end processing. *Current Opinion in Cell Biology*, **9**, 329-336.

Kellis, M., Birren, B.W. and Lander, E.S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature*, **428**, 617-624.

Kelly, W.G., Dahmus, M.E. and Hart, G.W. (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem*, **268**, 10416-10424.

Kessler, M.M., Zhao, J. and Moore, C.L. (1996) Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. *J Biol Chem*, **271**, 27167-27175.

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A. and Moore, C.L. (1997) Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev*, **11**, 2545-2556.

Kim, H. and Lee, Y. (2001) Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I. *Biochem Biophys Res Commun*, **289**, 513-518.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F. and Buratowski, S. (2004a) Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J*, **23**, 354-364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F. and Buratowski, S. (2004b) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, **432**, 517-522.

Komarnitsky, P., Cho, E.J. and Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, **14**, 2452-2460.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J. and Nogues, G. (2004) Multiple links between transcription and splicing. *RNA*, **10**, 1489-1498.

Kotovic, K.M., Lockshon, D., Boric, L. and Neugebauer, K.M. (2003) Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol Cell Biol*, **23**, 5768-5779.

Kramer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem*, **65**, 367-409.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C. and Hampsey, M. (2004) Ssu72 Is an RNA polymerase II CTD phosphatase. *Mol Cell*, **14**, 387-394.

Kyburz, A., Sadowski, M., Dichtl, B. and Keller, W. (2003) The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic Acids Res*, **31**, 3936-3945.

Kyburz, A., Friedlein, A., Langen, H. and Keller, W. (2006) Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol Cell*, **23**, 195-205.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*, **121**, 713-724.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Le, H., Tanguay, R.L., Balasta, M.L., Wei, C.C., Browning, K.S., Metz, A.M., Goss, D.J. and Gallie, D.R. (1997) Translation initiation factors eIF-iso4G and eIF-4B interact with the poly(A)-binding protein and increase its RNA binding activity. *J Biol Chem*, **272**, 16247-16255.

Leipe, D.D., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol*, **317**, 41-72.

Leroux, M.R. and Hartl, F.U. (2000) Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. *Curr Biol*, **10**, R260-264.

Leuschner, P.J., Ameres, S.L., Kueng, S. and Martinez, J. (2006) Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Rep*, **7**, 314-320.

Lewis, J.D., Gunderson, S.I. and Mattaj, I.W. (1995) The influence of 5' and 3' end structures on pre-mRNA metabolism. *J Cell Sci Suppl*, **19**, 13-19.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B. and Bentley, D.L. (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell*, **9**, 1101-1111.

Lingner, J., Kellermann, J. and Keller, W. (1991a) Cloning and expression of the essential gene for poly(A) polymerase from *S. cerevisiae*. *Nature*, **354**, 496-498.

Lingner, J., Radtke, I., Wahle, E. and Keller, W. (1991b) Purification and characterization of poly(A) polymerase from *Saccharomyces cerevisiae*. *J Biol Chem*, **266**, 8741-8746.

Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr. and Shenk, T. (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A*, **84**, 8306-8310.

Long, R.M., Elliott, D.J., Stutz, F., Rosbash, M. and Singer, R.H. (1995) Spatial consequences of defective processing of specific yeast mRNAs revealed by fluorescent in situ hybridization. *RNA*, **1**, 1071-1078.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P. and Pringle, J.R. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, **14**, 953-961.

Lopez, P.J. and Seraphin, B. (1999) Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. *RNA*, **5**, 1135-1137.

Lorentzen, E. and Conti, E. (2006) The exosome and the proteasome: nano-compartments for degradation. *Cell*, **125**, 651-654.

Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. and Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. *Science*, **268**, 533-539.

Lowell, J.E., Rudner, D.Z. and Sachs, A.B. (1992) 3'-UTR-dependent deadenylation by the yeast poly(A) nuclease. *Genes Dev*, **6**, 2088-2099.

Luo, Y. and Goss, D.J. (2001) Homeostasis in mRNA initiation: wheat germ poly(A)-binding protein lowers the activation energy barrier to initiation complex formation. *J Biol Chem*, **276**, 43083-43086.

Luo, W. and Bentley, D. (2004) A ribonucleolytic rat torpedoes RNA polymerase II. *Cell*, **119**, 911-914.

