Characterization of factors involved in the coupling of 3' end processing and splicing and in the 3' end formation of mRNA precursors

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Andrea Martina Kyburz Kooznetsoff

aus Unterkulm (AG)

Basel, 2006

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Walter Keller und Prof. Dr. Bernhard Dichtl.

Basel, den 4.7.2006

Prof. Dr. Hans-Jakob Wirz

Dekan der Philosophisch-Naturwissenschaftlichen Fakultät

Table of contents

Table of contents 3

Summary 4

Chapter1: Introduction to RNA processing 7

Chapter 2: Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing 36

Chapter 3: Knock down of CPSF100, SF3b130 and SF3b155 by RNAi 77

Chapter 4: Rse1p and its potential role in coupling splicing and 3' end processing in yeast 91

Chapter 5: Characterization of the tRNA endonuclease complex and its involvement in 3' end processing 117

Chapter 6: The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3' end formation 134

Final Discussion 157

References 159

Acknowledgements 195

Curriculum vitae 196

Summary

Eukaryotic mRNA precursors are processed at their 5' and 3' ends and are spliced prior to their export from the nucleus to the cytoplasm. Although all three processing reactions can be studied separately *in vitro,* they are coupled *in vivo*.

3' end processing of most mammalian pre-mRNAs involves endonucleolytic cleavage followed by polyadenylation of the upstream cleavage product. Cleavage and polyadenylation specificity factor (CPSF) is a multiprotein complex, which together with cleavage factor I_m and II_m (CF I_m, CF II_m), cleavage stimulatory factor (CstF), poly(A) polymerase (PAP) and nuclear poly(A) binding protein 1 (PABPN1) is required for 3' end formation.

We have found that the U2 snRNA and subunits of the splicing factors 3a and 3b (SF3a, SF3b), which are components of the U2 snRNP, were also present in highly purified CPSF fractions. GST pull-down experiments indicated a direct interaction of CPSF subunits with SF3b49 and SF3b130. Furthermore, antibodies directed against CPSF100 co-immunoprecipitated subunits of SF3a and SF3b and the U2 snRNA Taken together our results show that subunits of CPSF and the U2 snRNP directly interact with each other.

In order to analyze whether this interaction plays a role in the coupling of 3' end processing and splicing, we depleted CPSF subunits from HeLa cell nuclear extract and tested the extracts for splicing activity. CPSF100-depleted extract showed no detectable cleavage activity and its splicing activity was significantly reduced in coupled assays but not in un-coupled assays. Moreover, pre-mRNAs containing mutations in the binding site of SF3b were not only less efficiently spliced but they also showed reduced cleavage activity. Interestingly, efficient cleavage required the presence of the U2 snRNA in coupled but not in un-coupled assays. Based on our studies, we propose that the interactions between CPSF and U2 snRNP contribute to the coupling of splicing and 3' end formation.

Furthermore, we depleted CPSF100 and the U2 snRNP subunits SF3b155 and SF3b130 by means of RNAi. We observed that knock down of both SF3b proteins caused high lethality of the cells indicating that these polypeptides are essential. However, depletion of CPSF100 did not result in a significant increase in cell

mortality, suggesting that the protein is either not essential that the knock down was not efficient enough to result in cell lethality or that CPSF100 shares redundant functions with another protein. We were able to show that SF3b155 and SF3b130 are required for efficient splicing *in vivo* but did not detect a splicing deficiency in CPSF100 depleted cells. Knock down of neither of the proteins resulted in an observable 3' end processing deficiency. Further work is required to address the question if the U2 snRNP and CPSF couple splicing and 3' end processing *in vivo*.

Splicing and 3' end formation are highly conserved mechanisms from mammals to yeast and the two organisms share homologues of most of the proteins involved in the two reactions. To test whether the coupling mechanism mediated by CPSF and the U2 snRNP is conserved between different organisms, we focused on the yeast system. The essential protein Rse1p is the yeast homolog of SF3b130. We show that the *rse1-1* strain is sensitive to cordycepin, which suggests that Rse1p might be involved in 3' end processing. Furthermore, Rse1p and 3' end processing factors interacted genetically and Northern blot analysis suggested that strains carrying mutations in Rse1p and subunits of CPF had increased levels of unspliced pre-mRNA at restrictive temperature compared to the single mutants. We therefore suggest that the coupling of 3' end processing and splicing mediated by CPSF and U2 snRNP is conserved between mammals and yeast.

Precursor tRNAs (pre-tRNAs) must undergo a number of processing steps before they become mature tRNAs and some tRNAs contain introns. tRNA splicing is a three step reaction and each step requires an individual set of proteins. In the first step the pre-tRNA is cut at its two splice sites. This reaction is catalyzed by the so called tRNA splicing endonuclease complex. Recently this complex was purified from mammalian cells and interestingly hClp1 (a subunit of the 3' end processing factor CF II_m) was identified as one of its components. Furthermore, hSen2 a subunit of the endonuclease complex was shown to be required for efficient 3' end processing *in vivo*. A model was proposed suggesting that the tRNA endonuclease complex is involved in 3' end processing. In collaboration with S. Paushkin and C. Trotta we continued to investigate if this model is indeed correct. We found that biochemically purified CF II_m indeed carried tRNA endonuclease activity. However, tRNA endonuclease complexes were not able to reconstitute cleavage activity of CF II_m-depleted HeLa nuclear extract, unless they were purified with His-Flag tagged

hClp1. Taken all our results into account we cannot exclude that the tRNA endonuclease complex is indeed involved in 3' end processing. However, we believe that the evidence supporting this model is rather weak. We think it is more likely that hClp1 is part of the tRNA endonuclease complex as well as a subunit of CF II_m, and that the two complexes are not functionally related.

As mentioned earlier, 3' end processing is highly conserved form mammals to yeast. In *S. cerevisiae* cleavage and polyadenylation factor (CPF) is a multiprotein complex, which together with the cleavage factor IA (CF IA) and the cleavage factor IB (CF IB) is required for both the cleavage and the polyadenylation step of the 3' end formation reaction.

Ydh1p/Cft2p is an essential component of CPF. Cleavage and polyadenylation reactions revealed that the protein is required for both reactions to occur *in vitro*. Previously, it was demonstrated that an important function of CPF lies in the recognition of poly(A) site sequences and previous RNA binding analyses with recombinant Ydh1p/Cft2p suggested that the protein may interact with the *CYC1* poly(A) site region. In accordance, we found that mutant *ydh1* strains were deficient in recognition of the *ACT1* cleavage site *in vivo*.

Transcription by RNA polymerase II (RNAP II) and 3' end processing reactions are tightly linked. The C-terminal domain (CTD) of RNAP II plays a major role in coupling the two events, as it tethers the factors involved in polyadenylation to the polymerase. We provide evidence that Ydh1p/Cft2p interacts with the CTD, several subunits of CPF and with Pcf11p, a component of CF IA. We propose that Ydh1p/Cft2p contributes to the formation of important interaction surfaces that mediate the dynamic association of CPF with RNAP II, the recognition of poly(A) site sequences and the assembly of the polyadenylation machinery on the RNA substrate.

Chapter1: Introduction to RNA processing

1.1 Pre-mRNA processing **8**

1.1.1. 3' end processing 9

1.1.1.1 Mammalinan *cis* acting elements 10

1.1.1.2 Yeast *cis* acting elements 12

1.1.1.3 Mammalina *trans* acting factors 13

1.1.1.4 Trans acting factors in yeast 18

1.1.1.5 Histone 3' end processing 24

1.1.2 Splicing 25

1.1.2.1 Splice site recognition and *cis* acting elements 25

1.1.2.2 The spliceosome 26

1.1.2.3 Splicing in yeast 30

1.2 Coupling between transcription, capping, splicing and 3' end processing **30**

1.2.1 Coupling of capping with splicing and 3' end formation 30

1.2.2 Coupling between splicing and 3' end processing 31

1.2.3 Coupling between transcription and pre-mRNA processing 31

1.3 tRNA splicing **34**

1.1 Pre-mRNA processing

The genome of eukaryotes is transcribed by three different highly related RNA polymerases (RNAP). The ribosomal RNAs (rRNAs) are transcribed by RNAP I whereas RNAP III transcribes several RNAs including the small 5S rRNA and the transfer RNAs (tRNAs). Messenger RNA precursors (pre-mRNAs) are transcribed by RNAP II.

Pre-mRNAs have to undergo a number of processing steps in the nucleus before they are exported to the cytoplasm and translated by the ribosome. The 5' end of the pre-mRNA is capped as soon as the transcript has reached a length of about 20 to 25 nucleotides (Jove and Manley, 1984; Rasmussen and Lis, 1993). A 7-methyl guanine cap is added to the 5'-triphosphate end of the nascent RNA by a series of three enzymatic reactions. The 5'-triphosphate end of the pre-mRNA is hydrolyzed to a 5'-diphosphate by an RNA triphosphatase, then capped with GMP by an RNA guanylyltransferase, and methylated by an RNA (guanine-N7) methyltransferase (reviewed in Gu and Lima, 2005).

The coding sequence of RNAs (exons) in eukaryotic cells are interrupted by non coding sequences (introns), which are removed by the spliceosome in a lariate structure and the exons are ligated together. The mammalian introns average a length of $3x10^3$ nucleotides (nt) whereas the exons are much shorter, in average less than 300 nt in length. The spliceosome consists of small nuclear ribonucleoprotein particles (snRNP), which assemble anew onto each individual intron in a stepwise fashion in order to form the catalytically active spliceosome. Alternative splicing enables eukaryotes to increase the coding potential of their genome. In humans more than 59% of the genes seem to be alternatively spliced (Reed, 2000; Hastings and Krainer, 2001; Will and Lührmann, 2001; Jurica and Moore, 2003).

All eukaryotic primary mRNAs (with the exception of replication-dependent histone transcripts in higher eukaryotes) are endonucleolytically cleaved, followed by poly(A) addition to the upstream fragment of the cleavage reaction, and the downstream fragment is rapidly degraded. *Cis* and *trans* acting factors are involved in the two steps of the 3'-end processing reaction which will be discussed in further detail in the following chapters.

The poly(A) tails, and the proteins associated with them, affect the stability, turnover, export to the cytoplasm and translation of the mRNA (reviewed in Zhao et al., 1999a; Edmonds, 2002). Figure 1.1 shows a schematic representation of the three different pre-mRNA processing steps.

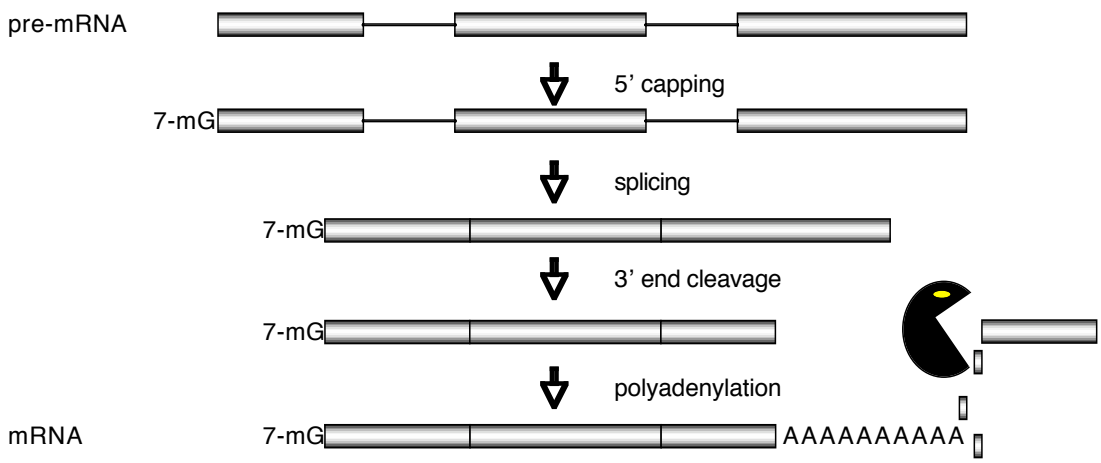


Figure 1.1 The three pre-mRNA processing steps.

The gray cylinders represent the exons and the lines the introns of the pre-mRNA. The black "head" to the right symbolizes the exonuclease.

1.1.1 3' end processing

3' end processing is highly conserved form mammalians to yeast and most of the mammalian subunits have homologues in yeast. Figure 1.2. shows a schematic representation of *cis* and *trans* acting elements required for 3' end processing in mammals and yeast. The different RNA sequence elements and the 3' end processing factors will be discussed in detail in the following paragraphs.

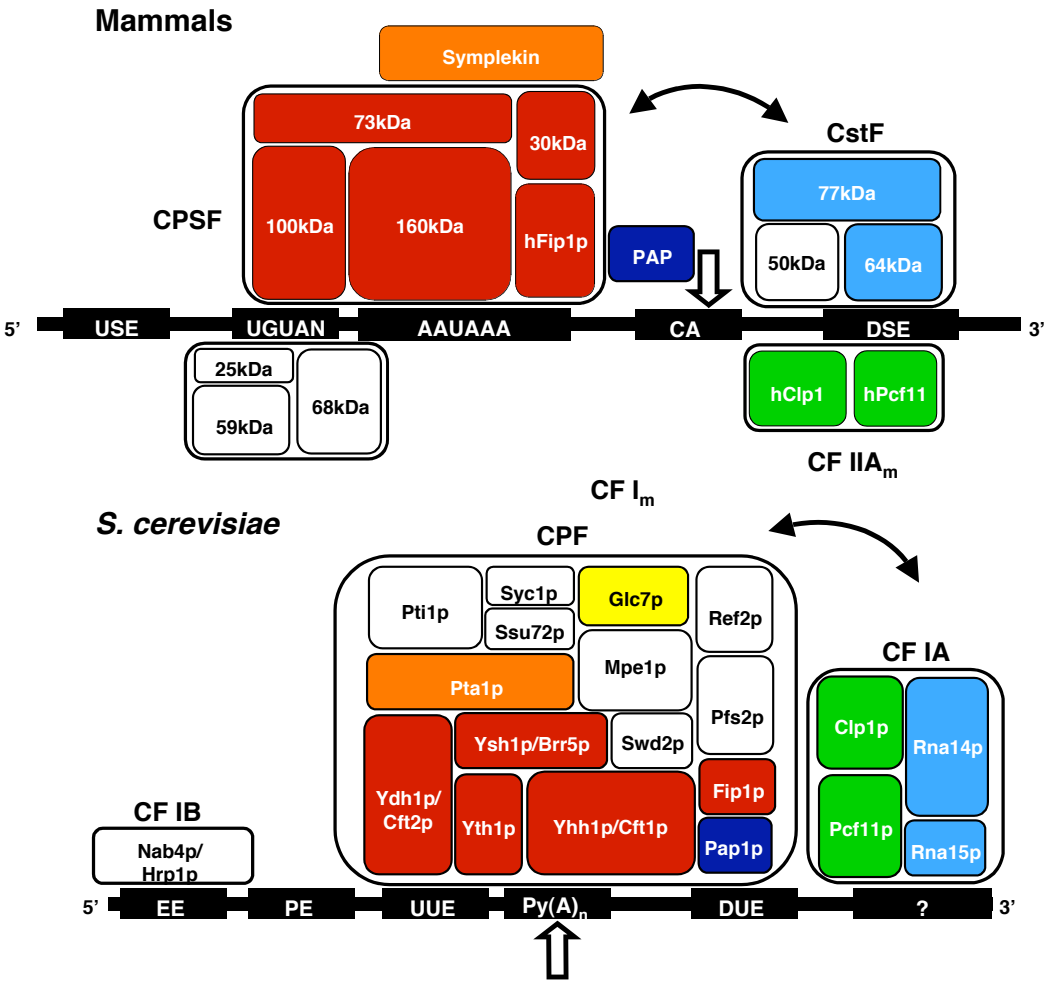


Figure 1.2 Schematic representation of the mammalian and yeast pre-mRNA cleavage complexes.

With the exception of CstF the arrangement of the subunits within the different factors is not known. Homologous subunits are shown in the same color. Subunits for which no homologues have been identified are in white. Double-headed arrows represent protein-protein contacts. Modified from a slide provided by B. Dichtl.

1.1.1.1 Mammalinan *cis* acting elements

In mammalian cells the core polyadenylation signal consists of three sequence elements, the highly conserved AAUAAA hexanucleotide located about 10 – 30

nucleotides upstream of the cleavage site, the G/U rich sequence element downstream of the cleavage site and the cleavage site itself (Figure 1.2).

The AAUAAA consensus sequence can be found in almost all polyadenylated mRNAs (Proudfoot and Brownlee, 1976; Wahle and Kuhn, 1997). *In vivo* deletion mutations confirmed that the signal is required for 3' end formation (Fitzgerald and Shenk, 1981) and *in vitro* point mutations and RNA modification experiments showed that alteration of any of the six nucleotides strongly effected 3' end processing. The only variant to the AAUAAA sequence that showed significant activity is the AUUAAA hexanucleotide (Zarkower et al., 1986; Conway and Wickens, 1987; Bardwell et al., 1991).

The downstream sequence element (DSE) is located approximately 30 nucleotides downstream of the poly(A) site. This element if poorly conserved and two different types have been described; the U-rich element is a short run of U residues (Chou et al., 1994; Gil and Proudfoot, 1987) and the G/U rich element consists of the consensus sequence YGUGUUYY (Y = pyrimidine; McLauchlan et al., 1985). In agreement with its poor conservation large deletions of the DSE are required to result in an inhibition of 3' end processing (Zarkower and Wickens, 1988). However, the distance of the DSE to the poly(A) site was shown to effect 3' end cleavage efficiency and poly(A) site selection (Mason et al., 1986; McDevitt et al., 1986; Gil and Proudfoot, 1987; MacDonald et al., 1994;)

The sequences surrounding the polyadenylation site are not conserved. In 70% of the pre-mRNAs the cleavage site contains an adenosine and the penultimate nucleotide is most often a C residue (Sheets et al., 1990). The poly(A) site selection is determined mainly by the distance between the AAUAAA and the DSE (Chen et al., 1995). Recently it was shown that CF I_m preferentially binds the sequence UGUAN (Brown and Gilmartin, 2003) and it was proposed that these sequence elements are also important for poly(A) site selection (Venkataraman et al., 2005).

U-rich upstream sequence elements (USEs) are auxiliary sequences that can modulate the efficiency of 3' end formation but are not essential for processing.

1.1.1.2 Yeast *cis* acting elements

Polyadenylation signals in yeast are not as highly conserved as in higher eukaryotes; however, common elements are found in the untranslated regions of genes, which direct the site of cleavage and polyadenylation. These sequences are the efficiency element (EE), positioning element (PE), U-rich elements located immediately up- and downstream of the cleavage site and the poly(A) site itself (Figure 1.2.)

The EE is found at variable distances upstream of the cleavage site. Computer analyses revealed that more than half of the 1000 yeast nuclear genes examined contain UAUAUA sequences in their 3' region (Graber et al., 1999), in comparison to related EE sequences, this motive seems to be most efficient for mRNA 3' end formation (Irniger and Braus, 1994). In addition, the U residues at the first and fifth positions were found to be the most crucial nucleotides in this sequence (Guo and Sherman, 1995). Other genes carry the related sequences UAUUUA, UAUGUA, UAUUUA or UUUUUAUA (Egli et al., 1995; Guo and Sherman, 1995).

The PE is often located approximately 20 nucleotides upstream of the cleavage site and consists of an A-rich sequence (Russo et al., 1991). The sequences UUAAGAAC, AAUAAA or AAAAAA are the most efficient positioning elements (Guo and Sherman, 1995). The PE contributes to the efficiency of 3' end processing and to poly(A) site selection (Abe et al., 1990; Hyman et al., 1991; Irniger et al., 1992; Duvel et al., 1999).

U-rich elements are located immediately up- and downstream of the poly(A)-site (Graber et al., 1999). The U-rich element upstream of the cleavage site consists of the sequence UUUUCU or UUUUUU. It contributes to cleavage site selection and enhances 3' end processing efficiency (Dichtl and Keller, 2001).

The poly(A) site often contains $Py(A)_n$ sequences (Py=pyrimidine; Heidmann et al., 1994). Many yeast genes contain a cluster of poly(A) sites downstream of the efficiency and positioning elements (Aranda et al., 1998).

1.1.1.4 Mammalina *trans* acting factors

Biochemical purification from HeLa cells and calf thymus identified the factors involved in 3' end processing. The cleavage and polyadenylation specificity factor (CPSF), cleavage stimulatory factor (CstF), Cleavage Factor I_m and II_m (CF I_m and CF II_m) and the poly(A) polymerase (PAP) are required for the cleavage reaction. CPSF, PAP and the nuclear poly(A) binding protein (PABPN1) are required for processive polyadenylation and for controlling the length of the poly(A) tails.