Luo, W., Johnson, A.W. and Bentley, D.L. (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev*, **20**, 954-965.

MacDonald, C.C., Wilusz, J. and Shenk, T. (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol*, **14**, 6647-6654.

Mandart, E. and Parker, R. (1995) Effects of mutations in the *Saccharomyces cerevisiae RNA14, RNA15,* and *PAP1* genes on polyadenylation *in vivo*. *Mol Cell Biol*, **15**, 6979-6986.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L. and Tong, L. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*, **444**, 953-956.

Mangus, D.A., Smith, M.M., McSweeney, J.M. and Jacobson, A. (2004) Identification of factors regulating poly(A) tail synthesis and maturation. *Mol Cell Biol*, **24**, 4196-4206.

Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499-506.

Manley, J.L., Yu, H. and Ryner, L. (1985) RNA sequence containing hexanucleotide AAUAAA directs efficient mRNA polyadenylation *in vitro*. *Mol Cell Biol*, **5**, 373-379.

Manley, J.L. and Takagaki, Y. (1996) The end of the message - another link between yeast and mammals. *Science*, **274**, 1481-1482.

Martin, G. and Keller, W. (1996) Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. *EMBO J*, **15**, 2593-2603.

Martin, G., Keller, W. and Doublie, S. (2000) Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. *EMBO J*, **19**, 4193-4203.

Martin, J. and Hartl, F.U. (1997) Chaperone-assisted protein folding. *Curr Opin Struct Biol*, **7**, 41-52.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D.L. (1997a) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, **11**, 3306-3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M. and Bentley, D.L. (1997b) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357-361.

McCracken, S., Rosonina, E., Fong, N., Sikes, M., Beyer, A., O'Hare, K., Shuman, S. and Bentley, D. (1998) Role of RNA polymerase II carboxy-terminal domain in coordinating transcription with RNA processing. *Cold Spring Harb Symp Quant Biol*, **63**, 301-309.

Meinhart, A. and Cramer, P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, **430**, 223-226.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. and Cramer, P. (2005) A structural perspective of CTD function. *Genes Dev*, **19**, 1401-1415.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F. and Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A*, **98**, 12902-12907.

Millhouse, S. and Manley, J.L. (2005) The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol Cell Biol*, **25**, 533-544.

Minvielle-Sebastia, L., Winsor, B., Bonneaud, N. and Lacroute, F. (1991) Mutations in the yeast *RNA14* and *RNA15* genes result in an abnormal mRNA decay rate; Sequence analysis reveals an RNA-binding domain in the RNA15 protein. *Mol Cell Biol*, **11**, 3075-3087.

Minvielle-Sebastia, L., Preker, P.J. and Keller, W. (1994) RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. *Science*, **266**, 1702-1705.

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 premessenger RNA 3'-end formation. *Proc Natl Acad Sci U S A*, **94**, 7897-7902.

Minvielle-Sebastia, L., Beyer, K., Krecic, A.M., Hector, R.E., Swanson, M.S. and Keller, W. (1998) Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO J*, **17**, 7454-7468.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M. and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell*, **91**, 457-466.

Mizrahi, N. and Moore, C. (2000) Posttranslational phosphorylation and ubiquitination of the *Saccharomyces cerevisiae* Poly(A) polymerase at the S/G(2) stage of the cell cycle. *Mol Cell Biol*, **20**, 2794-2802.

Moore, C.L. and Sharp, P.A. (1985) Accurate cleavage and polyadenylation of exogenous RNA substrate. *Cell*, **41**, 845-855.

Morris, D.P. and Greenleaf, A.L. (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, **275**, 39935-39943.

Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A. and Berezney, R. (1996) A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A*, **93**, 8253-8257.

Muller, S.A., goldie, k.n., burki, r., haring, r. and Engel, A. (1992) factors influencing the precision of quantitative scanning transmission electron microscopy. *Ultramicroscopy*, **46**, 317-334.

Muller, S.A. and Engel, A. (2001) Structure and mass analysis by scanning transmission electron microscopy. *Micron*, **32**, 21-31.

Munroe, D. and Jacobson, A. (1990a) mRNA poly(A) tail, a 3' enhancer of translational initiation. *Mol Cell Biol*, **10**, 3441-3455.