Cleavage and polyadenylation specificity factor

CPSF is required for the cleavage and the polyadenylation steps of the 3' end processing reactions. The factor was shown to bind to the highly conserved AAUAAA hexanucleotide signal and all six nucleotides are necessary for binding (Bardwell et al., 1991; Bienroth et al., 1991; Keller et al., 1991). CPSF binds only weakly to the AAUAAA signal but the binding can be enhanced by a cooperative interaction with CstF bound to the downstream signal sequence (Wilusz et al., 1990; Weiss et al., 1991; Gilmartin and Nevins, 1991; MacDonald et al., 1994). Purification of CPSF activity from HeLa cells and calf thymus lead to the identification of its five main subunits: CPSF160, CPSF100, CPSF73, CPSF30 and hFip1 (Barabino et al., 1997; Bienroth et al., 1991; Murthy and Manley, 1992; Jenny et al., 1994; Jenny and Keller, 1995; Jenny et al., 1996; Kaufmann et al., 2004).

CPSF160 was shown to bind preferentially to AAUAAA containing RNAs (Moore et al., 1988; Murthy and Manley, 1995). This is in agreement with the idea that this subunit is crucial for recognition of the conserved hexanucleotide. However, recombinant CPSF160 binds weaker to the AAUAAA signal compared to purified CPSF, suggesting that other CPSF subunits are involved in the recognition of this sequence as well (Murthy and Manley, 1995). CPSF160 was shown to interact with CstF77 and PAP which is consistent with the cooperative interactions of CPSF with CstF or PAP in forming stable complexes on the RNA precursor (Murthy and Manley, 1995).

CPSF100 is 23% identical and shares 49% similarity with CPSF73 (Jenny et al., 1996). Like CPSF73, CPSF100 shares similarity with the metallo-β-lactamase / β-CASP enzymes which cleave nucleic acids with a distinctive structure that

coordinates two zinc ions (Callebaut et al., 2002). However, unlike CPSF73, CPSF100 is predicted to be inactive in endonuclease activity as it lacks part of all the conserved amino acids that should be involved in the enzymatic function. The function of CPSF100 is largely unknown. It can be UV cross-linked to RNA (Edwalds-Gilbert and Milcarek, 1995), however RNA-binding experiments did not reveal a role of CPSF100 in RNA binding (Bernhard Dichtl, unpublished results).

CPSF73 is the candidate for the long sought endonuclease. As stated above CPSF73 is a member of a subfamily of metallo-β-lactamase / β-CASP enzymes (Callebaut et al., 2002). The ELAC proteins, also members of the metallo-β-lactamase family, share sequence similarity with CPSF73 (Aravind, 1999; Tavtigian et al., 2001) and ELAC2 was shown to be an endonuclease that cleaves 3' extensions from pre-tRNAs in mammalian cells (Takaku et al., 2003). Ryan and coworkers have recently shown that the putative active site of CPSF73 is essential for viability of yeast cells and a protein with the approximate size of CPSF73 was reported to UV cross-link to the cleavage site (Ryan et al., 2004). However, direct evidence for the catalytic activity of CPSF73 is still missing. Interestingly not all CPSF73 proteins present in a cell are associated with CPSF as determined by gel filtration columns. This could indicate additional functions of CPSF73 independent of CPSF.

The antibody directed against CPSF30 coimmunoprecipitates the other CPSF subunits and immunodepletes cleavage and polyadenyltion activity from HeLa cell nuclear extract (Barabino et al., 1997). This protein contains five zinc finger repeats and a zinc knuckle. Both sequences have been implied with binding nucleic acids. In agreement with this CPSF30 binds to RNA with a preference to poly(U) (Barabino et al., 1997). Interestingly CPSF30 has also been proposed to be the endonuclease since its *Drosophila* homologue clipper (clp) showed an endonucleolytic activity (Bai and Tolias, 1996). However, a similar enzymatic activity could not be demonstrated with recombinant CPSF30 or its yeast homologue Yth1p (M. Ohnacker, S. Barabino and W. Keller, unpublished results).

hFip1 was only recently identified as a subunit of CPSF. Antibodies directed against hFip1 co-precipitated CPSF and CstF subunits and depleted cleavage and polyadenylation activity from HeLa nuclear extract. GST pull-down experiments showed that hFip1 interacts directly with CPSF30, CstF77 and PAP. Interestingly, hFip1 binds preferentially to U-rich RNA sequences and can stimulate the

polyadenylation activity of PAP in an AAUAAA independent manner. Upon addition of PABPN1 to the polyadenylation reaction hFip1 stimulates the reaction to a similar extent as observed with CPSF, PAP and PABPN1 (Kaufmann et al., 2004).

Cleavage stimulation factor

CstF binds to the downstream element and is required for the cleavage but not the polyadenylation reaction. The factor contains three polypeptides of 77, 64 and 50 kDa (Gilmartin and Nevins, 1991; Takagaki et al., 1990)

CstF77 contains so called HAT repeats (half a TPR) which might be involved in mediating protein-protein interactions (Preker and Keller, 1998). CstF77 interacts with CstF64 and CstF50, whereby the factor arranges in a linear fashion (Takagaki and Manley, 1994). It furthermore interacts with CPSF160 and this interaction is believed to stabilize the CPSF-CstF-RNA complex (Murthy and Manley, 1995).

CstF64 contains an N-terminal RNA binding domain (RBD) which binds to the downstream element. CstF64 was originally implicated in binding to the AAUAAA signal (Moore et al., 1988; Wilusz and Shenk, 1988) However, SELEX experiments with the recombinant protein or only its RBD showed that CstF binds to GU-rich downstream elements (Beyer et al., 1997; Takagaki and Manley, 1997). CstF64 was also found to interact with symplekin, a protein suggested to be involved in mammalian 3' end processing (Takagaki and Manley, 2000; Hofmann et al., 2002). It was proposed that CstF77 and symplekin compete for the same binding site on CstF64.

CstF50 contains seven WD-40 (β-transducin) repeats, which were implicated in binding to the phosphorylated C-terminal domain of RNA polymerase II (CTD) and the BRCA1-associated protein BARD1 (McCracken et al., 1997; Kleiman and Manley, 1999; Takagaki and Manley, 2000). CstF50 also binds to itself and to CstF77 (Takagaki and Manley, 2000).

Cleavage factor I_m

The subunits of CF I_m have no known yeast homologues. The factor is composed of a small subunit of 25 kDa and a large subunits of either 57, 68 or 72 kDa (Rüegsegger et al., 1996). The 59 and 68 kDa subunits are encoded by paralogous genes, whereas the nature of the 72 kDa subunits remains unclear. The CF I_m59 and CF I_m68 kDa

proteins contain an N-terminal RNP-type RNA-binding domain (RBD), a central proline-rich domain and a C-terminal RS-like domain which is similar to that of the SR proteins involved in splicing (Graveley, 2000). The CF I_m 68/25 kDa heterodimer has been shown to be sufficient to reconstitute CF I_m activity i*n vitro* (Rüegsegger et al., 1998) and SELEX analysis has indicated that the 68/25 kDa heterodimer preferentially binds the sequence UGUAN (Brown and Gilmartin, 2003). The 68 kDa on its own does not seem to bind RNA even though it contains an RNA recognition motif. This subunit interacts with Srp20, 9G8 and hTra2β, members of the SR family of splicing factors (Dettwiler et al., 2004).

The 25 kDa subunit can bind RNA and was shown to interact with the large subunits of CF I_m, with PAP (Kim and Lee, 2001) and PABPN1 (Dettwiler et al., 2004).

Cleavage factor II_m

Upon purification of CF II_m from HeLa cells the factor separated into an essential fraction CF IIA and a stimulatory fraction CF IIB of the cleavage reaction. Mass spectrometry analysis of the CF IIA fraction revealed amongst other proteins homologues of the yeast 3' end processing subunits Clp1 and Pcf11. Accordingly, the two human subunits were called hClp1 and hPcf11 (de Vries et al., 2000).

The N-terminus of hPcf11 contains a DSI-consensus motif, which has been shown to bind to the CTD in yeast (Sadowski et al., 2003). Two zinc-finger motifs and 30 repeats of the consensus sequence (LRFDG) can also be found in the sequence of hPcf11. Depletion of hPcf11 with antibodies directed against the protein decreased the cleavage but not the polyadenylation activity of the extract. The activity could be reconstituted upon addition of recombinant hPcf11 (Isabelle Kaufmann, unpublished results).

hClp1 contains Walker A and B motifs. These motifs have been implicated to function in nucleotide binding. Indeed hClp1 can be UV crosslinked to ATP and GTP and these binding activities are dependent on metal ions (Henk de Vries, unpublished results). HeLa nuclear extract depleted of hClp1 is inactive in cleavage but not in polyadenylation confirming the role of CF II_m in the cleavage reaction (de Vries et al., 2000).

Poly(A) polymerase

PAP is the enzyme that catalyzes the poly(A) addition to the upstream cleavage product. PAP on its own is a template-independent polymerase; however, its activity is low due to its low affinity for the RNA primer. PAP has a higher activity when Mg^{2+} is substituted with Mn^{2+} (Wahle, 1991b; Wittmann and Wahle, 1997). PAP specifically polyadenylates AAUAAA-containing RNAs in the presence of CPSF, which binds to the enzyme and the RNA. Under these conditions polyadenylation also becomes more efficient (Bienroth et al., 1993).

The RNA binding domain of bovine PAP was shown to overlap with the nuclear localization signal and three aspartates essential for catalysis could be identified. The three aspartates ligate two of the three active site metals. One of these metals also contacts the adenine ring (Martin and Keller, 1996; Martin et al., 2000). The crystal structure of mammalian PAP showed that the catalytic domain at the N-terminus of the protein shares substantial structural homology with other nucleotidyl transferases. The C-terminal domain binds the RNA and folds into a compact structure. The central domain is located in between these two domains (Martin et al., 2000). The exact mechanism how PAP specifically selects for ATP is not known yet.

Nuclear poly(A) binding protein 1 (PABPN1)

PABPN1 (formerly called PAB II or PABP2) is a 33 kDa protein. It consists of an acidic N-terminus rich in glutamates, a RNP-motif, and an arginine-rich C-terminus (Wahle et al., 1993). The protein was identified based on the observation that after a certain lag-phase polyadenylation becomes AAUAAA-independent, but oligo(A)-dependent (Sheets and Wickens, 1989; Wahle, 1991a). In agreement with this, PABPN1 stimulates PAP activity and binds to the enzyme and to oligo(A). Furthermore, PABPN1 is believed to play an important role in poly(A) tail length control, probably through oligomerization on the growing poly(A) tail (Wahle, 1995; Keller et al., 2000) and may additionally contribute to export of mRNAs from the nucleus to the cytoplasm (Chen et al., 1999; Calado et al., 2000).

1.1.1.2 Trans acting factors in yeast

The pre-mRNA 3' end processing reaction in yeast requires the cleavage and polyadenylation factor IA (CF IA), cleavage and polyadenylation factor IB (CF IB), cleavage factor II (CF II), polyadenylation factor I (PF I) and the poly(A) binding protein (Pab1p; Chen and Moore, 1992; Kessler et al., 1996). It was shown that CF II and PF I form a functional unit *in vivo* (Ohnacker et al., 2000) and therefore this factor has been renamed as cleavage and polyadenylation factor (CPF). Site specific cleavage *in vitro* requires CF IA, CF IB and CF II. CF IA, CF IB, Pab1p and PF I are essential for specific polyadenylation *in vitro*. (For simplicity, subunits carrying two names will only be referred to with one of the names).

Cleavage and polyadenylation factor IA (CF IA)

CF IA consists of four subunits: Rna14p, Rna15p, Clp1p and Pcf11. The *RNA14* gene is essential and the protein is homologous to the mammalian CstF77 subunit (24% identity; Minvielle-Sebastia et al., 1991; Kessler et al., 1996). Extracts from yeast mutants carrying a temperature sensitive *rna14* or *rna15* allele show a defect in cleavage and polyadenylation (Minvielle-Sebastia et al., 1994) indicating that the affected proteins are involved in both steps of the 3' processing reaction. Rna14p interacts with Rna15p (Minvielle-Sebastia et al., 1994), Nab4p (Gross and Moore, 2001a) and Fip1p (Preker et al., 1995). Mutants in *RNA14* and *RNA15* displayed inefficient transcription elongation and genetically interacted with the THO-complex that acts at the interface between transcription and mRNA export (Luna et al., 2005).

RNA15 is an essential gene. The 38 kDa protein is the yeast homologue of CstF64 (43% identity in the RNA binding domain (RBD); (Takagaki and Manley, 1994). The RBD is located in the amino-terminal region and can be UV cross-linked to RNA (Minvielle-Sebastia et al., 1991; Takagaki and Manley, 1994; Kessler et al., 1996). The subunit is thought to bind specifically to the positioning element upon interaction with Rna14p which bridges Rna15p and Nab4p (Gross and Moore, 2001a).

Mutant yeast extracts carrying a temperature sensitive *pcf11* allele are defective in cleavage and polyadenylation. The 70 kDA protein interacts with both Rna14p and Rna15p (Amrani et al., 1997b). Pcf11 binds to the phosphorylated form of the CTD and is involved in transcription termination (Sadowski et al., 2003).

Furthermore the protein was shown to be involved in the dismantling of the RNA polymerase II elongation complex (Zhang et al., 2005). The CTD-interacting domain is located at the N-terminus of Pcf11. This domain can interact with RNA also and it was proposed that competition between the protein-CTD and the protein-RNA interaction is important for the disengagement of polyadenylation factors from RNAP II (Hollingworth et al., 2006).

The 50 kDa protein Clp1p is the yeast homologue of hClp1 (26% identity; de Vries et al., 2000). It contains a P-loop motif implicated in ATP/GTP binding (Preker et al., 1997).

Cleavage and polyadenylation factor IB (CF IB)

CF IB consists of the single 73 kDa protein Nab4p/Hrp1p (Kessler et al., 1996; Kessler et al., 1997) which is related to the A/B group of metazoan heterogeneous nuclear ribonucleoproteins (hnRNPs). Nab4p can be UV-crosslinked to RNA (Minvielle-Sebastia et al., 1998), and the efficiency element was shown to be the likely binding site of the protein (Kessler et al., 1997; Chen et al., 1998; Valentini et al., 1999).

It was shown that Nab4p is not required for the cleavage reaction but contributes to cleavage site selection by suppressing the use of alternative poly(A) sites (Minvielle-Sebastia et al., 1998). The protein is also found in the cytoplasm where it is directly involved in modulating the activity of the nonsense-mediated mRNA decay pathway (NMD; Gonzalez et al., 2000).

Cleavage and polyadenylation factor (CPF)

CPF contains all subunits of CF II and PF I and in addition Ref2p (60 kDa), Pti1p (47 kDa), Swd2 (37 kDa), Glc7p (36 kDa), Ssu72p (23 kDa) and Syc1 (21 kDa). By tagging different subunits of CPF it was shown that not all polypeptides pulled-down the entire CPF complex (Nedea et al., 2003). Therefore CPF was divided into one further sub-complex called associated with Pta1 (APT). Figure 1.3 shows an illustration of the different sub-complexes found in CPF.

Ref2p is essential for viability in some strain backgrounds (Dheur et al., 2003) and is required for normal growth but not for viability in others (Russnak et al., 1995). *REF2* encodes a nucleic acid-binding protein that stimulates the cleavage reaction,

particularly of pre-mRNAs with inefficient 3' processing signals (Russnak et al., 1995). It is furthermore involved in the length control of the poly(A) tail and for snoRNA 3' end maturation (Dheur et al., 2003; Mangus et al., 2004). Ref2p is also believed to be important for the recruitment of Glc7p and Swd2p to CPF (Nedea et al., 2003).

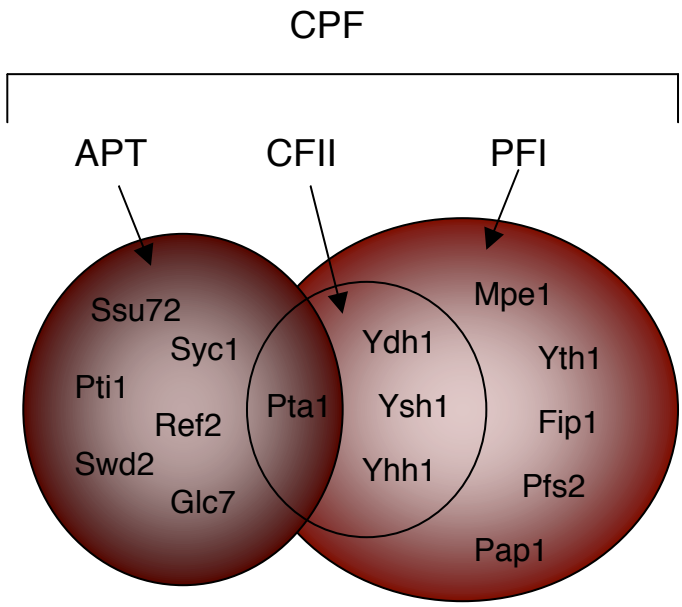


Figure 1.3 Schematic representation of the organization of CPF into different sub-complexes.

PTI1 is an essential gene that is well conserved in eukaryotes. It shares homology with CstF64 (Wolfgang Hübner unpublished results). It contains an RRM-type RNA-binding domain (RBD) and several potential phosphorylation sites. *PTI1* is a multicopy suppressor of a *pcf11* temperature-sensistive mutant. Furthermore, Pti1p plays a role in snoRNA 3' end formation (Dheur et al., 2003).

Swd2p is an essential WD repeat protein. It is a subunit of CPF as well as a subunit of the SET-complex, which modifies lysine 4 of histone H3. The protein is involved in 3' end formation of specific mRNAs and snoRNAs and *swd2* mutants were also deficient in transcription termination. Furthermore, methylations on lysine 4 of histone H3 were altered in *SWD2* mutant strains. Moreover, the role of the protein in 3' end formation and histone tail modification seams to be functionally independent (Cheng et al., 2004; Dichtl et al., 2004).

Glc7p is the catalytic subunit of type 1 protein phosphatase (PP1; Feng et al., 1991). The protein is involved in the regulation of many physiological processes. Glc7p itself has little substrate specificity (Lenssen et al., 2005). It obtains specificity by regulatory subunits that target Glc7p to different substrates (Cui et al., 2004). It is thought to be involved in the polyadenylation but not the cleavage step of 3' end processing. Depletion of Glc7p causes shortened poly(A) tails and accumulation of phosphorylated Pta1, its potential target (He and Moore, 2005). Npl3p, which is involved in mRNA export is activated by Glc7p-dependent dephosphorylation (Gilbert and Guthrie, 2004).

Ssu72p is an essential protein (Sun and Hampsey, 1996). It interacts with the general transcription factor TFIIB and RNAP II (Wu et al., 1999; Pappas and Hampsey, 2000). Ssu72p was suggested to bridge Pta1p and Ydh1p, TFIIB and RNAP II via Rpb2p. It is involved in 3' end cleavage as well as in transcription elongation and termination (Dichtl et al., 2002a; Ganem et al., 2003; Steinmetz and Brow, 2003). Ssu72p was furthermore shown to be a CTD phosphatase (Krishnamurthy et al., 2004).

SYC1 is a non-essential gene, but its removal improves the growth of other processing mutants at restrictive temperature. It is highly homologous to the C-terminal domain of Ysh1p, a part of the protein that shares no homology to the beta-lactamase family. Addition of recombinant Syc1p can restore processing activity to cleavage and polyadenylation-defective *ysh1-1* extract (Zhelkovsky et al., 2006).

Cleavage factor II

CF II has been purified by taking advantage of its ability to reconstitute the cleavage reaction in the presence of purified CF IA and CF IB (Zhao et al., 1997).

CF II consists of the four subunits Yhh1p/Cft1p (150 kDa; Stumpf and Domdey, 1996; Zhao et al., 1997), Ydh1p/Cft2p (105 kDa; Jenny et al., 1994), Ysh1p/Brr5p (100 kDa; Jenny et al., 1996) and Pta1p (90 kDa; Preker et al., 1997). The genes coding for the four subunits are essential.

Yhh1p is the yeast homologue of CPSF160 (24% identity; Jenny and Keller, 1995; Stumpf and Domdey, 1996) and interacts with multiple subunits of CPF and CF IA. Therefore, it might be important for the assembly of the processing machinery. The protein contains an RNA binding domain that is composed of β-propeller

forming repeats and interacts with poly(A) site sequences. In agreement with this fnding, Yhh1p acts in cleavage site selection and is essential for recognition of weak poly(A) sites (Dichtl et al., 2002b).

Ysh1p is the yeast homologue of CPSF73 (23% identity), over the first 500 aa it is even 53% identical to the bovine CPSF73 subunit (Chanfreau et al., 1996; Jenny et al., 1996). Ysh1p is a member of a subfamily of metallo-β-lactamase enzymes that cleaves nucleic acids (Callebaut et al., 2002). It has recently been shown that the putative active site of Ysh1p is essential for viability of yeast cells (Ryan et al., 2004) and the authors proposed that this protein is the endonuclease. Mutant yeast extracts carrying a conditional *ysh1* allele have been reported to be deficient in polyadenylation but not in cleavage (Jenny et al., 1996). This observation is not necessarily in discrepancy with the model that Ysh1p could be the endonuclease given that mutations that abolish the cleavage activity of the protein could be lethal at any temperature.