Munroe, D. and Jacobson, A. (1990b) Tales of poly(A): a review. *Gene*, **91**, 151-158.

Murthy, K.G. and Manley, J.L. (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev*, **9**, 2672-2683.

Myers, L.C., Lacomis, L., Erdjument-Bromage, H. and Tempst, P. (2002) The yeast capping enzyme represses RNA polymerase II transcription. *Mol Cell*, **10**, 883-894.

Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. and Miyawaki, A. (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol*, **20**, 87-90.

Nagy, P.L., Griesenbeck, J., Kornberg, R.D. and Cleary, M.L. (2002) A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3. *Proc Natl Acad Sci U S A*, **99**, 90-94.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L. and Greenblatt, J. (2003) Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *J Biol Chem*, **278**, 33000-33010.

Nemeth, A., Krause, S., Blank, D., Jenny, A., Jeno, P., Lustig, A. and Wahle, E. (1995) Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II. *Nucleic Acids Res*, **23**, 4034-4041.

Noble, C.G., Walker, P.A., Calder, L.J. and Taylor, I.A. (2004) Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic Acids Res*, **32**, 3364-3375.

Noble, C.G., Hollingworth, D., Martin, S.R., Ennis-Adeniran, V., Smerdon, S.J., Kelly, G., Taylor, I.A. and Ramos, A. (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol*, **12**, 144-151.

Noble, C.G., Beuth, B. and Taylor, I.A. (2007) Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. *Nucleic Acids Res*, **35**, 87-99.

Nonet, M., Sweetser, D. and Young, R.A. (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell*, **50**, 909-915.

O'Connor, J.P. and Peebles, C.L. (1992) *PTA1*, an essential gene of *Saccharomyces cerevisiae* affecting pre-tRNA processing. *Mol Cell Biol*, **12**, 3843-3856.

Ohnacker, M., Barabino, S.M., Preker, P.J. and Keller, W. (2000) The WD-repeat protein Pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. *EMBO J*, **19**, 37-47.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. (1985) Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature*, **313**, 762-766.

Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439-451.

Osheim, Y.N., Proudfoot, N.J. and Beyer, A.L. (1999) EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol Cell*, **3**, 379-387.

Pappas, D.L., Jr. and Hampsey, M. (2000) Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **20**, 8343-8351.

Passmore, L.A., Booth, C.R., Venien-Bryan, C., Ludtke, S.J., Fioretto, C., Johnson, L.N., Chiu, W. and Barford, D. (2005) Structural analysis of the anaphase-promoting complex reveals multiple active sites and insights into polyubiquitylation. *Mol Cell*, **20**, 855-866.

Peggie, M.W., MacKelvie, S.H., Bloecher, A., Knatko, E.V., Tatchell, K. and Stark, M.J. (2002) Essential functions of Sds22p in chromosome stability and nuclear localization of PP1. *J Cell Sci*, **115**, 195-206.

Perez-Canadillas, J.M. (2006) Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J*, **25**, 3167-3178.

Perumal, K. and Reddy, R. (2002) The 3' end formation in small RNAs. *Gene Expr*, **10**, 59-78.

Preker, P.J., Lingner, J., Minvielle-Sebastia, L. and Keller, W. (1995) The *FIP1* gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. *Cell*, **81**, 379-389.

Preker, P.J., Ohnacker, M., Minvielle-Sebastia, L. and Keller, W. (1997) A multisubunit 3'-end processing factor from yeast containing poly(A) polymerase and homologues of the subunits of mammalian cleavage and polyadenylation specificity factor. *EMBO J*, **16**, 4727 - 4737.

Proudfoot, N.J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem. Sci.*, **14**, 105-110.

Proudfoot, N. (1991) Poly(A) signals. *Cell*, **64**, 671-674.

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

Proudfoot, N. and O'Sullivan, J. (2002) Polyadenylation: a tail of two complexes. *Curr Biol*, **12**, R855-857.

Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.

Proudfoot, N. (2004) New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol*, **16**, 272-278.

Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218-229.

Radermacher, M., Wagenknecht, T., Verschoor, A. and Frank, J. (1987) Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc*, **146**, 113-136.

Radermacher, M. (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *J Electron Microsc Tech*, **9**, 359-394.

Ramaswamy, N.T., Li, L., Khalil, M. and Cannon, J.F. (1998) Regulation of yeast glycogen metabolism and sporulation by Glc7p protein phosphatase. *Genetics*, **149**, 57-72.

Rasmussen, E.B. and Lis, J.T. (1993) *In vivo* transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A*, **90**, 7923-7927.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, **17**, 1030-1032.

Rizzo, M.A., Springer, G.H., Granada, B. and Piston, D.W. (2004) An improved cyan fluorescent protein variant useful for FRET. *Nat Biotechnol*, **22**, 445-449.

Robert, F., Blanchette, M., Maes, O., Chabot, B. and Coulombe, B. (2002) A human RNA polymerase II-containing complex associated with factors necessary for spliceosome assembly. *J Biol Chem*, **277**, 9302-9306.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L. and Buratowski, S. (2000) Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol Cell Biol*, **20**, 104-112.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R. and Stewart, A.F. (2001) The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J*, **20**, 7137-7148.

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

Russnak, R., Nehrke, K.W. and Platt, T. (1995) REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation. *Mol Cell Biol*, **15**, 1689-1697.

Russo, P., Li, W.Z., Guo, Z. and Sherman, F. (1993) Signals that produce 3' termini in CYC1 mRNA of the yeast Saccharomyces cerevisiae. *Mol Cell Biol*, **13**, 7836-7849.

Ryan, K., Calvo, O. and Manley, J.L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*, **10**, 565-573.

Sachs, A.B., Bond, M.W. and Kornberg, R.D. (1986) A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. *Cell*, **45**, 827-835.

Sachs, A.B. and Davis, R.W. (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. *Cell*, **58**, 857-867.

Sachs, A. and Wahle, E. (1993) Poly(A) tail metabolism and function in eucaryotes. *J Biol Chem*, **268**, 22955-25968.

Sadhale, P.P. and Platt, T. (1992) Unusual aspects of *in vitro* RNA processing in the 3' regions of the GAL1, GAL7, and GAL10 genes in *Saccharomyces cerevisiae*. *Mol Cell Biol*, **12**, 4262-4270.

Sadowski, M., Dichtl, B., Hubner, W. and Keller, W. (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J*, **22**, 2167-2177.

Sander, B., Golas, M.M. and Stark, H. (2003a) Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *J Struct Biol*, **142**, 392-401.

Sander, B., Golas, M.M. and Stark, H. (2003b) Corrim-based alignment for improved speed in single-particle image processing. *J Struct Biol*, **143**, 219-228.

Sander, B., Golas, M.M., Makarov, E.M., Brahms, H., Kastner, B., Luhrmann, R. and Stark, H. (2006) Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol Cell*, **24**, 267-278.

Schroeder, S.C., Zorio, D.A., Schwer, B., Shuman, S. and Bentley, D. (2004) A function of yeast mRNA cap methyltransferase, Abd1, in transcription by RNA polymerase II. *Mol Cell*, **13**, 377-387.

Schroeder, S.C., Schwer, B., Shuman, S. and Bentley, D. (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev*, **14**, 2435-2440.

Senger, B., Simos, G., Bischoff, F.R., Podtelejnikov, A., Mann, M. and Hurt, E. (1998) Mtr10p functions as a nuclear import receptor for the mRNA-binding protein Npl3p. *EMBO J*, **17**, 2196-2207.

Setyono, B. and Greenberg, J.R. (1981) Proteins associated with poly(A) and other regions of mRNA and hnRNA molecules as investigated by crosslinking. *Cell*, **24**, 775-783.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E. and Tsien, R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol*, **22**, 1567-1572.

Shatkin, A.J. and Manley, J.L. (2000) The ends of the affair: capping and polyadenylation. *Nat Struct Biol*, **7**, 838-842.

Sheets, M.D., Ogg, S.C. and Wickens, M.P. (1990) Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res*, **18**, 5799-5805.

Shen, E.C., Henry, M.F., Weiss, V.H., Valentini, S.R., Silver, P.A. and Lee, M.S. (1998) Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev*, **12**, 679-691.

Shibagaki, Y., Itoh, N., Yamada, H., Nagata, S. and Mizumoto, K. (1992) mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylytransferase subunit from *Saccharomyces cerevisiae*. *J Biol Chem*, **267**, 9521-9528.