The essential protein Pta1p is involved in polyadenylation (Preker et al., 1997) and in cleavage (Zhao et al., 1999b). The phosphorylation status of Pta1p was implied to regulate polyadenylation (He and Moore, 2005).

Ydh1p is the 105 kDa subunit of CPF. It has 24.4% identity and 43% similarity with the mammalian protein CPSF100. It is also significantly related to Ysh1p and to CPSF73 (Jenny et al., 1994). Ydh1p is essential for cell viability (Preker et al., 1997). Its amino acid sequence was examined for known RNA binding motifs, but none could be found. The protein carries numerous potential serine/threonine phosphorylation sites, a possible tyrosine kinase site CYL at amino acid 32 as well as two putative single cluster type nuclear localization signals (Zhao et al., 1997). Even though no known RNA binding domain could be found, Ydh1p, in the presence of the other CF II subunits and ATP, was shown to UV-crosslink to wild-type GAL7-1 RNA substrate but not to a mutated version of the full-length precursor which lacks the $(UA)_6$ repeat upstream of the poly(A) site nor to precleaved GAL7-9 substrate lacking the sequences downstream of the cleavage site (Zhao et al., 1997). Furthermore, GST pull down experiments revealed that recombinant Ydh1p binds to full length CYC1 RNA, precleaved CYC1 lacking sequences downstream of the poly(A) site, and to CYC1-512, which has a 38nt deletion encompassing both EE and PE (Dichtl and Keller, 2001). The binding site of Ydh1p on CYC1 RNA substrate was determined by

RNase H protection experiments. The protein binds with high affinity at sequences surrounding the poly(A) site and weaker at the EE (Dichtl and Keller, 2001).

Poyadenylation factor I (PF I)

PF I was originally identified as an activity which supported poly(A) addition but not cleavage (Chen and Moore, 1992). It contains the four subunits of CF II, Yth1p (26 kDa), Mpe1p (58 kDa; Vo et al., 2001), Pfs2p (53 kDa), Fip1p (55 kDa) and Pap1p (64 kDa). All genes coding for these subunits are essential.

Yth1p is the yeast homologue of CPSF30 (40% identity; Barabino et al., 1997) and is required for both steps of 3' end processing and tethers Fip1p and the poly(A) polymerase to the rest of CPF. It binds to pre-mRNA around the poly(A) site and participates in the recognition of the cleavage site (Barabino et al., 2000).

Seven WD repeats can be found in the protein sequence of Pfs2p. This protein plays an essential role in cleavage and polyadenylation. It bridges CPF with CF IA through its interaction with Ysh1p, Fip1p and Rna14p (Ohnacker et al., 2000).

FIP1 (Factor Interacting with Poly(A) polymerase) is an essential gene. The protein can be phosphorylated by the protein kinase CK2 (Zielinski et al., 2006). *fip1* mutant extracts cleave pre-mRNA substrates *in vitro*, but are deficient in polyadenylation (Preker et al., 1995). The protein interacts both with Pap1p (with which it forms a 1:1 complex *in vitro*) and with Rna14p (Preker et al., 1995). Fip1p possibly tethers Pap1p to the RNA substrate. Furthermore, Fip1p might interact more weakly with phosphorylated than with unphosphorylated CPF, which in turn could influence polyadenylation efficiency (He and Moore, 2005).

The protein sequence of Mpe1p contains a putative RNA-binding zinc knuckle motif. The polypeptide interacts genetically with Pcf11p and is required for poly(A) site selection. Mpe1p is involved in polyadenylation (Vo et al., 2001). Controversial data exist on whether the protein plays a role in cleavage as well (Vo et al., 2001; Martin Sadowski, Dissertation 2002).

The yeast poly(A) polymerase (Pap1p) is 47% identical to mammalian PAP within the first 400 aa, a region which comprises the catalytic domain and includes the nucleotidyltransferase active site (Martin and Keller, 1996). An RNA binding site at the carboxyl terminus is believed to be essential for processive activity (Zhelkovsky et al., 1995). In addition two other contacts with the RNA substrate exist,

one is thought to recognize the last three nucleotides of the RNA primer and to help the enzyme discriminate against deoxyribonucleotide substrates, and another base-specific site is proposed to interact with the primer 12 to 14 nucleotides upstream of the 3' end (Zhelkovsky et al., 1995). At either end of Pap1 two specific domains were identified, which might be necessary to recruit Pap1 to the polyadenylation machinery by interacting with specificity factors and to regulate its activity (Zhelkovsky et al., 1995).

In contrast to the mammalian poly(A) polymerase yeast Pap1p is not required for cleavage (Mandart and Parker; 1995). The crystal structure of the protein has also been solved (Bard et al., 2000).

Poly(A) binding protein (Pab1p)

Pab1p associates with the poly(A) tails of mRNAs in the nucleus and the cytoplasm (Adam et al., 1986; Setyono and Greenberg, 1981). It was shown to limit the length of poly(A) tails as cells bearing a *pap1* mutation, conferring temperature-sensitive growth show an aberrantly long poly(A) tail *in vivo* and *in vitro* (Amrani et al., 1997a; Minvielle-Sebastia et al., 1997). Moreover, the protein is involved in the polyadenylation activity of the complex, as poly(A) tail addition in reconstituted assays only occurs in the presence of Pab1p. It also plays a crucial role in translation initiation (Tarun and Sachs, 1995; Tarun and Sachs, 1996; Tarun et al., 1997) and in deadenylation-dependent mRNA turnover (Caponigro and Parker, 1995).

1.1.1.4 Histone 3' end processing

In metazoans most histone transcripts are not polyadenylated but like the pre-mRNAs they undergo endonucleolytic cleavage. (Birchmeier et al., 1984). The *cis* elements involved in histone 3' end processing are a conserved stem-loop, positioned upstream of the cleavage site and a purine-rich histone downstream element (HDE) (for review see Marzluff and Duronio, 2002). The U7 snRNP binds through base pairing of the 5' end of the U7 snRNA to the HDE (Schaufele et al., 1986). The stem loop binding protein (SSLBP) binds to the stem loop and helps recruiting the U7 snRNP to the RNA (Dominski et al., 1999). ZFP100, a zinc finger protein

facilitates this recruitment by bridging between SLBP and Lsm11 (Dominski et al., 2002; Pillai et al., 2003). Lsm11 is a unique component of the U7-specific Sm core. The site of endonucleolytic cleavage is located 10 to 11 nucleotides upstream of the intermolecular U7/HDE RNA duplex (Scharl and Steitz, 1994; Scharl and Steitz, 1996). Recently a protein complex required for histone processing was purified that contained all five subunits of CPSF, CstF77, CstF64 and symplekin. Furthermore symplekin was identified as the heat-labile factor (Kolev and Steitz, 2005). Moreover, it was shown that CPSF73 cross-links to the upstream and downstream fragment of the cleavage products and the authors suggested that CPSF73 is both the endonuclease and 5'-3' exonuclease in histone pre-mRNA processing (Dominski et al., 2005a).

1.1.2 Splicing

The excision of the intron and subsequent ligation of the exons is achieved in two transesterification reactions. First the 2'OH of the branchpoint adenosine acts as a nucleophile attacking the 5' exon-intron border. This results in a free 5' exon and a lariate-shaped intron that is still attached to the 3' exon. In the second step, the 3' OH of the freed 5' exon attacks the intron-3' exon border. This second transesterification results in the ligation of the exons and release of the intron as a shorter lariat. The lariat is then subsequently degraded.

1.1.2.1 Splice site recognition and *cis* acting elements

In mammals the spliceosome has to recognize the relatively short exons amid much longer introns. Despite of this challenge, the *cis* acting elements in vertebrates are poorly conserved and the splice sites sequences in mammals are even less conserved than their yeast counterparts (Guthrie, 1991; Ruby and Abelson, 1991).

The most common sequences involved in splice site selection are the 5' splice site, the branchpoint sequence (BP), the polypyrimidine tract (py) and the 3' splice site. The overwhelming majority of the 5' splice site contains the consensus A<u>G</u>↓<u>GUPu</u>A<u>G</u>U (Pu=Purine). The arrow marks the exon-intron junction and the underlined positions mark the most highly conserved residues. In most introns the

5' splice site is followed thousands of basepairs downstream, by the BP and Py-tract sequences. The BP sequence in humans is not highly conserved. A common consensus sequence often found is CUPuAPy, the underlined A is highly conserved. The Py-tract is a run of pyrimidines, which can vary significantly in length and sequence composition. The average length however is eight nucleotides. The 3' splice site is often located immediately downstream of the Py-tract. The distance between the highly conserved A of the BP sequence and the 3' splice site is usually eighteen bases. However, there are many introns where the distance between the BP and the 3' splice site can be hundreds of nucleotides (Smith and Nadal-Ginard, 1989). The 3' splice site often contains the consensus sequence PyAG↓Pu. The arrow marks the exon-intron boundary and the underlined nucleotides are highly conserved; all reviewed in (Sharp and Burge, 1997; Goldstrohm et al., 2001).

Although many sequences within the mammalian transcripts match the consensus splice sites most of them are not used. It is believed that many of these pseudo-splice sites have multiple defects and are inhibited by surrounding splicing silencer sequences (Sun and Chasin, 2000). Whether a splice site is used or not is probably dependent on the arrangement of positive and negative *cis* acting elements. The SR proteins and hnRNP proteins are believed to be involved in splice site selection by binding to these regulatory elements.

The exon definition model proposes that the exons are recognized by the splicing machinery which searches for a pair of closely spaced splice sites in an exonic polarity. The exon is then defined by the binding of the U1 and U2 snRNP and associated splicing factors to the RNA and their interaction with each other (reviewed in Berget, 1995).

1.1.2.2 The spliceosome

The spliecosome consists of several small nuclear ribonucleoproteins (snRNPs) and numerous non-snRNP splicing factors. Each snRNP consists of a snRNA, seven Sm or Sm-like proteins and several particle-specific proteins (Will and Lührmann, 1997). The five spliceosomal snRNAs U1, U2, U4, U5 and U6 exist as individual RNA-protein complexes; however, the sequences of U4 and U6 are quite complementary

and these snRNAs are mainly found base-paired together in a U4/U6 di-snRNP (Forne et al., 1996).

The precursors of U1, U2, U4 and U5 snRNAs are transcribed by RNAP II in the nucleus. The primary transcripts are then transported to the cytoplasm, where they are bound by the core proteins and a trimethyl guanosine cap is added to their 5' ends. These core snRNPs are re-imported to the nucleus, where the snRNAs are internally modified at several positions, primarily by pseudouridylation and 2'-O-methylation. Finally the snRNP-specific proteins bind to the core snRNP. The U6 snRNA is transcribed by RNAP III and has a γ-monomethyl phosphate cap. The U6 snRNA does not contain an Sm site and formation of the U6 snRNP involves the association of seven so-called Sm-like proteins. Assembly of the U6 snRNP is believed to take place entirely in the nucleus (Kambach et al., 1999; Will and Lührmann, 2001).

In addition to the snRNPs many proteins involved in splicing are members of the serine/arginine-rich (SR) protein family. At their N-terminus SR proteins contain one or more RRMs. The C-terminus consists of an arginine (R) and serine (S) rich domain, the so called RS domain, which is essential for splicing *in vitro* and might have a role in nuclear localisation. The serines can be phosphorylated which might alter the RNA or protein binding specificity (Will and Lührmann 2001).

The spliceosome assembles on the pre-mRNA in a stepwise manner. In an ATP-independent step the U1 snRNP recognizes and binds the 5' splice site. This interaction is mediated by the U1 snRNA base-pairing with the pre-mRNA as well as protein-RNA interactions. The U2AF proteins bind to the Py-tract and SF1 to the BP sequence. Binding of SF1 to the BP sequence is weak but its affinity is significantly increased by a simultaneous interaction with U2AF65 (Will and Lührmann, 1997; Selenko et al., 2003). SR proteins promote E-complex formation and apparently mediate a functional interaction between the 5' and 3' splice sites, forming a protein link between the U1 snRNP and U2AF65. The U2 snRNP is also associated with this so-called E-complex by protein-protein but not by RNA-protein interactions (Das et al., 2000).

The A-complex is formed in an ATP-dependent step in which the U2 snRNP is bound to the branch site, displacing SF1. This interaction is facilitated by snRNA-pre-mRNA and protein-RNA interactions, involving SF3a and SF3b subunits, as well as the splicing factors U2AF and SF1. The U2 snRNA base pairs with the BP

sequence and in doing so buldges out the branchpoint adenosine and enhances the first nucleophilic attack. U2AF and the U1 snRNP are thought to be less tightly bound in the A-complex than in the E-complex (reviewed in Will and Lührmann, 2001).

According to the currently accepted model the B-complex is formed upon addition of the U4/U6-U5 tri-snRNP. However, recent data indicated that the tri-snRNP recognized the 5' splice site together with U1 snRNP at the earliest stages of spliceosome assembly (Maroney et al., 2000). The U1 snRNP-5' splice site interaction is weakened or disrupted and often the U1 snRNP is missing from this complex.

The complex B undergoes many conformational changes leading to the formation of an intricate RNA-RNA network within the spliceosome. At this point the U2 snRNA, U6 snRNA and the pre-mRNA seem to interact with each other and form a catalytic core. By doing so they place the 5' splice site and branchpoint in close proximity. This complex is catalytically active and is reffered to as complex C (Madhani and Guthrie, 1992; Staley and Guthrie, 1998; Will and Lührmann, 2001). Figure 1.4 shows a schematic representation of the different spliceosome assembly steps.

In addition to the finding mentioned above that the tri-snRP was found already in the E-complex, preparation of cell-free extracts from *S. cerevisiae* led to the discovery of a high molecular-mass spliceosome complex, containing all five snRNAs and more than 85% of all splicing factors identified, that apparently assembled functionally on a pre-mRNA as a whole (Stevens et al., 2002). This favors a model in which a pre-assembled splicing machinery recognizes the pre-mRNA and then rearranges into a catalytically active form. However, it has recently been reported that cotranscriptional spliceosome assembly occurs in a step wise fashion (Gornemann et al., 2005; Tardiff and Rosbash, 2006).

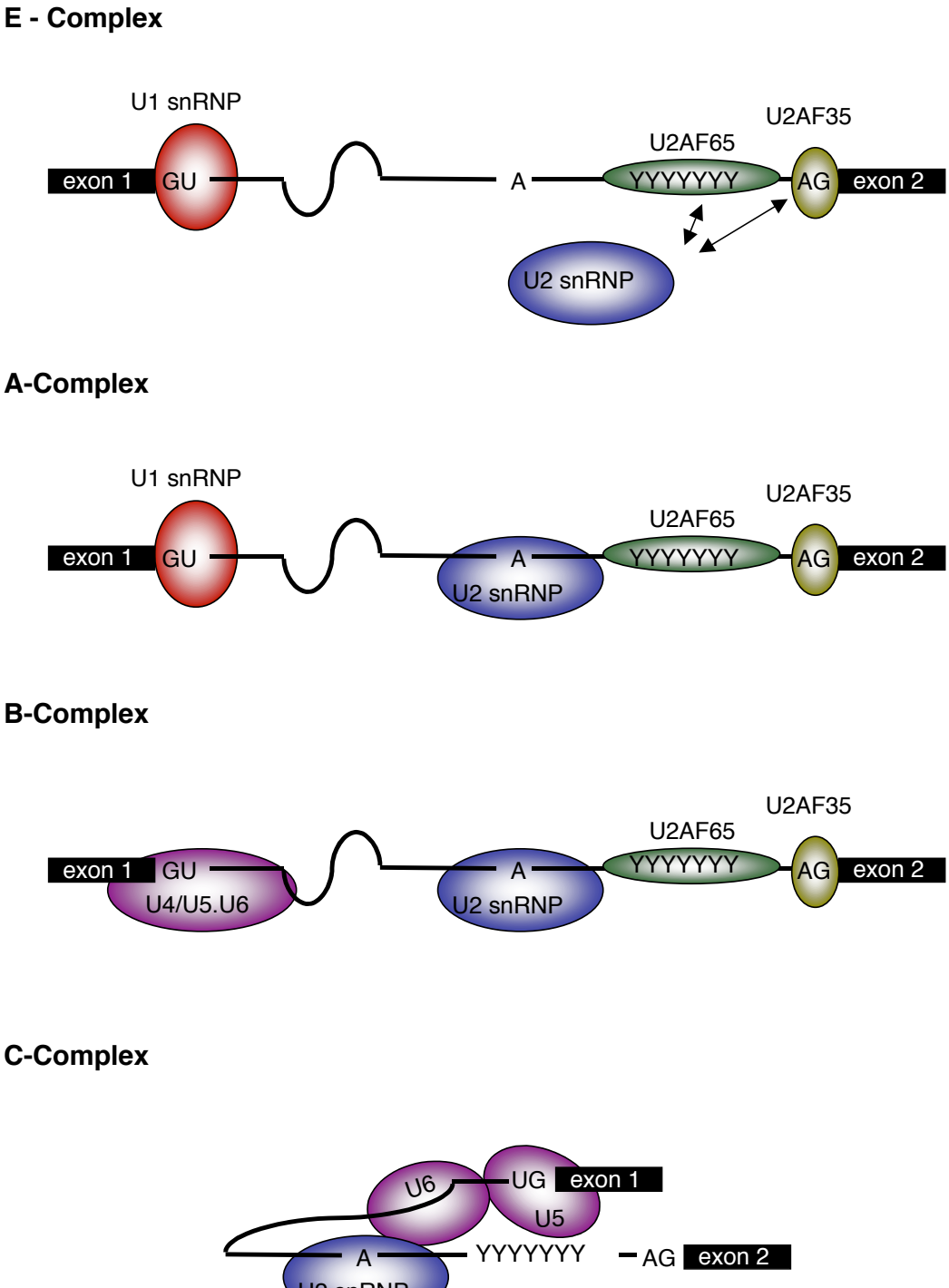


Figure 1.4 Schematic representation of the different steps in spliceosome assembly.

1.1.2.3 Splicing in yeast

A mammalian nucleus contains 10^5 – 10^6 snRNPs compared to about 100 found in a yeast cell. In *S. cerevisiae*, only 5% of the genes contain introns. Usually they have only one single intron located near the 5' end of the gene (Lopez and Seraphin, 1999; Spingola et al., 1999).

In yeast the *cis* acting RNA elements are more conserved than in humans. According to the more stringent splice site selection no alternative splicing has been found in these organisms yet. The 5' splice site contains the sequence GUAUGU. The branchpoint sequence consists of the highly conserved nucleotides UACUAAC; the underlined A marks the branchpoint andenosine. The 3' splice site consensus sequence is PyAG. In most genes a Py-tract between the BP and the 3' splice site can be found (Kuhn and Kaufer, 2003).

The splicing apparatus, spliceosome assembly and mechanism are strongly conserved from higher eukaryotes to yeast. However there are variations that probably reflect functional distinctions such as alternative splicing.

1.2 Coupling between transcription, capping, splicing and 3' end processing

Transcription, capping, splicing and 3' end processing can be studied separately from each other *in vitro*. However, it is now clear that all these reactions are coupled and influence one another *in vivo*.

1.2.1 Coupling of capping with splicing and 3' end formation

That 5' end capping has an influence on splicing became clear when it was shown, that in an *in vitro* system the addition of a cap structure to the pre-mRNA can enhance splicing. The cap was shown to have a bigger influence on the first than on the internal introns. It was found that the cap binding complex (CBC) is involved in mediating the coupling of the two reactions, when depletion of the 80 kDa subunit of

the CBC resulted in inhibition of splicing. Furthermore, the interaction of the U1 snRNP with the 5' splice site is enhanced by the CBC (Izaurralde et al., 1994; Lewis et al., 1996; Lewis and Izaurralde, 1997). The CBC was also suggested to facilitate the U6 snRNA 5' splice site interaction possibly by affecting the displacement of the U1 snRNP (O'Mullane and Eperon, 1998). Furthermore the coupling of capping and splicing was also shown in yeast, suggesting that these interconnections are conserved in evolution (reviewed in Shatkin and Manley, 2000).

The cap was also demonstrated to have a positive effect on 3' end processing. The presence of the CBC enhanced the 3' end cleavage but not the polyadenylation reaction and similar to splicing, the CBC appeared to have a stabilizing effect on the polyadenylation complex and its target pre-mRNA (Flaherty et al., 1997; Lewis and Izaurralde, 1997).