Skaar, D.A. and Greenleaf, A.L. (2002) The RNA Polymerase II CTD Kinase CTDK-I Affects Pre-mRNA 3' Cleavage/Polyadenylation through the Processing Component Pti1p. *Mol Cell*, **10**, 1429-1439.

Smith, T.F., Gaitatzes, C., Saxena, K. and Neer, E.J. (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci*, **24**, 181-185.

Stark, H., Mueller, F., Orlova, E.V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R. and van Heel, M. (1995) The 70S Escherichia coli ribosome at 23 A resolution: fitting the ribosomal RNA. *Structure*, **3**, 815-821.

Stark, H., Orlova, E.V., Rinke-Appel, J., Junke, N., Mueller, F., Rodnina, M., Wintermeyer, W., Brimacombe, R. and van Heel, M. (1997) Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy. *Cell*, **88**, 19-28.

Stark, H. and Luhrmann, R. (2006) Cryo-electron microscopy of spliceosomal components. *Annu Rev Biophys Biomol Struct*, **35**, 435-457.

Stark, M.J. (1996) Yeast protein serine/threonine phosphatases: multiple roles and diverse regulation. *Yeast*, **12**, 1647-1675.

Steinmetz, E.J. and Brow, D.A. (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol*, **16**, 6993-7003.

Steinmetz, E.J. and Brow, D.A. (2003) Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol*, **23**, 6339-6349.

Stiller, J.W. and Cook, M.S. (2004) Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell*, **3**, 735-740.

Stumpf, G. and Domdey, H. (1996) Dependence of yeast pre-mRNA 3'-end processing on CFT1: a sequence homolog of the mammalian AAUAAA binding factor. *Science*, **274**, 1517-1519.

Sun, Z.W. and Hampsey, M. (1996) Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Mol Cell Biol*, **16**, 1557-1566.

Tacahashi, Y., Helmling, S. and Moore, C.L. (2003) Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. *Nucleic Acids Res*, **31**, 1744-1752.

Takagaki, Y., Ryner, L.C. and Manley, J.L. (1989) Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev*, **3**, 1711-1724.

Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J. and Shenk, T. (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev*, **4**, 2112-2120.

Takagaki, Y. and Manley, J.L. (1994) A polyadenylation factor subunit is the human homologue of the *Drosophila suppressor of forked* protein. *Nature*, **372**, 471-474.

Tan, Y.S., Morcos, P.A. and Cannon, J.F. (2003) Pho85 phosphorylates the Glc7 protein phosphatase regulator Glc8 in vivo. *J Biol Chem*, **278**, 147-153.

Tarun, S.Z., Jr. and Sachs, A.B. (1995) A common function for mRNA 5' and 3' ends in translation initiation in yeast. *Genes Dev*, **9**, 2997-3007.

Tarun, S.Z., Jr. and Sachs, A.B. (1996) Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J*, **15**, 7168-7177.

Tarun, S.Z., Jr., Wells, S.E., Deardorff, J.A. and Sachs, A.B. (1997) Translation initiation factor eIF4G mediates in vitro poly(A) tail-dependent translation. *Proc Natl Acad Sci U S A*, **94**, 9046-9051.

Thuresson, A.C., Astrom, J., Astrom, A., Gronvik, K.O. and Virtanen, A. (1994) Multiple forms of poly(A) polymerases in human cells. *Proc Natl Acad Sci U S A*, **91**, 979-983.

Tsien, R.Y. (1998) The green fluorescent protein. *Annu Rev Biochem*, **67**, 509-544.

Valentini, S.R., Weiss, V.H. and Silver, P.A. (1999) Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. *RNA*, **5**, 1-9.

van Heel, M. (1984) Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, **13**, 165-183.

Van Heel, M. (1987) Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, **21**, 111-123.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R. and Schatz, M. (1996) A new generation of the IMAGIC image processing system. *J Struct Biol*, **116**, 17-24.

van Hoof, A. (2005) Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics*, **171**, 1455-1461.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G. and Keller, W. (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol*, **3**, e189.

Vo, L.T., Minet, M., Schmitter, J.M., Lacroute, F. and Wyers, F. (2001) Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. *Mol Cell Biol*, **21**, 8346-8356.

Wach, A., Brachat, A., Alberti-Segui, C., Rebischung, C. and Philippsen, P. (1997) Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in Saccharomyces cerevisiae. *Yeast*, **13**, 1065-1075.

Wahle, E. (1991) Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *J Biol Chem*, **266**, 3131-3139.

Wahle, E. and Keller, W. (1992) The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu Rev Biochem*, **61**, 419-440.

Wahle, E. and Kuhn, U. (1997) The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. *Prog Nucleic Acid Res Mol Biol*, **57**, 41-71.

Wahle, E. and Ruegsegger, U. (1999) 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev*, **23**, 277-295.

Walsh, E.P., Lamont, D.J., Beattie, K.A. and Stark, M.J. (2002) Novel interactions of *Saccharomyces cerevisiae* type 1 protein phosphatase identified by single-step affinity purification and mass spectrometry. *Biochemistry*, **41**, 2409-2420.

Wassarman, K.M. and Steitz, J.A. (1993) Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP? *Genes Dev*, **7**, 647-659.

West, S., Gromak, N. and Proudfoot, N.J. (2004) Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522-525.

Whitelaw, E. and Proudfoot, N. (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J*, **5**, 2915-2922.

Will, C.L. and Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, **13**, 290-301.

Williams-Hart, T., Wu, X. and Tatchell, K. (2002) Protein phosphatase type 1 regulates ion homeostasis in *Saccharomyces cerevisiae*. *Genetics*, **160**, 1423-1437.

Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem*, **68**, 1-8.

Wolfe, K.H. and Shields, D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, **387**, 708-713.

Wu, W.H., Pinto, I., Chen, B.S. and Hampsey, M. (1999) Mutational analysis of yeast TFIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIB. *Genetics*, **153**, 643-652.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M. and Corden, J.L. (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A*, **93**, 6975-6980.

Zacharias, D.A., Violin, J.D., Newton, A.C. and Tsien, R.Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*, **296**, 913-916.

Zaret, K.S. and Sherman, F. (1982) DNA sequence required for efficient transcription termination in yeast. *Cell*, **28**, 563-573.

Zhang, J. and Corden, J.L. (1991) Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J Biol Chem*, **266**, 2297-2302.

Zhang, Z., Fu, J. and Gilmour, D.S. (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev*, **19**, 1572-1580.

Zhao, J., Kessler, M.M. and Moore, C.L. (1997) Cleavage factor II of *Saccharomyces cerevisiae* contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. *J Biol Chem*, **272**, 10831-10838.

Zhao, J., Hyman, L. and Moore, C. (1999a) 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, J., Kessler, M., Helmling, S., O'Connor, J.P. and Moore, C. (1999b) Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. *Mol Cell Biol*, **19**, 7733-7740.

Zhelkovsky, A., Helmling, S. and Moore, C. (1998) Processivity of the *Saccharomyces cerevisiae* poly(A) polymerase requires interactions at the carboxyl-terminal RNA binding domain. *Mol Cell Biol*, **18**, 5942-5951.

Zhelkovsky, A., Tacahashi, Y., Nasser, T., He, X., Sterzer, U., Jensen, T.H., Domdey, H. and Moore, C. (2006) The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in *Saccharomyces cerevisiae*. *RNA*, **12**, 435-445.

Zielinski, R., Hellman, U., Kubinski, K. and Szyszka, R. (2006) Fip1-an essential component of the *Saccharomyces cerevisiae* polyadenylation machinery is phosophorylated by protein kinase CK2. *Mol Cell Biochem*, **286**, 191-197.

7. Acknowledgements

I deeply thank Prof. Walter Keller for giving me the opportunity and freedom to do my Ph.D. thesis in his laboratory and for his enthusiasm about my projects. I would like to thank him for always supporting me even when results did not come as fast as expected.

I am also grateful to Prof. Andreas Engel for accepting to be the co-referee of my thesis.

I would like to thank Prof. Ueli Aebi for being the president of my thesis committee.

I would like to thank very much my collaborators Holger Stark, Shirley Müller, Philippe Ringler, Françoise Erne-Brand and Salome Röck for many discussions, comments and advices.