1.2.2 Coupling between splicing and 3' end processing

See introduction to Chapter 2

1.2.3 Coupling between transcription and pre-mRNA processing

All three processing reactions are taking place co-transcriptionally and there is strong evidence that the processing reactions are coupled to transcription. The connection between transcription and pre-mRNA processing starts as early as at the promoter. Specific types of promoters seam to recruit specific processing activities that determine how the pre-mRNA is processed. It was shown that capping, splicing and 3' end formation were prevented or partially repressed when a RNAP II promoter was replaced by RNAP I or RNAP III promoters. Similarly, exchanging the promoter of a protein encoding gene with a promoter of an snRNA gene impaired pre-mRNA processing (Proudfoot et al., 2002).

The main player in coupling transcription to pre-mRNA processing events is the C-terminal domain (CTD) of RNAP II. The CTD is separated from the main body of RNAP II and is attached to it by a relatively unstructured linker sequence (Cramer et al., 2001). The CTD consists of a heptad amino acid sequence that is repeated 52 times in mammals and 26 times in yeast. The repeat contains two serine residues (ser2 and ser5) and they show different phosphorylation states depending on the progress of

the transcription reaction (Dahmus, 1996). At the initiation of transcription the CTD is not phosphorylated and the RNAP II interacts with a number of general transcription factors. Upon transcription elongation however, the CTD is phosphorylated, generating the elongation-competent RNAP II (Reines et al., 1996; Price, 2000). Ser5 of the heptad repeat is phosphorylated mainly at early stages of elongation whereas phosphorylated ser2 is found at later stages (Komarnitsky et al., 2000). Proteins of all three processing reactions are interacting with the phosphorylated CTD and travel along with the elongating RNAP II.

The capping enzymes associate with the CTD at the beginning of transcription but are not found any more as the elongation progresses further into the gene (Komarnitsky et al., 2000). It was also shown that the phosphorylated CTD can enhance the capping reaction (Ho et al., 1999). Interestingly, the polymerase pauses transcription at around the time when the capping reaction occurs. It is therefore possible that this pausing occurs to allow time for 5' capping, and that the RNAP II will not proceed elongation until the protective modification has been added (Woychik and Hampsey, 2002).

It is an established fact now that splicing occurs co-transcriptionally (Beyer and Osheim, 1988; Misteli et al., 1998; Misteli and Spector, 1998; Gornemann et al., 2005; Das et al., 2006). It was shown that the phosphorylated form of the CTD can enhance splicing efficiency (Corden and Patturajan, 1997; Hirose et al., 1999) and exon definition might even require the RNAP II given that phospho-CTD only activated RNA substrates with complete exons bordered by introns (Zeng and Berget, 2000). A series of protein factors sharing homology to SR proteins and a subunit of the U1 snRNP are known to directly interact with the phosphorylated CTD (Corden and Patturajan, 1997; Morris and Greenleaf, 2000). Given that splicing occurs co-transcriptionally the 5' splice site appears before the 3' splice site on the nascent RNA. It was proposed that the 5' splice site might pair with the first 3' splice site to appear or if more 3' splice sites appear in short succession more than one choice may be possible. Also, it was shown that the elongation rate of RNAP II can influence the 3' splice site choice. If RNAP II transcribes more slowly the more proximal site is favored over the more distal 3' splice site (Howe et al., 2003; de la Mata et al., 2003). These results suggest that elongation rate and the lag time between the appearance of different splice sites can modulate alternative splicing. It was also recently suggested

that the exons are tethered to the transcribing RNAP II, facilitating splice site selection (Dye et al., 2006).

In mammals the CTD was proposed to facilitate efficient 3' end cleavage also in the absence of transcription (Hirose and Manley, 1998; Ryan et al., 2002). The phosphorylated CTD was shown to interact with the 3' end formation proteins and surprisingly the 3' end processing factors were found associated with the transcribing RNAP II from the 5' end till the 3' end of the genes (Gall et al., 1999; Licatalosi et al., 2002; Kim et al., 2004). There is also a network of interactions between 3' end processing subunits and general transcription factors localized at the promoters. CPSF subunits were shown to interact with the transcription factor TFIID (Dantonel et al., 1997). The transcriptional coactivator PC4/Sub1, which can act to inhibit CTD phosphorylation by CDKs (Schang et al., 2000), interacts with CstF64 and its yeast homologue Rna15 (Calvo and Manley, 2001). Furthermore, Ssu72, a subunit of CPF interacts with the general transcription factor TFIIB and RNAP II (Wu et al., 1999; Pappas and Hampsey, 2000), Ssu72 was also shown to be involved in transcription elongation and termination, as well as in 3' end cleavage (Dichtl et al., 2002a; Ganem et al., 2003; Steinmetz and Brow, 2003). Ssu72p was furthermore shown to be a CTD phophatase (Krishnamurthy et al., 2004).

Transcription termination and 3' end processing are strongly coupled. First evidence for this coupling emerged when it was shown that termination requires a correct poly(A) site (Zaret and Sherman, 1982; Whitelaw and Proudfoot, 1986; Logan et al., 1987). Furthermore, yeast strains carrying mutations in the CF I subunits Pcf11p, Rna14p and Rna15p (Birse et al., 1998; Sadowski et al., 2003) as well as in the CPF subunits Yhh1p, Ssu72p and Swd2p were shown to be deficient in transcription termination (Dichtl et al., 2002a; Dichtl et al., 2002b; Ganem et al., 2003; Steinmetz and Brow, 2003; Cheng et al., 2004; Dichtl et al., 2004). Pcf11p appears to play a very interesting part in transcription termination, as it was proposed to be involved in the dismantling of the RNA polymerase II elongation complex (Zhang et al., 2005; Hollingworth et al., 2006). Recently RNAP II pause sites were also suggested to be involved in transcription termination (Gromak et al., 2006). However, it is not clear at present what exactly triggers transcription termination. Two main models exist, the so called "torpedo" model and the "antiterminator" model. The torpedo model suggests that after 3' end cleavage the exonucleases

degrading the downstream cleavage product eventually catch up with the still transcribing polymerase and that this event subsequently triggers transcription termination. The antitermiator model proposes that the 3' end processing factors induce conformational changes within RNAP II upon recognition of the poly(A) site, which leads to termination (Figure 1.5; reviewed in Hirose and Manley, 2000; Proudfoot et al., 2002) Recently a combination of both models has also been proposed (Luo et al., 2006).

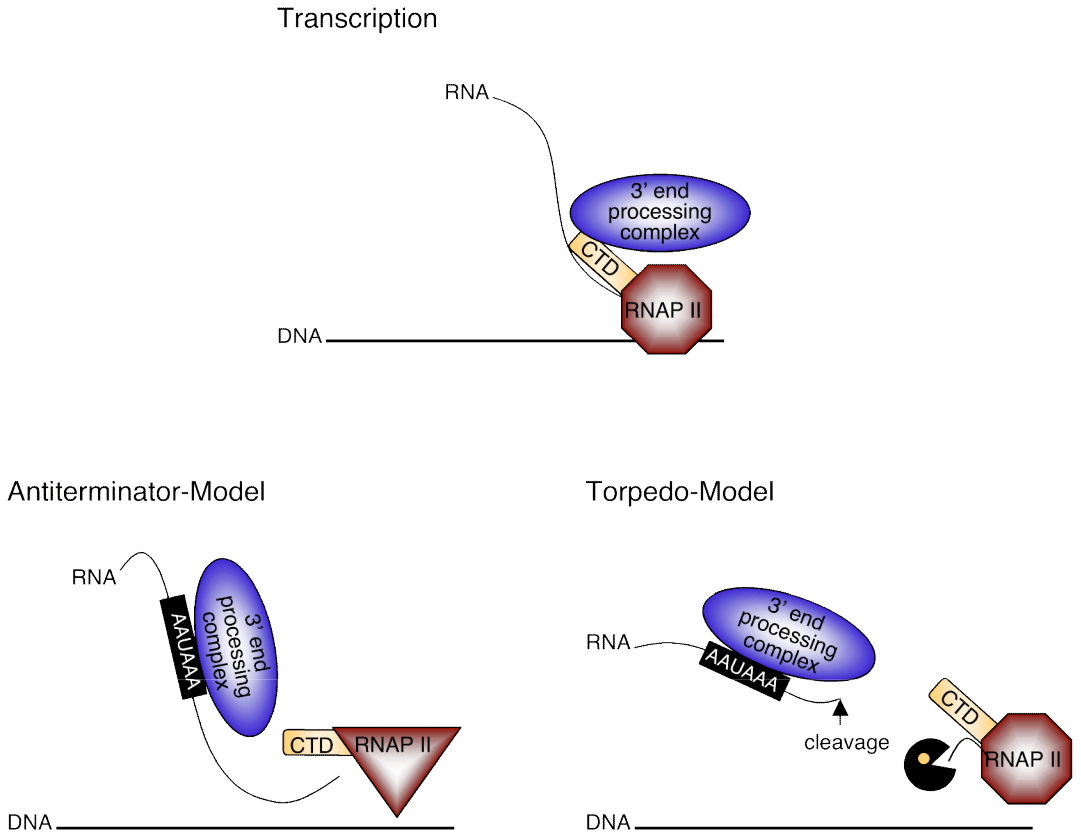


Figure 1.5 Schematic representation of the Antiterminator and the Torpedo model. The drawing at the top shows the transcribing RNAP II.

1.3 tRNA splicing

In order to generate mature tRNAs the precursor molecules must go through several processing steps. The 5' leader is removed by the RNase P (Frank and Pace, 1998; Xiao et al., 2002) and the 3' trailer is cut off by the ELAC2 enzyme (Takaku et al., 2003) and a CCA trinucleotide is added to the cleaved 3' end. Furthermore, numerous

nucleotides of the pre-tRNA are modified (reviewed in Hopper and Phizicky, 2003). In addition, some pre-tRNAs contain introns, which need to be spliced out (reviewed in Abelson et al., 1998).

In humans, only 6% and in lower eukaryotes approximately 25% of all tRNA genes contain introns (Lowe and Eddy, 1997; Trotta et al., 1997). The introns are 14 to 60 nucleotides long and interrupt the anticodon loop immediately 3' of the anticodon (Ogden et al., 1984). The 3' splice site is invariably located in a bulged loop but otherwise there is no sequence conservation at the splice junction (Baldi et al., 1992).

tRNA splicing is a three-step reaction and each step requires an individual set of proteins. In the first step the pre-tRNA is cut at its two splice sites, leading to two tRNA half molecules and the linear intron. The 90 kDa tRNA ligase subsequently ligates the two tRNA half molecules (Phizicky et al., 1986; Westaway et al., 1988). After the ligation reaction a 2'-phosphate remains at the spliced junction and must be removed to complete the splicing reaction (McCraith and Phizicky, 1990; McCraith and Phizicky, 1991; Culver et al., 1997).

The yeast tRNA splicing endonuclease contains four essential subunits: Sen2, Sen34, Sen54 and Sen15 (Rauhut et al., 1990; Trotta et al., 1997). The archeal and yeast Sen2 and Sen34 proteins are highly conserved, indicating an enzymatic role for both proteins. Indeed, yeast strains carrying mutations in Sen2p resulted in deficient cleavage of the 5' splice site (Ho et al., 1990) and mutations in Sen34p did not cleave the 3'splice site (Trotta et al., 1997). Sen2p interacts with Sen54p and Sen34p interacts with Sen15p (Trotta et al., 1997). The Sen54p and Sen15p subunits carry at their C-terminus a domain required for the dimerization of the two yeast heterodimers Sen45p-Sen2 and Sen15p-Sen34p (Lykke-Andersen and Garrett, 1997; Li et al., 1998). In yeast tRNA splicing is believed to take place in the cytoplasm (Yoshihisa et al., 2003: Huh et al., 2003), whereas in humans the reaction is likely occurring in the nucleus (Paushkin et al., 2004). Nethertheless the tRNA endonuclease subunits are conserved from yeast to humans (Paushkin et al., 2004).

Chapter 2: Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing

(This chapter is a manuscript that will appear in Molecular Cell)

2.1 Summary **38**

2.2 Introduction **38**

2.3 Results **41**

2.3.1 Subunits of the U2 snRNP are associated with CPSF 41

2.3.2 Subunits of SF3b directly interact with components of CPSF 43

2.3.3 Depletion of CPSF100 inhibits splicing in a coupled splicing and 3' end cleavage assay 45

2.3.4 Mutations in the U2 snRNA binding site inhibit 3' end cleavage in a coupled splicing and 3' end cleavage assay 48

2.3.5 Complexes involved in the coupling of 3' end processing and splicing do not form efficiently on mutant RNA substrates deficient in splicing and 3' end cleavage 51

2.3.6 Digestion of the U2 snRNA results in a decrease of 3' end cleavage activity in coupled assays 52

2.4. Discussion **56**

2.5. Experimental Procedures **59**

2.5.1 Purification of cleavage and polyadenylation specificity factor from calf thymus and MALDI-TOF MS analysis 59

2.5.2 Northern analysis 60

2.5.3 Co-immunoprecipitations and immunodepletions of HeLa cell nuclear extract 60

2.5.4 Expression of recombinant proteins in *Escherichia coli* 60

2.5.5 GST-pulldown experiments 60

2.5.6 Preparation of RNA substrates 61
2.5.7 Polyadenylation assay 61
2.5.8 Coupled cleavage and splicing assay 61
2.5.9 Complex formation analysis 61
2.5.10 Oligodeoxynucleotide-directed RNase H cleavage 62
2.5.11 Plasmids 62

2.6 Supplemental material 63
2.6.1 Results and discussion 63
2.6.1.1 Mutations in AAUAAA 63
2.6.1.2 Deletion of the U1 and U4 snRNA resulted in impaired 3' end cleavage 65
2.6.1.3 Two-hybrid interactions 68
2.6.2 Experimental procedures 76
2.6.2.1 Mutant AAUAAA RNA substrates 76
2.6.2.2 Oligodeoxynucleotide-directed RNase H cleavage 76
2.6.2.3 Plasmids 76

2.1 Summary

Eukaryotic pre-mRNAs are capped at their 5' ends, polyadenylated at their 3' ends and spliced before being exported from the nucleus to the cytoplasm. Although the three processing reactions can be studied separately *in vitro*, they are coupled *in vivo*. We identified subunits of the U2 snRNP in highly purified CPSF and showed that the two complexes physically interact. We therefore tested whether this interaction contributes to the coupling of 3' end processing and splicing. We found that CPSF is necessary for efficient splicing activity in coupled assays and that mutations in the pre-mRNA binding site of the U2 snRNP resulted in impaired splicing and in much reduced cleavage efficiency. Moreover, we showed that efficient cleavage required the presence of the U2 snRNA in coupled assays. We therefore propose that the interaction between CPSF and the U2 snRNP contributes to the coupling of splicing and 3' end formation.

2.2 Introduction

Eukaryotic messenger RNA precursors (pre-mRNAs) are capped at the 5' end, spliced, and polyadenylated at the 3' end before they are exported from the nucleus to the cytoplasm as mature messenger RNA (mRNA).

3' end processing is a two-step reaction. First, the pre-mRNA is endonucleolytically cleaved, followed by poly(A) addition to the upstream fragment of the cleavage reaction, and the downstream fragment is rapidly degraded. *Cis* elements on the pre-mRNA direct the factors involved in 3' end processing to the correct cleavage site. The highly conserved AAUAAA hexanucleotide, located about 10-30 nucleotides upstream of the cleavage site, is bound by the cleavage and polyadenylation specificity factor (CPSF; Keller et al., 1991). The cleavage stimulation factor (CstF) binds the G/U-rich sequence element downstream of the cleavage site (MacDonald et al., 1994). In addition, the 3' end processing reaction requires cleavage factor I_m (CF I_m), which binds to UGUAN sequences ((Brown and Gilmartin, 2003; Venkataraman et al., 2005), cleavage factor II_m (CF II_m; (de Vries et

al., 2000), poly(A) polymerase (PAP), and nuclear poly(A)-binding protein 1 (PABPN1; reviewed in Zhao et al., 1999a; Edmonds, 2002).

Purified CPSF contains five polypeptides: CPSF160, CPSF100, CPSF73, CPSF30 and hFip1 (Bienroth et al., 1991; Murthy and Manley, 1992; Jenny et al., 1994; Jenny and Keller, 1995; Jenny et al., 1996; Barabino et al., 1997; Kaufmann et al., 2004). CPSF is required for both steps of the 3' end processing reaction and plays a crucial role in poly(A) site recognition by binding to the AAUAAA hexanucleotide (Keller et al., 1991) and also to upstream sequence elements (USEs; Gilmartin et al., 199); Graveley et al., 1996; Brackenridge and Proudfoot, 2000).

Splicing involves the excision of introns and the subsequent ligation of the exons by the spliceosome, which is composed of the U1, U2, U4, U5 and U6 small ribonucleoprotein particles (snRNPs) and many non-snRNP proteins (reviewed in Will and Lührmann, 2001; Hastings and Krainer, 2001). Binding of U1 snRNP to the 5' splice site initiates spliceosome assembly by forming the E-complex. The E-complex also contains the U2 snRNP, which at this stage associates via protein-protein interactions (Das et al., 2000). Formation of the A-complex is achieved upon base pairing of the U2 snRNA with the branch site of the pre-mRNA. The B-complex is generated upon the addition of the U4/U6.U5 tri-snRNP and this complex undergoes an additional conformational change that leads to the catalytically active C-complex (reviewed in (Will and Lührmann, 2001; Hastings and Krainer, 2001).

The 12S U2 snRNP consists of the seven Sm proteins common to the spliceosomal snRNPs and the U2-specific proteins U2-A' and U2-B''(Will and Lührmann, 2001). Binding of the splicing factor 3b (SF3b) to the 12S U2 snRNP results in formation of the 15S particle. The active 17S U2 snRNP is formed after subsequent binding of the splicing factor 3a (SF3a; Brosi et al., 1993; Kramer et al., 1999). Subunits of SF3a and SF3b contribute to branch site recognition by interacting with the branch site adenosine and surrounding sequences. Both SF3a and SF3b are thought to play an essential role during pre-spliceosome assembly (Gozani et al., 1996; Query et al., 1996; Gozani et al., 1998; Will et al., 2001).

Transcription, capping, splicing and 3' end processing can be studied separately *in vitro*; however, in the last years it became apparent that all these processes are coupled events *in vivo*. For example, transcription itself has been linked

to all three pre-mRNA processing reactions and the CTD of the RNA polymerase II (RNAP II) seems to play a critical role in mediating this coupling by serving as a protein assembly platform. Furthermore, all three pre-mRNA processing reactions were shown to be coupled to each other (reviewed in Proudfoot et al., 2002; Zorio and Bentley, 2004).

Evidence that 3' end formation and splicing influence each other emerged when it was shown that mutations in *cis* acting elements involved in splicing affected pre-mRNA splicing and also 3' end processing. For example, mutations of the 3' splice site or the pyrimidine tract decreased splicing efficiency and 3' end processing activity *in vitro* (Cooke and Alwine, 1996; Vagner et al., 2000). Similarly, mutations in the highly conserved polyadenylation signal AAUAAA and downstream sequences decreased polyadenylation activity as well as splicing efficiency (Niwa and Berget, 1991; Cooke et al., 1999). It was proposed that splicing near the 3' end of a transcript is prevented until the RNA is released from the site of transcription by poly(A) site cleavage (Bird et al., 2005). In addition, *in vivo* studies confirmed that introns affect 3' end formation and that polyadenylation may play a role in splice site selection (Chiou et al., 1991; Nesic et al., 1993; Nesic and Maquat, 1994; Scott and Imperiale, 1996; Tsai et al., 1998). The mechanism by which this coupling takes place is not well understood; however a number of studies suggested the involvement of the U1 snRNP in coupling splicing and 3' end formation. Crosslinking of the U1 snRNP to cleavage and polyadenylation efficiency elements correlated with 3' end processing efficiency (Wassarman and Steitz, 1993). In agreement with this study, the U1A protein of the U1 snRNP was shown to stimulate polyadenylation via its interaction with CPSF160 (Lutz et al., 1996). On the other hand, the U1A protein also inhibits polyadenylation of its own pre-mRNA by interacting with PAP (Gunderson et al., 1994); similarly, the U1 70 protein of the U1 snRNP was shown to inhibit polyadenylation by interacting with PAP (Gunderson et al., 1998). Furthermore it has been reported that the U1 snRNP interacts with CF I_m (Awasthi and Alwine, 2003).

Here we present experiments that show an interaction between CPSF and the U2 snRNP. We found that proteins of the U2 snRNP are associated with CPSF and we demonstrate that subunits of SF3b directly interact with subunits of CPSF. We also provide evidence that both the U2 snRNP and CPSF are involved in coupling

splicing and 3' end processing *in vitro*. We propose that their direct interaction mediates this coupling.

2.3 Results

2.3.1 Subunits of the U2 snRNP are associated with CPSF

To identify proteins associated with CPSF we purified the factor by chromatography on five columns. CPSF-containing fractions were identified by their polyadenylation activity and Western blot and pooled after each column. We identified the proteins by mass spectrometry and found that the complex contained all the previously identified CPSF subunits: CPSF160, CPSF 100, CPSF73, hFip1 and CPSF30 (Figure 2.1.A). Surprisingly, we also detected several subunits of SF3b, namely SF3b155, SF3b130 and SF3b49. Western blot analysis revealed the presence of several other polypeptides of the U2 snRNP: SF3b14/14a, SF3a120, SF3a60 and the U2-specific protein B'' (results not shown). We did not find SF3b125 by Western blotting. In addition, the CPSF-containing fractions contained U2 snRNA as shown by Northern blotting (Figure 2.1.C, lowest panel, load).