I also thank my lab-colleagues with whom it was a pleasure to work. I thank all of them for their advices, their help and for the good atmosphere they created in the lab: Diana Blank, Monika Garas, Sophie Jaeger, Anne Knoth, Andrea Kyburz, Georges Martin, Cedric Meury, Erica Oesch, Christiane Rammelt, Salome Röck, Salvo San Paolo, Marie-Joëlle Schmidt, Stepanka Vanacova, Verena Widmer.
Special thanks to Georges Martin (the Pope of PAP), Christiane Rammelt, Stepanka Vanacova and Marie-Joëlle Schmidt for critically reading this thesis-manuscript.

Other special thanks to Monika, Salome and Marie-Joëlle for mutual support, pleasant atmosphere in the lab and especially for being also good friends outside of the lab.

I don't forget the former members of the lab who welcomed me when I arrived: Sabine Dettwiler, Bernhard Dichtl, Wolfgang Hübner, Isabelle Kaufmann, Martin Sadowski, Myriam Schaub and Jeannette Wolf.
Many thanks to Wolfgang Hübner, Martin Sadowski and Myriam Schaub for the good times we had in lab 275.

I also would like to thank Bernhard Dichtl for his help, comments and suggestions about my work. Bernhard, I hope you will be as successful with your own group as you were in Walter's lab.

I also think about my long time friends, Romain Montard, Maud Cassard and Isabelle Mounie, who are still present in my life despite sometimes the distance. Their friendship means for me much more than what they might think. Romain, "mon vieux pote", we know each other since we are 3 years old. We grew up together for almost 30 years, but I hope this is just the beginning of our friendship. I also want to tell you that I'm convinced that you will become a famous and talented ophthalmologist. Isa, "ma cocotte", it is also a long time we know each other, but I am still amazed by your kindness and your humanity, quite rare nowadays. I hope that life in Lisbon will become comfortable. Maud, my irreplaceable "binôme", we went through university studies together as a perfect duo and have experienced a lot of extremely good times during this period that I will never forget. I promise I will have more time now to visit you and Augusta more often.

I am deeply grateful to my parents who always supported and encouraged me during my life, whatever the decision I made. This work was possible thanks to your love.

The last thank goes to Vero, my "little butterfly" who was incredibly patient with me and who helped me a lot during the time I wrote the manuscript of this thesis. Your support and love were and are invaluable. Thanks for being the person you are.

8. Curriculum vitae

Name:	Paguet
First name:	Bertrand
Date of birth:	January 22, 1977
Place of birth:	Besançon, France
Nationality:	French
Marital status:	single

Education

June 1995:	French "Baccalauréat" from the Lycée Saint-Jean in Besançon, France
Sept. 1995-June 1997	Medical studies at the Medical Faculty of Besançon
Sept. 1997-June 1998	DEUG (Diplôme d'Etudes Universitaires Générales; first two years of university degree) in Life Sciences at the University of Besançon, France
Sept. 1998-June 1999	Licence in Biochemistry (Bachelor of Sciences equivalent) at the University of Besançon, France
Sept. 1999-June 2000	Maîtrise in Biochemistry (Master's degree equivalent) at the University of Besançon, France
Sept. 2000-Oct. 2001	DEA in Molecular and Cellular Biology (Diplôme d'études approfondies; Diploma thesis equivalent) in the laboratory of Pr. Jean-Marc Egly at the IGBMC, University Louis Pasteur, Strasbourg, FRANCE. Research topic: Role of TFIIH in androgen response.
Since Nov. 2001	Ph.D. studies in the laboratory of Pr. Dr. Walter Keller, at the Department of Cell Biology, Biozentrum, University of Basel, Switzerland.

Communication

RNOMICS meeting, IMM, Lisbon, march 2005
Oral presentation: *Interactions between the CTD of RNA polymerase II and the 3' end processing factors Yhh1p and Pcf11p*

Messenger RNA 3' ends: Interconnections with transcription and mRNA turnover meeting, Oxford, march 2005
poster: *Seeing is believing: Structural and dynamic in vivo approaches to understand the function of pre-mRNA 3' end processing factors*

SWISS RNA workshop, Bern, 2002-2003-2004-2005
Poster presentation every year

I declare that I wrote this thesis

"Biochemical characterization and three-dimensional structure analysis of the yeast cleavage and polyadenylation factor CPF"

with the help indicated and only handed it in to the faculty of science of the University of Basel and to no other faculty and no other university.

Bertrand Paguet

Basel, 2 may 2007