The following proteins were also present in our purified fractions: fibronectin, SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily E member 1 (SMARCE1), SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily B member 1 (SMARCB1), actin beta and pyrroline 5 carboxylate reductase. At present we do not know if these proteins are truly associated with CPSF.

To test whether polypeptides of the U2 snRNP are stably associated with CPSF, we fractionated the purified factor on a gel filtration column, which separates molecules according to their sizes. The column was calibrated with proteins of known molecular weight as indicated at the top of Figure 1B. We assayed the individual fractions for their polyadenylation activity in reconstitution assays (Figure 2.1.B) and checked the elution profile of the proteins by Western blotting (Figure 2.1.C). The polyadenylation activity peaked between fractions 57 and 69, correlating with the elution profile of CPSF100. SF3b130, SF3a60 and the U2 snRNA were found in the

same fractions as the CPSF100 polypeptide, indicating that they are stably associated with CPSF. In contrast fibronectin eluted between fractions 63 to 68.

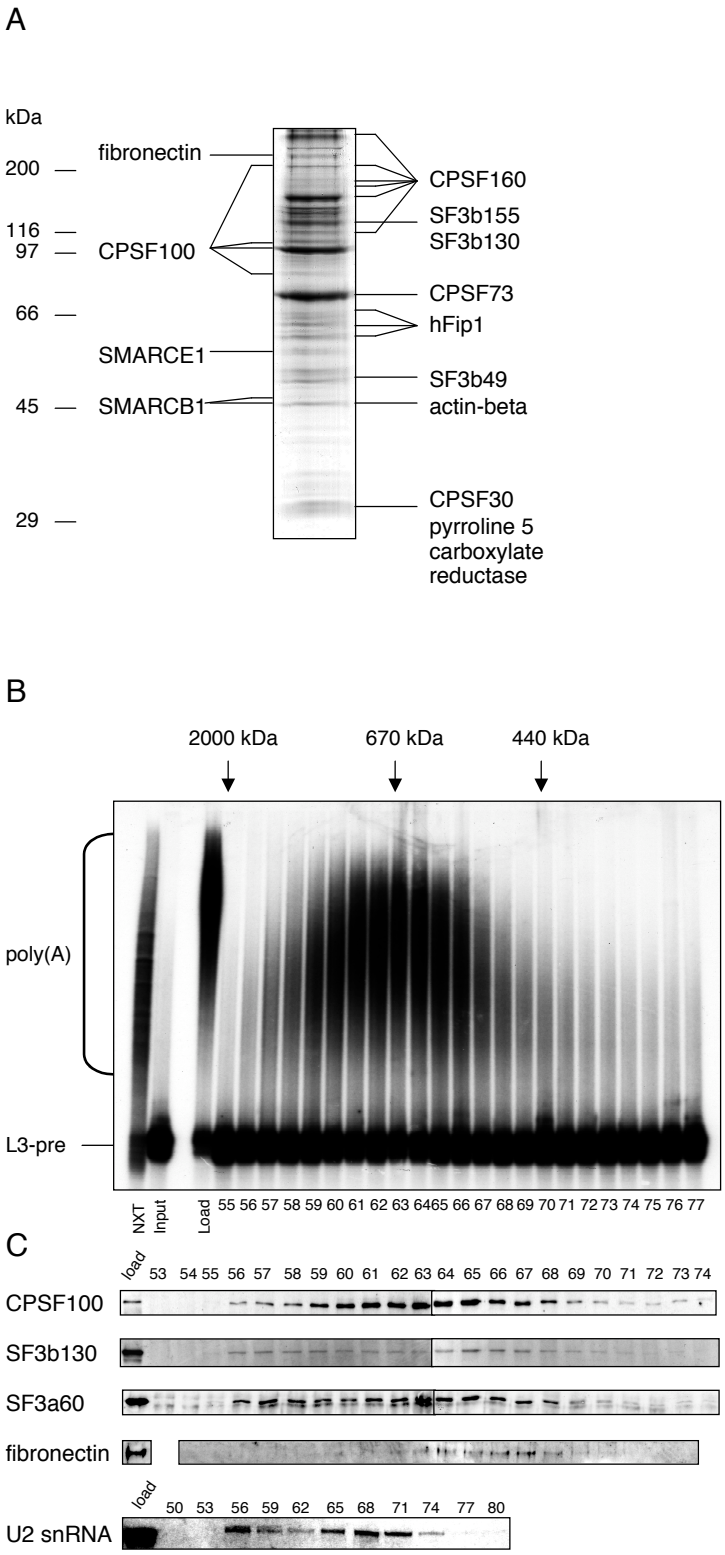


Figure 2.1 Subunits of the U2 snRNP co-purifiy with CPSF

A) Polypeptide composition of CPSF that was purified on DEAE-Sepharose followed by Blue-Sepharose, Heparin-Sepharose, Mono-Q and Poly(A) Sepharose columns. The proteins were separated on a 10% SDS gel, stained with colloidal blue and analyzed by mass-spectrometry. The migration of the molecular weight markers in kilodaltons are indicated.
B) Separation of CPSF by gel filtration chromatography. Purified CPSF was run on a S200 gel filtration column and equal amounts of the load (load) and the individual fractions (indicated at the bottom of the gel) were tested in a polyadenylation assay with 1 ng of recombinant PAP. In the lane labeled NXT, nuclear extract was added to the reaction mixture instead of the fractions. Input lanes represent mock-treated reactions. The migration positions of the internally [^{32}P]-labeled L3-pre RNA (L3-pre) and the polyadenylated product (poly(A) are indicated on the left. The molecular weight markers of the gel filtration column are indicated at the top.
C) Western blot analysis of the S200 fractions (indicated on top of the gel) with the antibodies indicated at the left. The lowest panel shows a Northern blot probed with a radioactive DNA oligo directed against the U2 snRNA.

2.3.2 Subunits of SF3b directly interact with components of CPSF

To confirm that subunits of the U2 snRNP interact with polypeptides of CPSF, we carried out co-immunoprecipitation experiments using HeLa cell nuclear extract. As shown in Figure 2.2A (left panel), antibodies directed against CPSF100 but not the pre-immune serum efficiently co-immunoprecipitated SF3b155, SF3b130, SF3a120 and SF3a66 in an RNA and DNA independent manner. Furthermore, the CPSF100 antibody pulled down the U2 snRNA (right panel) whereas the pre-immune serum or the Protein A-Sepharose beads without antibodies did not. Because the RNase A treatment digested the U2 snRNAs present in the nuclear extract, these co-immunopreciptitation experiments were done without RNase A.

To investigate if polypeptides of the U2 snRNP and CPSF directly interact with each other we carried out three independent GST-pull down experiments with recombinantly expressed GST-tagged SF3b49 and SF3b130. Both proteins interacted with CPSF160; CPSF100 interacted with SF3b130 and more weakly with SF3b49; CPSF30 was pulled down by SF3b49 and to a lower extent by SF3b130 (Figure 2.2B). Neither SF3b49 nor SF3b130 pulled down significant amounts of CPSF73 or PAP. As has been previously shown, SF3b130 interacted with SF3b155 (Das et al., 1999). All proteins interacted in an RNA and DNA independent manner. Taken together, our results show that CPSF and the U2 snRNP physically interact with each other.

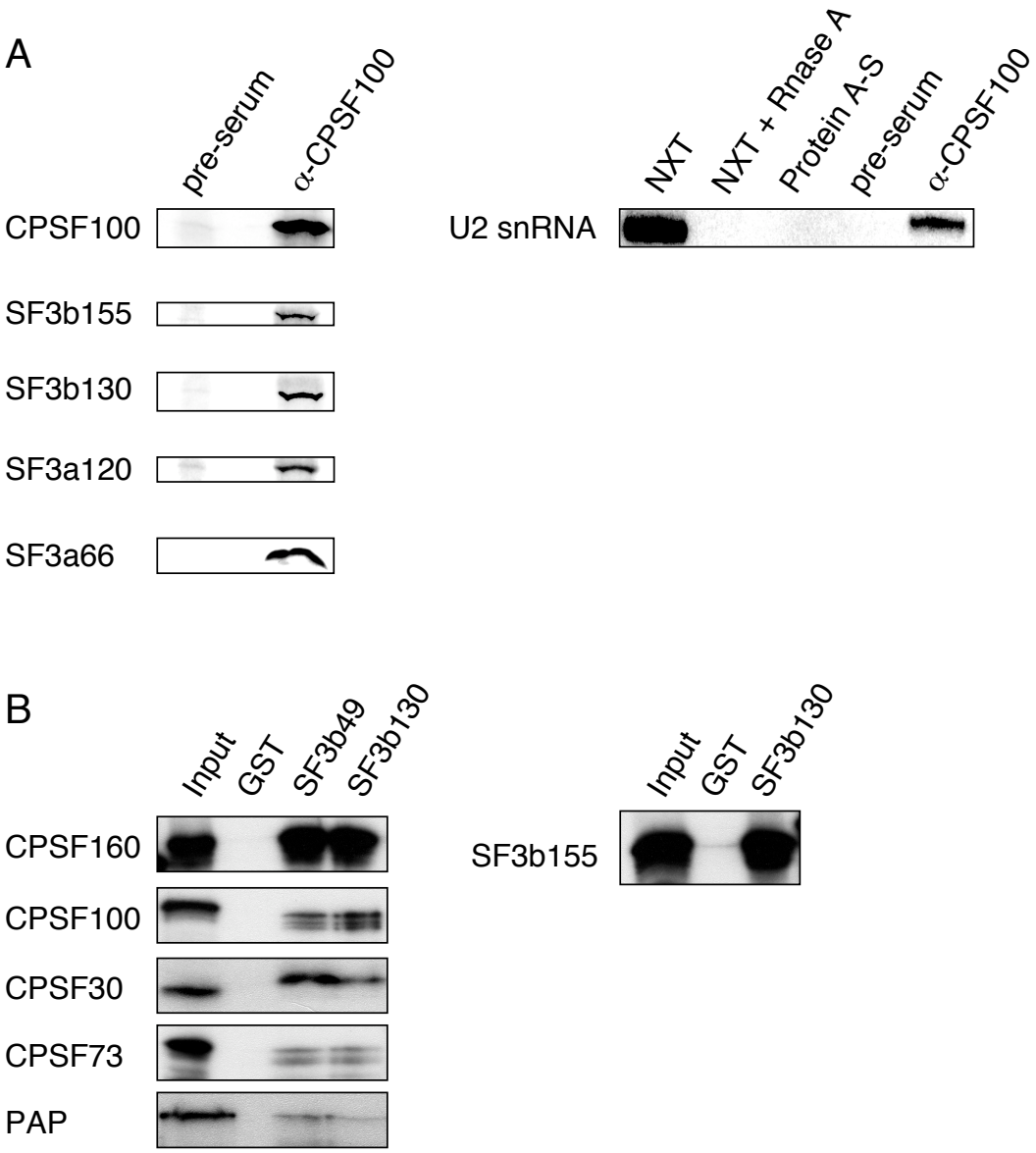


Figure 2.2 Subunits of the U2 snRNP directly interact with subunits of CPSF

A) The left panel shows a co-immunoprecipitation experiments with polyclonal antibodies directed against CPSF100 (α-CPSF100) or pre-immune serum (pre-serum) as indicated on top of the panel. The precipitates were separated on an 8% SDS gel and subjected to Western blot analysis with the antibodies indicated on the left of each panel.

The right panel shows a Northern blot of RNA extracted from nuclear extract (NXT), nuclear extract that was previously treated with RNase A and DNase I (NXT + RNase A) or co-immuoprecipitates of antibodies directed against CPSF100 (α-CPSF100), pre-immune serum (pre-serum) or Protein A-Sepharose only (Protein A-S) in the absence of RNase A and Dnase I.

B) GST pull-down experiments with 1 μg GST, 100 ng recombinant GST-tagged SF3b49 or SF3b130 (as indicated on top of each lane) with *in vitro* translated [^{35}S]methionine labeled proteins (indicated on

the left). The first lanes (Input) show 10% of the *in vitro* translated protein used in the binding reactions.

2.3.3 Depletion of CPSF100 inhibits splicing in a coupled splicing and 3' end cleavage assay

To examine if CPSF is required for efficient splicing we depleted HeLa cell nuclear extract of CPSF100 and tested these extracts in a coupled assay using the MXSVL RNA substrate (Niwa et al., 1990); Figure 2.3A). This RNA contains two exons separated by an intron and the second exon carries a SV40 polyadenylation signal. S-C- represents the unprocessed MXSVL substrate; S−C+ represents RNA that has been cleaved but not spliced; S+C− refers to RNA that has been spliced but not cleaved; and S+C+ represents RNA that has been fully processed. CPSF100-depleted extract showed no cleavage activity (lane 4, S−C+, S+C+) and its splicing activity was significantly reduced (lane 4, S+C−, S+C+). In comparison, extract depleted of hPcf11 (a subunit of CF II_m) resulted in a decrease of cleavage activity (lane 5, S−C+, S+C+) but spliced the substrate efficiently (lane 5, S+C−). Untreated nuclear extract (lane 1), extract which had been incubated with Protein A-Sepharose only (lane 2), and extract that was incubated with pre-immune serum (lane 3) were active in cleavage (S−C+, S+C+) and in splicing (S+C−, S+C+). These results indicate that it is not the cleavage reaction itself that is required for stimulation of splicing but rather the proteins involved in 3' end processing.

Importantly, the CPSF100-depleted extract (α-CPSF100) contained significantly lower amounts of CPSF100 and CPSF73 (Figure 2.3B) compared to the extract that was treated with pre-immune serum (pre-serum). However the amounts of SF3b155, SF3b130, SF3a120 and SF3a66 were not significantly lowered, indicating that the splicing deficiency of this extract is indeed due to the lack of CPSF and not a result of lower amounts of U2 snRNP subunits.

Furthermore, addition of partially purified CPSF to CPSF100-depleted extract was able to reconstitute the cleavage (Figure 2.3C, S−C+, lane 4-6) and the splicing activity (S+C−, lane 4-6). CPSF on its own showed no processing activity (lane 7). We were not able to reconstitute the cleavage activity nor the splicing activity by addition of highly purified CPSF. Depletion of CPSF100 could result in co-depletion

of other factors of the 3' end processing machinery which are most likely not present in highly purified CPSF fractions.

We also tested the splicing activity of CPSF100-depleted extracts on the MINX RNA substrate. This RNA contains the same exons and intron as the MXSVL RNA but lacks the SV40 polyadenylation cassette. Interestingly, CPSF100-depleted extract showed no significant reduction in its splicing activity when tested on the MINX substrate (Figure 2.3D, compare lane 6 with lane 5; intron and mRNA levels) but had reduced splicing activity when assayed with the MXSVL RNA (compare lane 3 with lane 2; S+C+ and intron levels). Taken together these results imply that CPSF is required for the efficient splicing in coupled but not in uncoupled reactions.

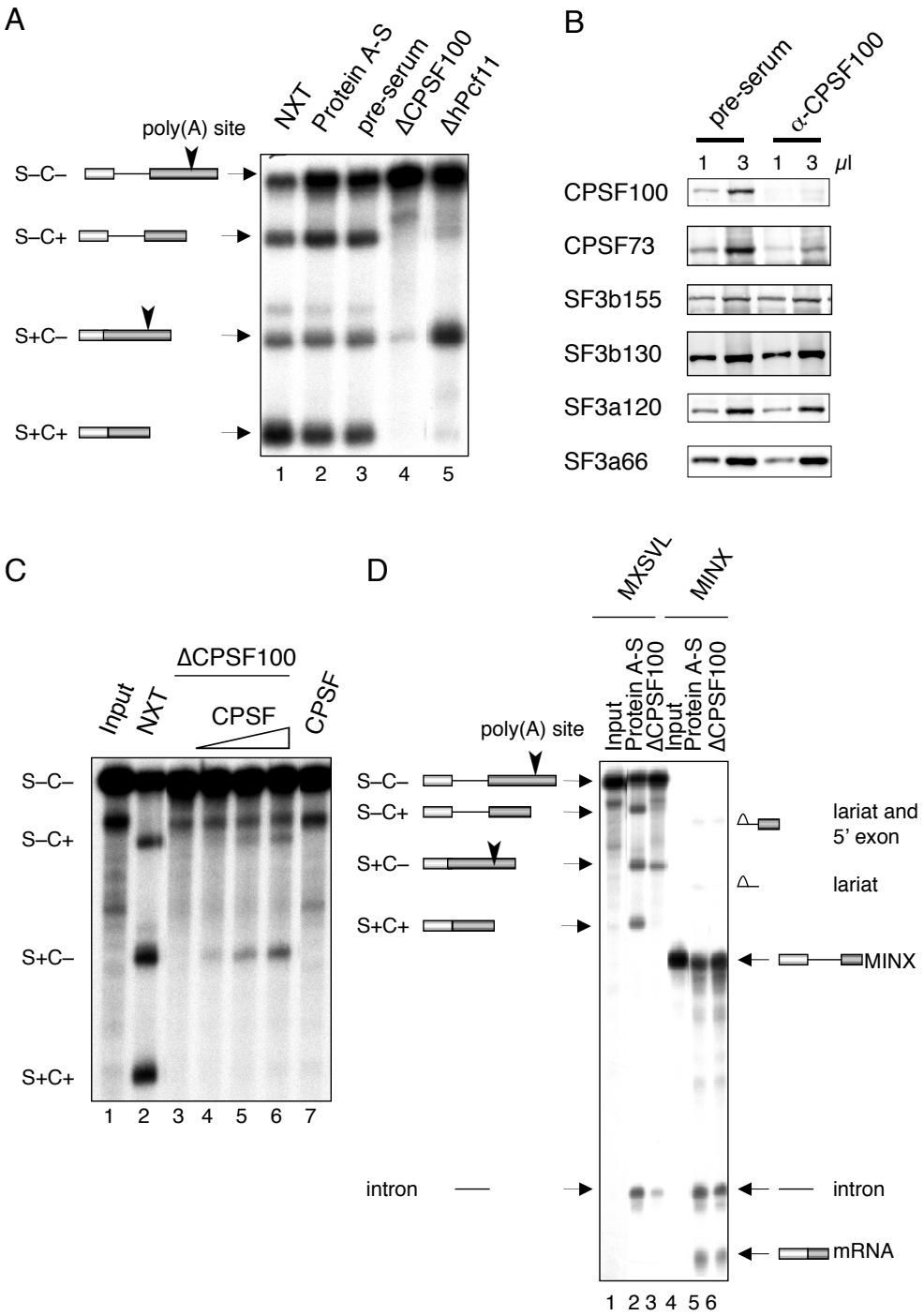


Figure 2.3 CPSF is required for efficient splicing of the MXSVL substrate *in vitro*

A) Coupled 3' end cleavage and splicing assay with the internally [^{32}P]-UTP-labeled MXSVL RNA. The migration position of the unprocessed S–C–, cleaved but not spliced S–C+, spliced but not cleaved S+C– and spliced and cleaved S+C+ RNAs are indicated on the left. Nuclear extract (NXT) was either mock treated with Protein A-Sepharose (Protein A-S) or treated with pre-immune serum (pre-serum), or depleted of CPSF100 (ΔCPSF100) or hPcf11 (ΔhPcf11) as indicated at the top and incubated with the RNA as described in Experimental Procedures. The RNA products were separated on a 4% polyacrylamide gel.

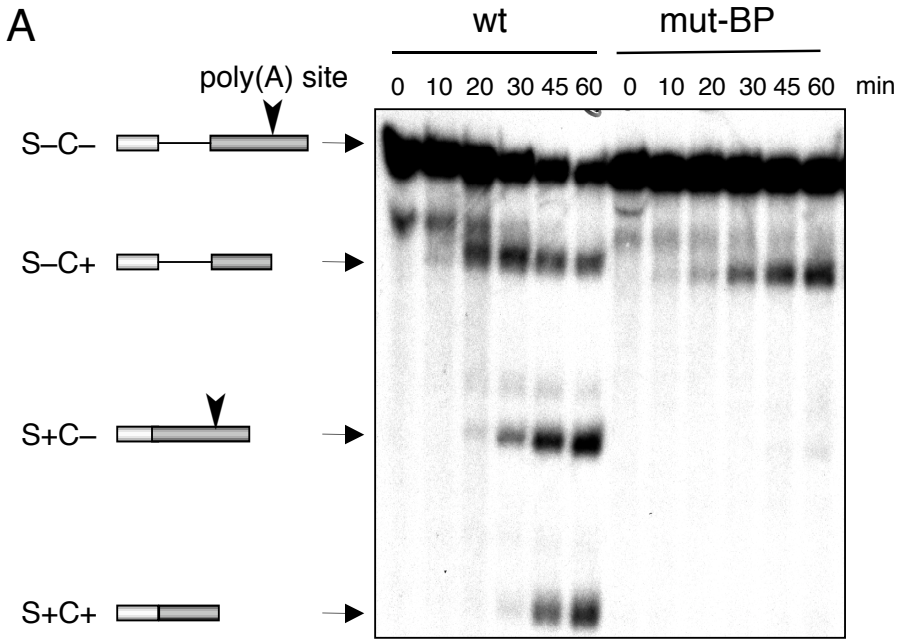
B) Western blot analysis of NXT depleted of CPSF100 (ΔCPSF100) or NXT treated with pre-immune serum (pre-serum) as indicated at the top. The antibodies used for the Western blot are indicated at the left.
C) Coupled splicing and 3' end cleavage assay. Input represents mock treated RNA. The cleavage and splicing activity of CPSF100-depleted extract (lanes 3-6) was reconstituted by addition of partially purified CPSF (lanes 4-6; S−C+, S+C−). CPSF represents RNA that has been incubated with CPSF alone (lane 7). The RNA products were separated on a 4% polyacrylamide gel.
D) Comparison between the splicing activity of CPSF100-depleted extract on the MXSVL versus the MINX RNA substrate. MXSVL (lanes 1-3) and MINX (lanes 4-6) RNAs were incubated with nuclear extract that was mock treated with Protein A-Sepharose (lanes 2 and 5) or depleted of CPSF100 (lanes 3 and 6). Input lanes represent mock treated reactions (lanes 1 and 4). For explanations of the migration positions of the MXSVL processing products see legend to panel A. The scheme to the right represents the migration positions of the MINX RNA processing products.

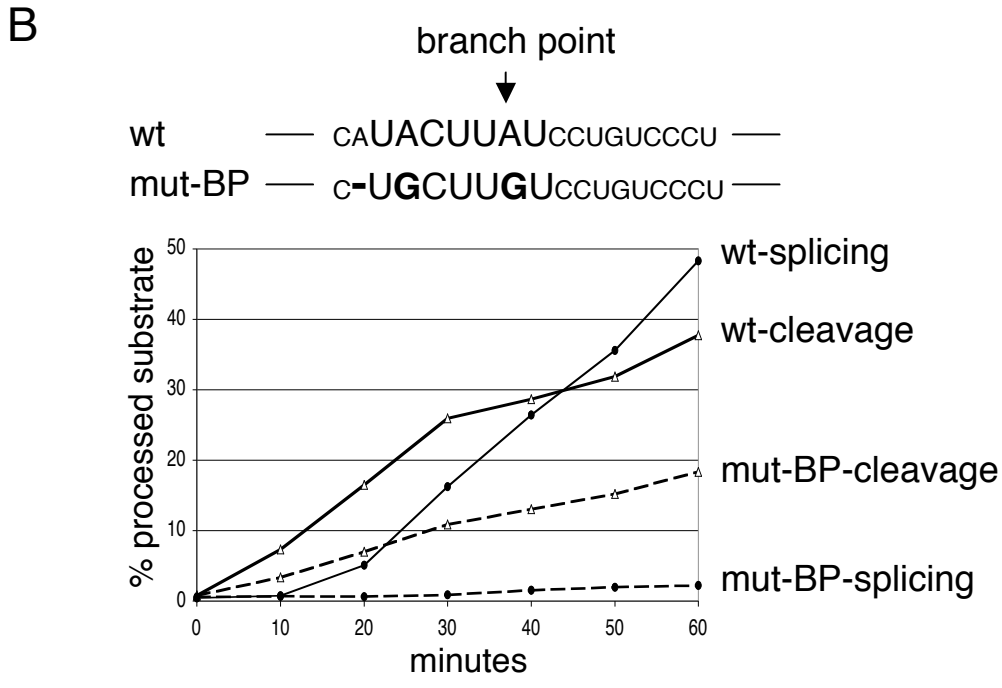
2.3.4 Mutations in the U2 snRNA binding site inhibit 3' end cleavage in a coupled splicing and 3' end cleavage assay

The interaction of CPSF with the U2 snRNP implied that components of the U2 snRNP might be involved in pre-mRNA 3' end processing. To address this hypothesis, we mutated the U2 snRNP binding site on the MXSVL substrate. We generated two pre-mRNAs with mutations in the branch point sequence (Figure 2.4B and 2.4D, mut-BP and mut-BPII) and one RNA with mutations downstream of the branch point sequence (Figure 2.4C, down-BP).

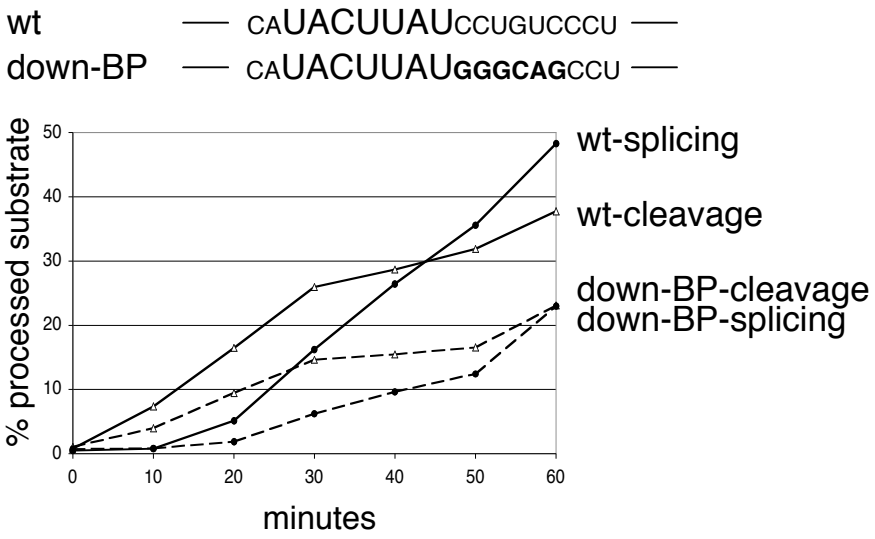
Figure 2.4A shows a time-course assay with the wild-type (wt; MXSVL) and the mut-BP substrate. We quantified the RNA products of three individual experiments and calculated the percentage of splicing (sum of S+C− and S+C+) and cleavage (sum of S−C+ and S+C+) activities in relation to the total RNA used in the assay (sum of S−C− and S−C+ and S+C− and S+C+). As can be seen in Figure 2.4B, the mut-BP RNA was not spliced and showed an average reduction in cleavage activity of 52% with a standard deviation (SD) of 10% compared to the wild-type substrate after one hour incubation at 30°C. Figure 2.4C shows the splicing and the cleavage activity of the down-BP RNA. This mutation resulted in an average reduction of the splicing activity of 55% (SD 15%) and in an average decrease of the cleavage activity of 40% (SD 14%) compared to the wild-type RNA after one hour incubation at 30°C.

Interestingly, not all mutations in the branch point sequence affected 3' end processing. The mut-BPII mutations had a similar effect on splicing as the down-BP mutation but did not result in a significant reduction of the cleavage activity (Figure 2.4D). This indicates that it is not the reduced splicing activity that leads to the decrease of 3' end processing, but rather the proteins involved in the splicing and 3' end formation reactions and potentially their interaction with each other.





C



D

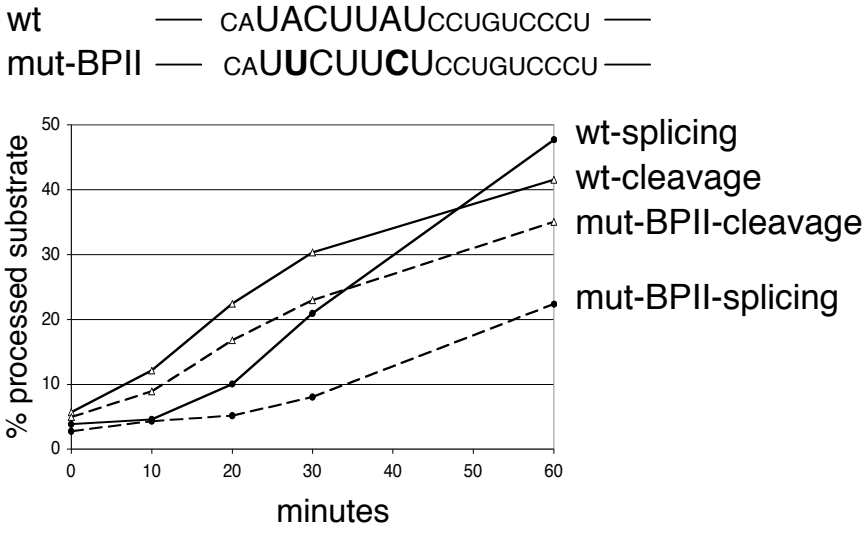


Figure 2.4 Mutations in the branch point sequence inhibit 3' end cleavage activity *in vitro*

A) Time course experiment of coupled splicing and cleavage assay with MXSVL (wt) and mut-BP (mut-BP) RNA substrate (indicated on top of the gel). The products were separated on a 4% polyacrylamide gel. For explanations of the migration positions of the RNA products see legend to figure 2.3.A

B) The scheme at the top represents the wild-type branch point sequence of the MXSVL substrate. The larger letters mark the conserved branch point sequence. Below is the sequence of the mut-BP RNA substrate; the mutations are indicated in bold. Three gels from individual experiments were quantitated and the splicing and 3' end cleavage activities were calculated as follows: percent splicing activity = 100/(S–C– + S–C+ + S+C– + S+C+) x (S+C– + S+C+); percent cleavage activity = 100 / (S–C– + S–C+ + S+C– + S+C+) x (S–C+ + S+C+).

C) See legend to panel B, the assays were carried out with the substrate down-BP.

D) See legend to panel B, the assays were carried out with the substrate mut-BPII.

2.3.5 Complexes involved in the coupling of 3' end processing and splicing do not form efficiently on mutant RNA substrates deficient in splicing and 3' end cleavage

It has been shown before that specific RNA-protein complexes correlating with the coupling of 3' end processing and splicing form on the MXSVL RNA (Cooke and Alwine, 2002). We therefore carried out RNA band shift experiments with our mutant substrates in order to find out if the mutations had an effect on complex formation. As shown in Figure 2.5A the nonspecific complex N is formed on the RNAs at the time point zero. After one minute incubation at 30°C the so called A_C complex appears which is thought to be a precursor of the coupling complex B_C. Note that the A_C and B_C complexes differ from the A and B complexes that form on splicing substrates that contain no polyadenylation cassette.

We calculated the percentage of B_C complex formation compared to the amount of complex N present at time 0. The quantitative evaluation of four independent experiments can be seen in Figure 2.5B. The amount of B_C complex formation after 60 minutes did not change compared to the 30 minutes time point (results not shown). The mut-BPII mutation, which did not affect cleavage activity significantly, resulted in a decrease of B_C complex formation of only 25% (SD 7) compared to the wild-type. At present we do not know why this RNA was spliced less efficiently. The formation of the B_C complex was significantly reduced compared to the wild-type substrate: 45% (SD 5) with the mut-BP RNA and 47% (SD 11) with the down-BP substrate after 30 minutes. This agrees with the observation that both mutants are spliced and cleaved with lower efficiency.

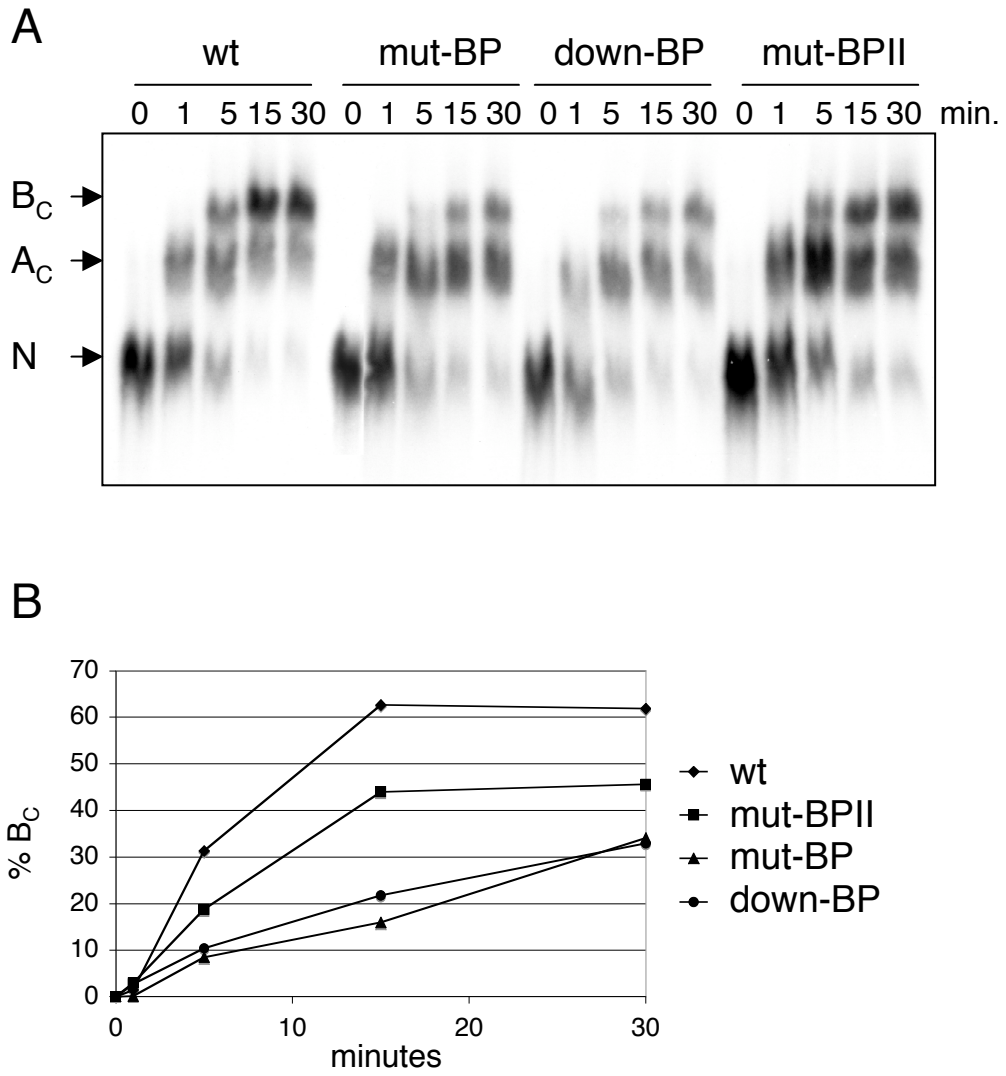


Figure 2.5 The coupling complex B_C forms less efficiently on the mut-BP and down-BP RNA substrates

A) Time course complex formation on the different RNA substrates indicated at the top. The migration positions of the complexes are shown on the left.

B) Quantitative evaluation of the B_C complex formation over time. The percentage of the B_C complex formation was calculated in comparison to the amount of complex N at time zero.

2.3.6 Digestion of the U2 snRNA results in a decrease of 3' end cleavage activity in coupled assays

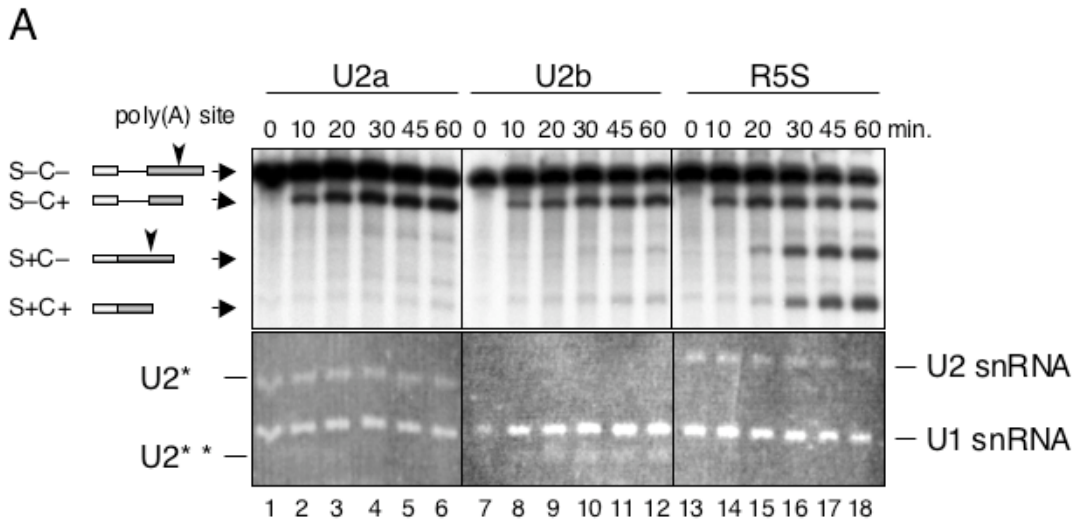
Given that mutations in the RNA substrate could affect not only the interaction of the U2 snRNP with the RNA but also the interactions of other proteins with the RNA or

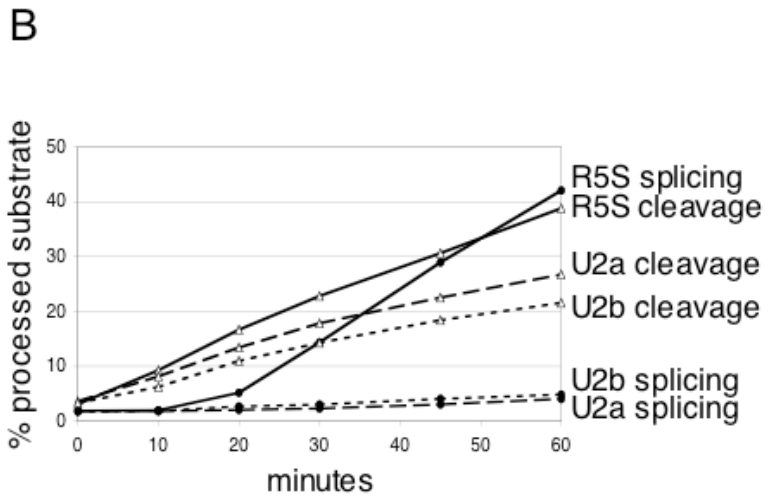
could influence the RNA secondary structure, we wanted to confirm that the U2 snRNP itself is involved in the coupling of 3' end processing and splicing. To this end, we digested the U2 snRNA of nuclear extract with RNase H and two different DNA oligo nucleotides complementary to the U2 snRNA (U2a, U2b; Black et al., 1985) and carried out a coupled cleavage and splicing assay (Figure 2.6A, upper panels). The control DNA oligo R5S is complementary to putative single-stranded regions of 5S ribosomal RNA (Black et al., 1985). We determined the extent of the digestion of the U2 snRNA by staining the gel with ethidium bromide (lower panels). As has been shown before, the treatment of nuclear extact with the U2a DNA oligo and RNase H resulted in a truncated form of the U2 snRNA (Black et al., 1985; U2*, lanes 1-6). The treatment of nuclear extract with RNase H and the U2b DNA oligo resulted also in digestion of the U2 snRNA and a faint band representing the digestion product is visible on the gel (Black et al., 1985, U2**, lanes 7-12). In comparison, nuclear extract treated with the R5S DNA oligo and RNase H had no effect on the levels of the U2 snRNA present in the extract (lanes 13-18). All extracts contained the same amount of U1 snRNA. Figure 2.6B shows the quantitative evaluation of two independent experiments. As expected, the U2 snRNA-depleted extract lost its splicing activity and its cleavage activity was significantly reduced compared to the control extract to an average of 30% (SD 4%) with U2a, and 45% (SD 2%) with U2b after one hour incubation at 30°C.

If the different effects on splicing and 3' end cleavage of the mutated RNA substrates shown in Figure 2.4 are indeed resulting form the involvement of the U2 snRNP in coupling splicing and 3' end processing one would expect that the different substrates show the same splicing and cleavage activity when assayed with U2 snRNA-depleted extract. We therefore tested U2 snRNA-depleted nuclear extract in coupled assays with the different mutant RNAs. Figure 2.6C shows the quantitative evaluation of three independent experiments. In this set of experiments treatment of the nuclear extract with RNase H resulted in an inhibition of 3' end cleavage activity on the wild-type RNA of 46% (SD 15) with the U2a oligo and 54% (SD 7) with the U2b oligo after one hour incubation at 30°C compared to extract treated with the control oligo 5RS. Upon U2 snRNA digestion the wild-type and the different mutant RNAs all showed similar splicing and cleavage activities. This observation

strengthens the hypothesis that the mut-BP and down-BP mutations affect the role of the U2 snRNP in coupling splicing and 3' end formation.

U2 snRNA-depleted extracts were not deficient in 3' end cleavage when we tested them on the SV40 RNA substrate (Figure 2.6D). This RNA contains no introns but the SV40 polyadenylation cassette – the same sequence which is present in the MXSVL RNA. This shows that the U2 snRNA is required for efficient 3' end cleavage in coupled but not in uncoupled reactions. In summary, these results support the notion that CPSF and the U2 snRNP are involved in mediating the coupling of 3' end processing and splicing by their direct interaction.





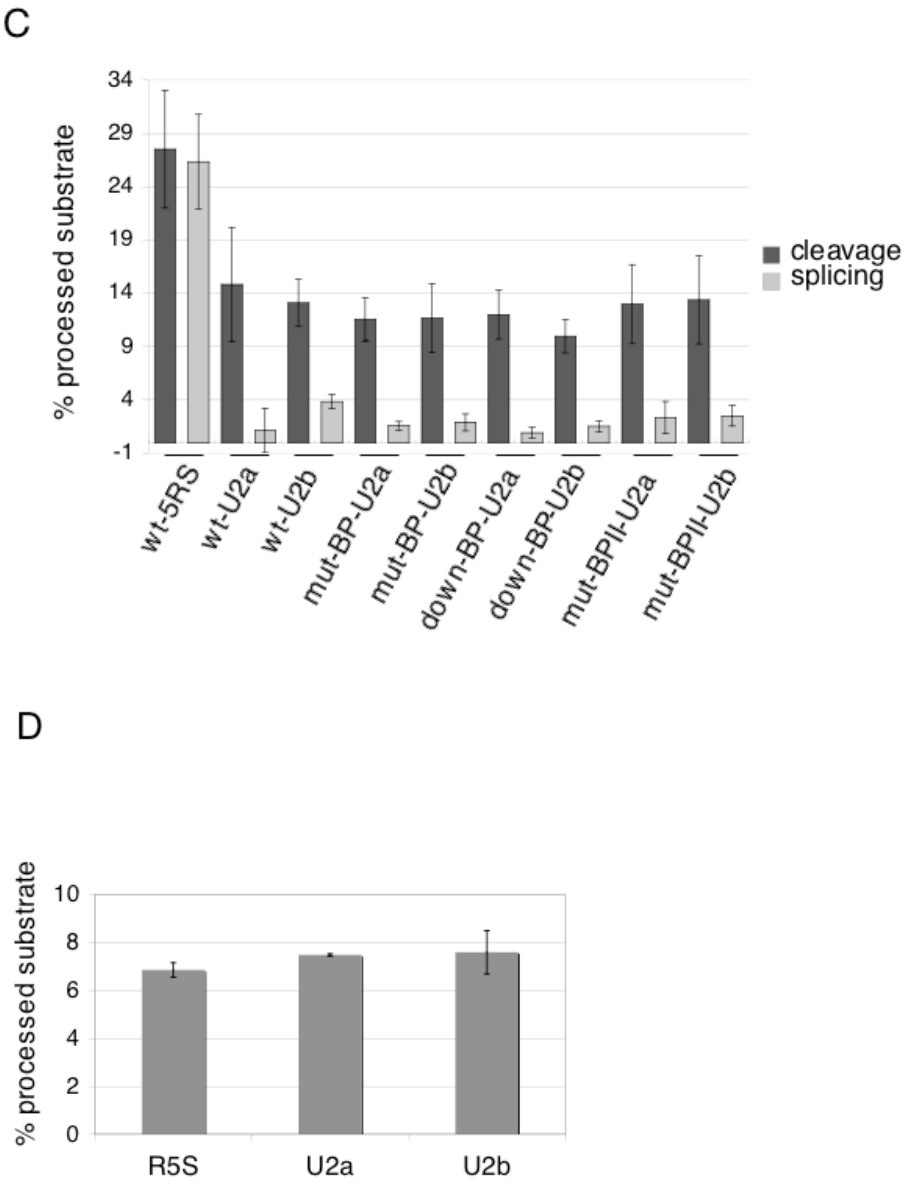


Figure 2.6 U2 snRNA is required for efficient 3' end cleavage in coupled assays

A) Upper panels: Coupled splicing and 3' end cleavage time course experiment with extract that was treated with RNase H and the U2a DNA oligo (U2a), U2b DNA oligo (U2b) or the R5S DNA oligo (R5S) as indicated on top. For explanation of the migration positions of the RNA substrate and products see legend to figure 3A. The products were separate on a 10% polyacrylamide gel.

Lower panels: Ethidium bromide staining of the gel shown in the upper panels. The migration positions of the U1 and U2 snRNAs are indicated on the right. The U2* and U2** indicate the migration position of the U2 snRNA digestion product resulting after treatment with the U2a or the U2b DNA oligo respectively.

B) Quantitative evaluation of the splicing and 3' end cleavage activities. The graph represents an average of two individual experiments. For explanation of the calculation see legend to figure 4B.

C) The graph represents the quantitative evaluation of three independent experiments of the 3' end cleavage and splicing activites of nuclear extract that has been treated with Rnase H and the DNA oligos indicated at the bottom. The different substrates used in the assays are also indicated at the bottom. The error bars represent the standard deviation.
D) The graph shows the quantitative evaluation of two independent experiments of the 3' end cleavage activity of nuclear extract that has been treated with Rnase H and the DNA oligos indicated at the bottom. SV40 was used as RNA substrate. The error bars represent the standard deviation.

2.4 Discussion

Splicing and 3' end formation are coupled processes. As has been reported earlier, the U1 snRNP, U2AF65, the SR-related nuclear matrix protein of 160 kDa (SRm160), PAP and CPSF30 have been implied to participate in this coupling (Wassarman and Steitz, 1993; Gunderson et al., 1994; Lutz et al., 1996; Blencowe et al., 1998; Gunderson et al., 1998; Vagner et al., 2000; Millevoi et al., 2002; Awasthi and Alwine, 2003). However, the U2 snRNP has so far not been known to play a role in coupling splicing and 3' end formation.

In order to identify proteins, which associate with CPSF we purified the factor over five columns and found that many subunits of the U2 snRNP co-purified with CPSF. We did not address the question if the entire U2 snRNP was present in these fractions. However, we detected the U2 snRNA and all the tested subunits of the U2 snRNP except SF3b125 in the purified complex. Interestingly SF3b125 is thought to dissociate from SF3b upon formation of the 17S U2 snRNP (Will et al., 2002) and is therefore not present in the fully assembled U2 snRNP. The U2 snRNP proteins were present in lower amounts compared to the CPSF polypeptides in our purified fractions. This implies that only a part of all the CPSF complexes present in a cell are associated with subunits of the U2 snRNP. This could mean that a subset of CPSF complexes could be permanently associated with U2 snRNPs; however we think it is more likely that the two complexes interact with each other in a dynamic fashion.

The purified complex was size fractionated on a gel filtration column on which SF3b130, SF3a60 and the U2 snRNA were found in the same fractions as CPSF100. We confirmed that CPSF interacts with the U2 snRNP by co-immunoprecipitation experiments and showed that this association is mediated by a direct interaction between subunits of SF3b and polypeptides of CPSF. To our

knowledge this is the first demonstration of a physical interaction between CPSF and the U2 snRNP. The previously known and the new protein-protein interactions between the splicing machinery and the 3' end processing factors on the last exon of a pre-mRNA are summarized in Figure 2.7

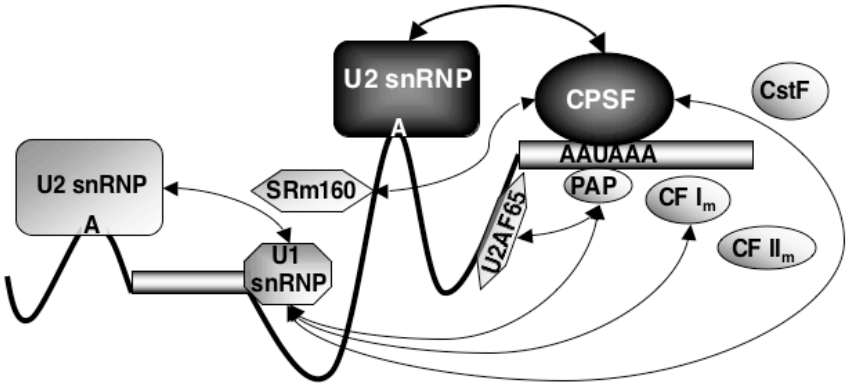


Figure 2.7 Model of protein-protein interactions on the last exon

Arrows indicate the previously known and the new protein-protein interactions of the splicing machinery with the 3' end processing factors on the last exon (see text for details).

The interaction between CPSF and the U2 snRNP suggested that the two complexes might be involved in mediating the coupling of 3' end processing and splicing. It has been shown earlier that mutations in the highly conserved CPSF binding site AAUAAA lead to decreased splicing efficiency (Niwa and Berget, 1991; Cooke et al., 1999). To confirm that this effect can be explained by the involvement of CPSF in splicing, we depleted CPSF100 from nuclear extract. We found that depletion of CPSF100 not only abolished cleavage but also significantly decreased the splicing activity in coupled assays. Both activities were reconstituted upon the addition of partially purified CPSF. Furthermore, depletion of the CF II_m subunit hPcf11 resulted in a strong decrease of 3' end cleavage but not in reduced splicing activity. This shows that it is not the cleavage reaction itself that is required for efficient splicing but that the presence of proteins involved in 3' end processing are responsible for the stimulatory effect.

SF3a and SF3b were shown to bind to the branch point and surrounding sequences during spliceosome assembly (Query et al., 1996; Gozani et al., 1996; Gozani et al., 1998; Will et al., 2001). We showed here that two independent mutations in these sequences significantly reduced splicing and cleavage activities. Protein complexes involved in the coupling of splicing and 3' end processing did not form efficiently on these mutated substrates**.** Furthermore, we found that one mutation in the branch point sequence reduced the splicing activity but had no significant effect on the cleavage activity. This indicates that it is indeed the proteins involved in splicing which affect 3' end processing and not the splicing reaction per se. In addition depletion of the U2 snRNA from nuclear extract also resulted in decrease of the splicing and cleavage activities in coupled assays. These results corroborate the unanticipated role of the U2 snRNP in mediating the coupling of splicing and 3' end formation. This is the first time that a snRNP binding in the proximity of the 3' splice site was shown to be involved in this coupling mechanism.

The exon definition model suggests that the boundaries of internal exons are defined by the binding of the U1 and U2 snRNPs and associated splicing factors. Interestingly, the role of the U2 snRNP is therefore to define the 5' border of the exon whereas the U1 snRNP demarcates the 3' border. In the case of the last exon, the 3' end processing complex is thought to be responsible to define the 3' border of the exon. This model implies that the coupling of 3' end processing and splicing comes primarily into play at the last exon (Berget, 1995; Cooke et al., 1999). In accordance with this, we found that depletion of CPSF100 reduced the splicing efficiency when assayed with the MXSVL but had no significant effect on the splicing efficiency when tested with the MINX RNA, which does not contain a polyadenylation cassette. Furthermore, we showed that digestion of the U2 snRNA resulted in inefficient 3' end cleavage of the MXSVL substrate but had no effect on 3' end cleavage of the intronless SV40 RNA substrate. This implies that CPSF and the U2 snRNP play a role in mediating the coupling of 3' end formation and splicing on the last intron but that CPSF is most likely not involved in the splicing of internal introns and that the U2 snRNP is not required for efficient 3' end cleavage of an intronless RNA precursor. However, further experiments are needed to prove that the coupling mechanism mediated by CPSF and the U2 snRNP is mainly required for the splicing of the last intron.

Given that the factors involved in splicing and in 3' end formation are recruited to the transcribing RNAP II by binding to the CTD (Gornemann et al., 2005; reviewed in Proudfoot et al., 2002; Zorio and Bentley, 2004) it is possible that CPSF and the U2 snRNP already interact with each other while traveling with the elongating RNA polymerase *in vivo*. Interestingly CPSF plays an important role in cleavage site selection (Keller et al., 1991; Gilmartin et al., 1995; Graveley et al., 1996; Brackenridge and Proudfoot, 2000) and the U2 snRNP is required for splice site selection (Gozani et al., 1996; Query et al., 1996; Gozani et al., 1998; Brackenridge and Proudfoot, 2000; Will et al., 2001). The fact that they directly interact with each other as well as with the pre-mRNA leads us to propose that this is a mechanism to promote the coupling between splicing and 3' end processing by facilitating their recruitment to the RNA and the mutual stabilization of the splicing and 3' end processing machineries.

2.5 Experimental Procedures

2.5.1 Purification of cleavage and polyadenylation specificity factor from calf thymus and MALDI-TOF MS analysis

CPSF was purified from one kilogram of calf thymus by chromatography on DEAE-Sepharose, Blue-Sepharose, Heparin-Sepharose, Mono-Q and Poly(A)-Sepharose essentially as described by(Bienroth et al., 1991). The peak fractions of the poly(A)-Sepharose column were pooled and the proteins were identified by mass spectrometry as described (de Vries et al., 2000).

Peak fractions of the Poly(A)-Sepharose were pooled, concentrated on a 5 ml DEAE column, and separated on a Superdex S200 gel-filtration column. The buffer used for the gel-filtration contained: 100 mM KCl, 50 mM Tris-HCl pH 8, 0.2 mM EDTA, 3 mM MgCl, 10% glycerol, 0.02% NP40, 0.5 mM DTT, 1 mM PMSF, 0.4 μg/ml leupeptin, 0.7 μg/ml pepstatin.

Partially purified CPSF used for the reconstitution of the cleavage and splicing activities were obtained from a Mono S fraction as described by ((Ruegsegger et al., 1996).

2.5.2 Northern analysis

Fractions of the Superdex S200 gel-filtration column and the co-immunoprecipitates were incubated with a final concentration of 100 mM Tris-HCl pH 7.9, 150 mM NaCl, 12.5 mM EDTA, 1% SDS, 1 mg/ml Proteinase K and 0.7 µg/µl glycogen at 42°C for 1.5 hrs. Two phenol-chloroform extractions were performed and the upper phases incubated with a final concentration of 85 mM NaOAc and 70% ethanol at -20°C overnight. The RNA was precipitated, washed with 70% ethanol and run on a 8.3 M urea/12% polyacrylamide gel and transferred to Hybond N+ (Amersham Pharmacia Biotech). Oligonucleotides anti-U2 (tattgtcctcggatagaggacgta) and anti-U2-2 (ttagccaaaaggccgagaagc) were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Both oligonucleotides recognized the same band in the Northern blots (results not shown).

2.5.3 Co-immunoprecipitations and immunodepletions of HeLa cell nuclear extract

Co-immunoprecipitations and immunodepletions of HeLa cell nuclear extract were performed as described (Barabino et al., 1997). For the co-immunoprecipitation experiments 50 µl nuclear extract was incubated with 0.1 mg/ml RNase A and 0.2 U DNase I (Promega) at 30°C for 30 min. prior to addition to the Protein A-Sepharose bound antibodies.

2.5.4 Expression of recombinant proteins in *Escherichia coli*

Recombinant proteins were expressed in *E.coli* BL21 Lys(S) and purified as described (Kyburz et al., 2003).

2.5.5 GST-pulldown experiments

GST-pulldown experiments were carried out as described (Kyburz et al., 2003) with binding buffer containing 0.1 mg/ml RNase A and 0.2 U DNase I (Promega).

2.5.6 Preparation of RNA substrates

RNA substrates were prepared as described (Rüegsegger et al., 1996). Mutations in the branch point and in downstream sequences were generated by site directed PCR mutagenesis, followed by DpnI digestion of the parent plasmid. The PCR products were transformed into *E. coli* and sequenced.

2.5.7 Polyadenylation assay

In vitro polyadenylation assays were carried out as described (Bienroth et al., 1991).

2.5.8 Coupled cleavage and splicing assay

Nuclear extract (NXT) was prepared as described (Ruegsegger et al., 1996). The NXT was dialyzed against 20 mM Tris-HCl pH 7.9, 20 mM $(NH_4)_2SO_4$, 3 mM $MgCl_2$, 20% glycerol, 0.2 mM EDTA, 0.02% NP40, 0.5 mM DTT, 0.5 mM PMSF, 0.4 μg/ml leupeptin, 0.7 μg/ml pepstatin. 15 fmol RNA substrate and 6 μl NXT were incubated with 0.25 mM ATP, 1 mM Cordycepin triphosphate, 20 mM creatine phosphate, 2.6% polyvinylalcohol, 1 mM $MgCl_2$, 0.01% NP40, 2 mM DTT, 0.01 mg/ml creatine kinase, 0.2 U RNase guard (Amersham) and 0.1 μg/μl tRNA in a total volume of 25 μl for one hour at 30°C unless indicated otherwise. The proteins were digested with Proteinase K and the RNA was separated by elecrophoresis on 4% or 10% acrylamide gels as described (Minvielle-Sebastia et al., 1994).

2.5.9 Complex formation analysis

15 fmol RNA substrate and 2 µl nuclear extract were incubated at 30°C in the reaction mixture used for the coupled cleavage and splicing assay in a total volume of 25 µl. The reaction was stopped by addition of 5 µl termination buffer containing 40% glycerol, 3 µg/µl heparin and 30 mM EDTA. The complexes were separated on native 3.5% polyacrylamide gels as described (Cooke and Alwine, 2002).

2.5.10 Oligodeoxynucleotide-directed RNase H cleavage

The synthetic oligonucleotides U2a (E15), U2b (L15) and R5S has been described previously (Black et al., 1985). 10 μl NXT was incubated with 4 μM of the oligonucleotides, 0.5 U RNase H (Invitrogen) and the reaction mixture described in the coupled cleavage and splicing assay without the tRNA in a total volume of 23 μl for one hour at 30°C. The reaction was started by addition of 15 fmol RNA substrate dissolved in 2 μl tRNA [0.1 μg/μl] and continued as explained in the coupled cleavage and splicing assay. The gel was stained with ethidium bromide and the snRNAs were visualized by illumination with UV light.

2.5.11 Plasmids

pAK 33 (GST-Tev-SF3b 49-His6): SF3b 49 was PCR amplified with the primers AGTATCCCGGGTATGGCTGCCGGGCCGATCT and TAGACTGCGGCCGCTCAGTGGTGGTGGTGGTGCTGAGGGAGAGGGCCTCGA followed by digestion with XmaI and NotI and ligated into pGDV1 (Dichtl and Keller, 2001). pAK 35 (GST-Tev-SF3b 130-His6): SF3b 130 was PCR amplified with the primers AGTATCCCGGGTATGTTTCTGTACAACTTAACCTTG and TAGACTGCGGCCGCTCAGTGGTGGTGGTGGTGGTGGAAGGCGTAGCGGGTCCGG digested with XmaI and NotI and ligated into pGDV1.

2.6 Supplemental material

2.6.1 Results and discussion

2.6.1.1 Mutations in AAUAAA

(Niwa and Berget, 1991; Cooke et al., 1999) showed, that mutations in the highly conserved CPSF binding site AAUAAA to AAGAAA and downstream sequences lead to decreased splicing efficiency and abolished the 3' end cleavage activity on these mutated MXSVL substrates. In order to confirm their results, we generated MXSVL RNA substrates with mutations in the AAUAAA hexanucleotide sequence (Figure 2.8.).

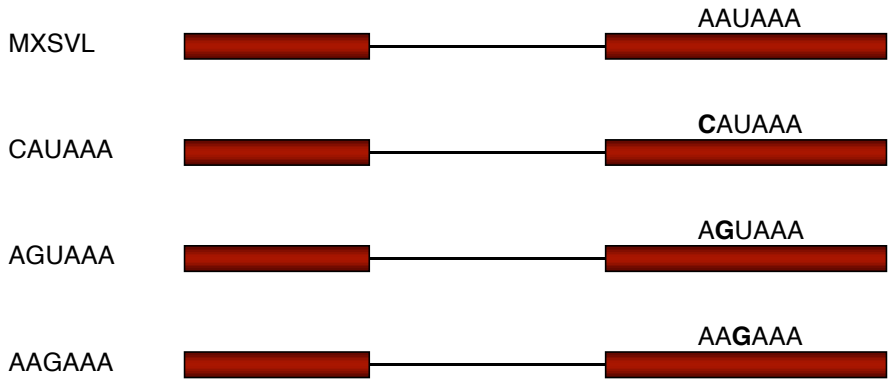


Figure 2.8 Schematic representation of RNA substrates with mutations in AAUAAA.
The scheme represents the different RNA substrates carrying mutations in the AAUAAA hexanucleotide sequence. The name of the substrates are indicated at the left. The mutations are shown in bold letters. MXSVL is the wild-type substrate.

Figure 2.9 shows the quantitative evaluation of two independent time course experiments with the wild-type and mutated RNAs. Of all three mutations the CAUAAA had the weakest effect on 3' end cleavage activity. The AGUAAA RNA mutation caused a strong reduction in the cleavage activity as did the AAGAAA mutation. Interestingly, the AAGAAA mutation did not completely abolish the

cleavage reaction. And most surprisingly, we found that none of the mutations affected the splicing activity.

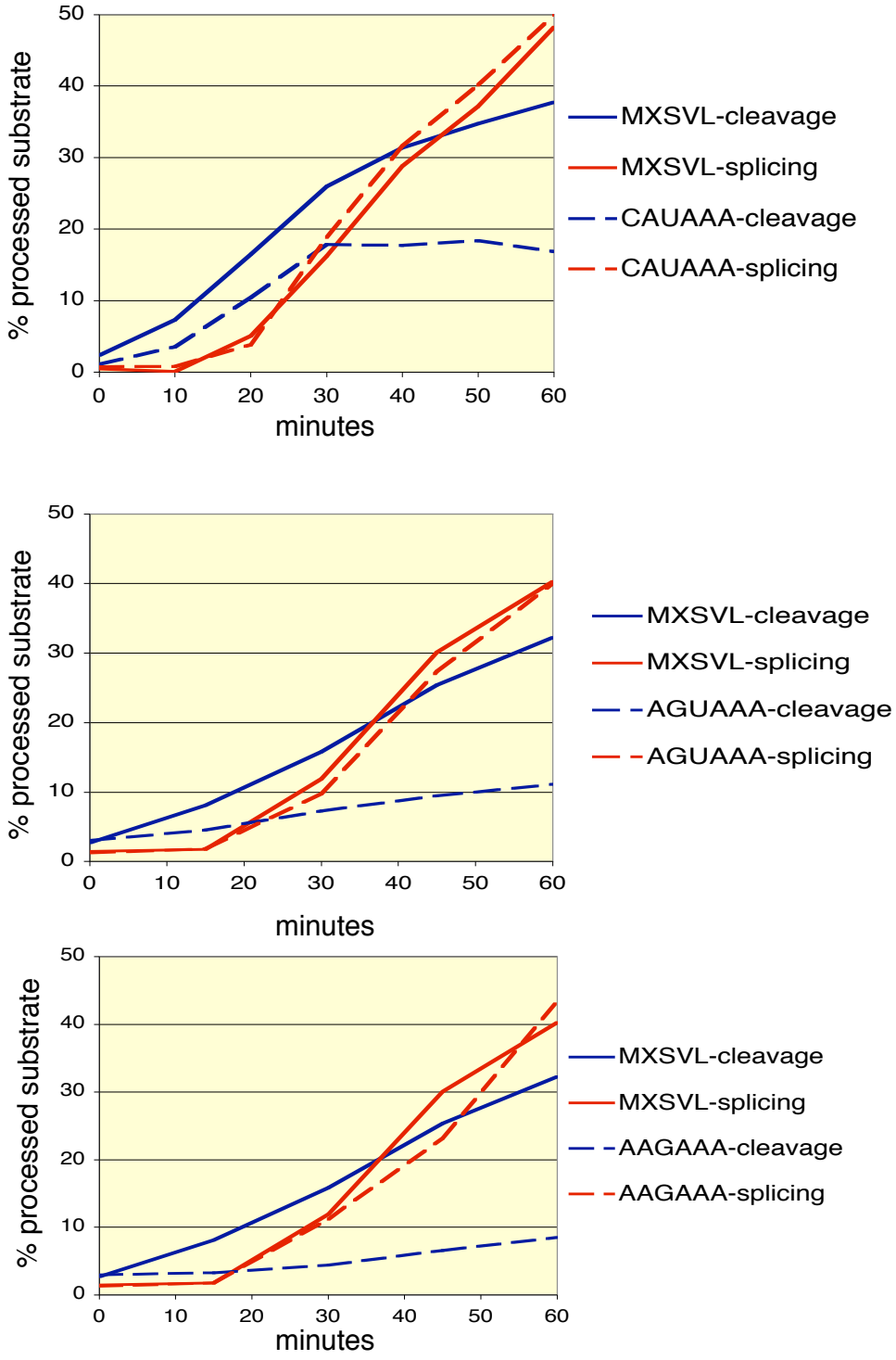


Figure 2.9 Mutations in AAUAAA affecte the 3' end processing but not the splicing activities

Quantitative evaluation of the splicing and 3' end processing activities on the different RNA substrates indicated at the right. For explanation of the calculation see legend to figure 2.4.

Niwa and Berget (1991) proposed that the coupling between 3' end formation and splicing is dependent on the $MgCl_2$ concentration in the reaction buffer. In order to find out if our AAGAAA mutation would display an effect on splicing activity under different $MgCl_2$ concentrations, we performed the assay with the $MgCl_2$ concentration ranging from 0.7 mM to 7.7 mM. However, under none of these conditions did the AAGAAA mutation influence the splicing efficiency (results not shown).

The MXSVL RNA substrate contains three USEs upstream of the AAUAAA hexanulceotide sequence. The RNA substrate that (Niwa and Berget, 1991) used in their assay was a shorter version of the MXSVL RNA substrate and lacked the first and half of the second USEs. Cooke and coauthors (1999) only tested the AAGAAA mutation in combination with a deletion of all three USEs. We therefore assume that the AAGAAA mutation only abolishes the cleavage activity on the MXSVL RNA in combination with a deletion of all or parts of the USEs. CPSF160 has been shown to bind to the USE of the HIV1 pre-mRNA (Gilmartin et al., 1995). It is not known if CPSF binds also to the USE of the MXSVL pre-mRNA. However, it is intriguing that this might be the case. If so, this could explain why we still observed some cleavage activity with the AAGAAA mutation whereas the two other groups did not. If CPSF binds to the USE, cleavage at aberrant cleavage sites could take place. Furthermore the presence of CPSF on the pre-mRNA could be sufficient to allow the coupling of splicing and 3' end processing which would explain why our AAGAAA RNA substrate was spliced efficiently.

2.6.1.2 Deletion of the U1 and U4 snRNA resulted in impaired 3' end cleavage

In addition to the U2 snRNA we depleted nuclear extract of the U1 and U4 snRNAs with RNaseH and DNA oligonucleotides complementary to the snRNAs. We tested these extracts in the coupled splicing and 3' end processing assay (Figure 2.10A). The specific digestion of the individual snRNAs was controled by staining the gel shown in the upper panel with ethidiumbromide (lower panel). The U1 snRNA was partially digested and the digestion product is indicated as U1*. The U4 snRNA stained very weakly with ethidium bromide. However, a faint band was visible which disappeared

in the extract that was treated with RNaseH and a DNA oligo complementary to the U4 snRNA. We do not know what the faint band represents which was running at approximately the hight of the U1* digestion product in the extract treated with the U4 DNA oligonucleotide

Figure 2.10.B shows the quantitative evaluation of two independent experiments. Depletion of the U1 snRNA resulted in an inhibition of splicing of approximately 50% and in a reduction of cleavage activity of only 25% compared to the untreated nuclear extract after one hour incubation at 30°C. Depletion of the U4 snRNA abolished the splicing activity and resulted in a reduction of 3' end cleavage of approximately 50% after one hour.

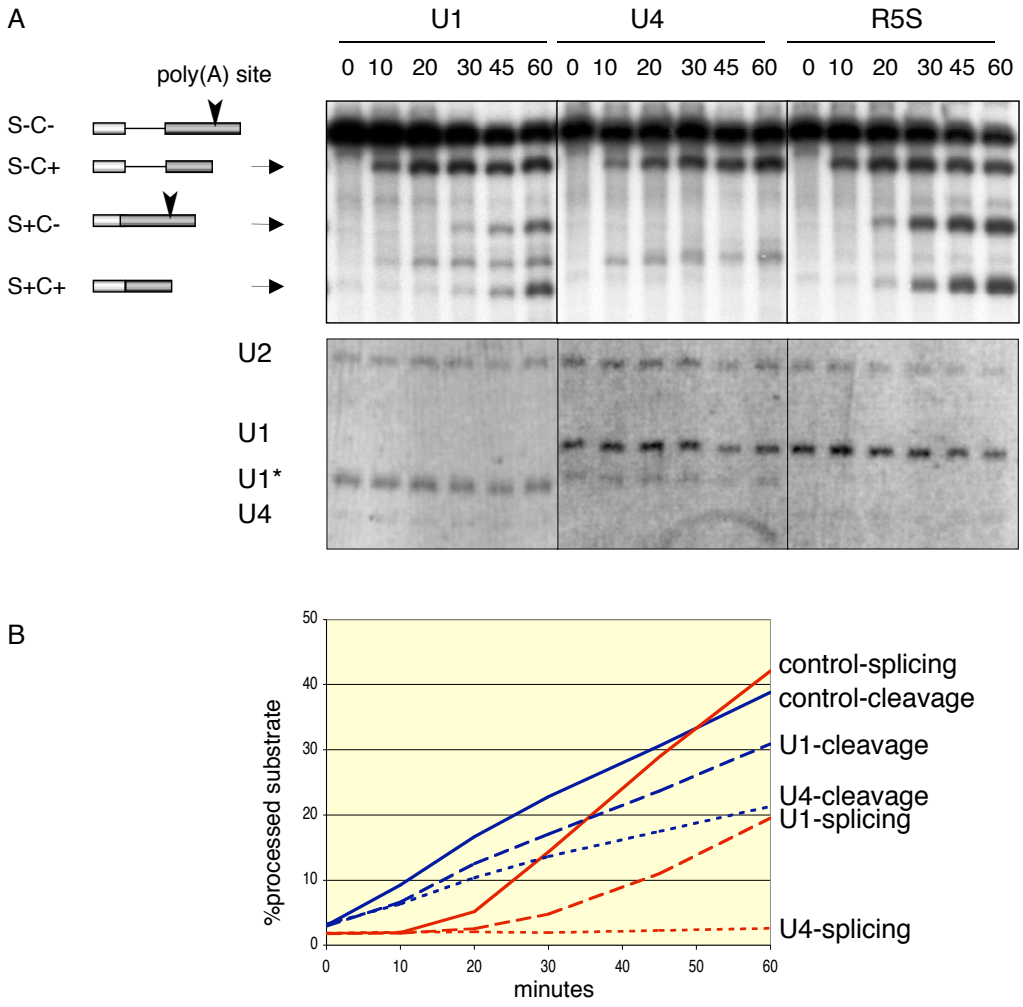


Figure 2.10 Digestion of the U1 snRNA and U4 snRNA resultes in decreased splicing and 3' end cleavage activity

A) The upper panel shows a time course coupled splicing and 3' end cleavage assay with extract treated with DNA oligonucleotides complementary to the U1 snRNA, U4 snRNA or 5S rRNA and Rnase H. For explanation of the migration positions of the RNA products see legend to figure 2.3.. The lower panel shows an ethidium bromide staining of the upper gel. To the left the migration position of the U1, U2 and U4 snRNAs are indicated. U1* indicates the migration position of the digestion product resulting from digestion of the U1 snRNA with RNase H and the oligonucleotide complementary to the U1 snRNA.
B) Quantitative evaluation of the splicing and 3' end cleavage activity. The graph shows the average of two independent experiments. For explanations to the calculations see legend to figure 2.4..

The U1 snRNP has previously been shown to be involved in coupling 3' end formation and splicing (Wassarman and Steitz, 1993; Gunderson et al., 1994; Lutz et al., 1996; Gunderson et al., 1998) and it was surprising that deletion of this snRNA did not result in a more dramatic effect on 3' end processing. However, given that the splicing activity was not abolished, it is possible that the U1 snRNA was not fully digested in the nuclear extract and the remaining level of intact U1 snRNPs could have been sufficient to maintain a cleavage activity of 75% compared to the untreated nuclear extract after one hour incubation at 30°C. On the other hand the U1 snRNP was shown to be mainly involved in the polyadenylation step of the 3' end prcessing reaction. Raising the possibility that this snRNP is not involved in coupling the splicing and 3' end cleavage reaction. It would be interesting to investigate the effect of U1 snRNA depletion on the polyadenylation activity in the future.

So far the U4 snRNP has not been shown to be involved in coupling 3' end processing and splicing. We found a strong effect on 3' end cleavage activity upon depletion of the U4 snRNA. This raises the possibility that this snRNP could also be directly involved in coupling 3' end processing and splicing. However, given that the different snRNPs interact with and influence each other, we cannot exclude that the effect of the U4 snRNA depletion on 3' end cleavage activity was an indirect effect, and future work is required to investigate if the U4 snRNP indeed plays a direct role in coupling 3' end processing and splicing.

Two-hybrid interactions

In the beginning of the project we were interested to find new interaction partners of the 3' end processing machinery with proteins involved in other reaction mechanisms. Therefore we performed two hybrid screens in collaboration with Hybrigenics S.A. (Paris) with CPSF100, CPSF73 and hClp1 as baits. To a later time point we wanted to investigate with which proteins the U2 snRNP subunits SF3b155, SF3b130 and SF3b49 interact. The results of the different screens are summarized below. Unfortunately the screen with SF3b130 did not provide any protein-protein interactions.

The number of hits show how often a certain protein was fished out during the screen. The score spans from "A" to "E" where "A" indicates that this interaction is very likely to be a true interaction, whereas interactions with an "E" score might not be.

Bait: CPSF100

Numbers of hits and score	Interacting proteins
3xB	**Methyl-Cpg binding domain protein 1(hMBD1):** Binds specifically to methylated DNA; can repress transcription from methylated gene promoters.
2xB	**SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily E, member 1 (SMARCE1):** Subunit of the BRG1-associated factors (BAF) which is thought to facilitate transcriptional activation of specific genes by antagonizing chromatin-mediated transcriptional repression. The complex contains an ATP-dependent nucleosome disruption activity that can lead to enhanced binding of transcription factors.
2xB	**Metal-regulatory transcription factor (MTF1):** Transcription factor that

	binds to metal-responsive promoter elements and is required for transcription of metallothioneins.
2xC	**Transcription termination factor, RNA polymerase I (hTTF1):** Transcription termination by RNA polymerase I is largely the result of binding of a transcription termination factor to a DNA element called a transcriptional terminator. TTF1 binds to the transcription terminatior sequences and thereby promotes transcription termination.
2xD	**Musculin / activated B-cell factor 1 (hMSC):** Contains the basic helix—loop-helix motif (bHLH); binds DNA and is involved in transcription repression.
1xD	**Transcription enhancer factor 5 (TEF5):** Binds to multiple functional sites in the CSH2 enhancer. Mutations in these sites that disrupted TEF5 binding also inactivated the transcription-enhancing activity of the element.
1xD	**Zinc finger protein 350 (hZNF350):** Interacts with BRCA1 which is implicated in the transcriptional regulation of DNA damage-inducible genes that function in cell cycle arrest. ZNF350 represses transcription in a BRCA1-dependent manner
1xD	**Hypothetical protein FLJ10713**
1xD	**Hypothetical protein HSPC195**
1xD	**Similar to hypothetical protein FLJ25467**
2xE	**hHairy-related transcription factor 3 (hHEYL):** Contains the basic helix-loop-helix (bHLH) and orange domains conserved in HEY proteins.
1xE	**Hairy/enhancer of split-related with YRPWmotif 1 (hHEY1):** Transcription factor implicated in cell fate decisions. It contains the basic helix-loop-helix motif (bHLH). HEY genes are direct transcriptional targets of the Notch signaling pathways in Drosophila and vertebrates.

It was striking how many proteins involved in transcription were found to interact with CPSF100. Suggesting that CPSF100 might play a role in transcription initiation, elongation and/or termination. It would be interesting in the future to verify which of the above interactions are indeed true interactions and investigate their functional

relevance. Interestingly, the protein SMARCE1 (a subunits of the BAF complex) was found in this two hybrid-screen but did also co-purify with CPSF (Figure 2.1.). Furthermore, the SMARCB1 protein also a subunits of the BAF complex was identified in purified CPSF. This indicates that CPSF and the BAF complex might indeed be associated with each other and it would be intriguing to analyze the functional role of this association.

Bait: CPSF73

Numbers of hits and score	Interacting proteins
64xA	**hSymplekin:** homologue of Pta1p. Was shown to interact with CstF and is also associated with CPSF. The protein was shown to be involved in 3' end processing.
3xB	**Human Plectin (PLEC1):** Interlinks Intermediate Filaments with Microtubles and Microfilaments and anchors Intermediate Filaments to Desmosomes or Hemidesmosomes. Could also bind actin to membrane complexes in muscles. Involved in crosslining, stabilization and regulation of the cytoskeletal intermediate Filaments network and possibly in the regulation of their dynamics. Loss causes epidermolysis bullose with muscular dystrophy.
2xC	**Proteasome 26S subunit, ATPase, 6 (hPSMC6):** The 26S protease is involved in the ATP-dependent degradation of ubiquitinated proteins. The regulatory (or ATPase) complex confers ATP dependency and substrate specificity to the 26S complex.
6xD	**Pleckstrin homology domain-containing protein, family M, Member 2 (PLEKHM2):** Binds to the Salmonella SifA protein localized to the cytosol in HeLa cells and was partially recruited to membranes upon ectopic expression of SifA.
2xD	**Step II splicing factor SLU7:** associates with the spliceosome relatively late in the assembly process. The association of SLU7 with the spliceosome is coupled to the catalysis of step 1. Depletion of SLU7

	resulted in the accumulation of novel intermediate C-type spliceosomal complexes that were blocked before step 2.
1xD	**hAXIN2:** regulates stability of beta-catenin an adherens juncion protein. Possibly involved in DNA mismatch repair
1xD	**DnaJ (Hsp49) homolog, subfamily C, member 3 (DNAJc3):** Inhibitor of the interferon-induced, dsRNA-activated protein kinase.
1xD	**Metal- regulatory transcription factor 2 (MTF2):** Binds to metal response elements in a zinc dependent manner..
1xD	**Steroid hormone receptor Ner-I (hNR1H2):** Nuclear receptor, binds to specific response elements upon activation.
1xD	**Transcription factor-like 4 (hTCFL4):** The protein contains basic helix-loop-helix and leucine zipper domains. It is a transcription factor with a role in proliferation, determination and differentiation.
1xD	**Hypothetical protein Hpp13624**
1xD	**Hypothetical protein DKFZp762L0311**
1xD	**Hypothetical protein FLJ10713**
6xE	**hHairy-related transcription factor 3 (hHEYL):** contains the basic helix-loop-helix and orange domains conserved in HEY proteins

Symplekin was previously shown to be associated with CPSF (Hofmann et al., 2002) however, a direct interaction was not known. The strong two hybrid interaction of symplekin with CPSF73 suggests that CPSF73 directly interacts with symplekin. Given that symplekin is not thought to be an integer part of CPSF (Hofmann et al., 2002) (Isabelle. Kaufmann, dissertation 2002) it is possible that this interaction could have a regulatory effect on CPSF73.

Another interesting interaction in the context of coupling 3' end processing and splicing was the finding that hSLU7 interacted with CPSF73. However to this point we have not investigated whether this interaction plays a role in the coupling mechanism.

Bait: hClp

Numbers of hits and score	Interacting proteins
36xA	**ADP-Ribosylation factor 1 (ARF1):** Small guanine nucleotide-binding protein. Guanine exchange-activated ARF1 at the Golgi membrane recruits and binds cytoplamic COP1 to the membranes.
24xA	**MET Protooncogene (MET):** MET is in the tyrosine kinase receptor family of oncogenes. Localized at the cell-surface. Receptor for hepatocyte growth factor (HGF).
9 x A	**Heat-Shock 27-kD Protein 1 (hHSPB1):** Synthesis has been shown to be correlated with the acquisiton of thermotolerance. Approximately 20% of its residues are susceptible to phosphorylation. The MAPKAPK5 stress activated kinase was shown to phosphorylate HSPB1.
3xB	**Protein Kinase, cAMP-dependent catalytic, Inhibitor Beta (PKIB):** May interact with the catalytic subunit of cAMP-dependent protein kinase and act as a competitive inhibitor.
2xD	**Tyrosine-protein kinase receptor Tie-1 (hTIE1):** Probable protein tyrosine-kinase transmembrane receptor.
2xD	**Gamma-Actin (ACTG):** Actins are highly conserved proteins that are involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells.
1xD	**Cyclin dependent kinase 2 (hCDK2):** Might play a unique role in cell cycle regulation of vertebrate cells. Probably involved in the control of the cell cycle. Interacts with cyclins A, B3, D, or E. Activity of CDK2 is maximal during S phase and G2.
1xD	**Chorionic somatomammotropin hormone-like 1 (hCSHL1):** May be a novel gestational hormone required to compensate for absence of other members of the GH/CS cluster during gestation.
1xD	**Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48**

	kDa subunit (hDDOST): Essential subunit of N-oligosaccharyl transferase enzyme which catalyzes the transfer of a high mannose oligosaccharide to an asparagine residue within an Asn-X-Ser/Thr consensus motif in nascent polypeptide chains.
1xD	**Solute carrier family 9 (sodium/hydrogen exchanger), isoform 9 (SL9A3):** Involved in pH regulation to eliminate acids generated by active metabolism or to counter adverse environmental conditions. Major proton extruding system driven by the inward sodium ion chemical gradient. Plays an important role in signal transduction.
1xD	**NK-tumor recognition protein (hNKTR):** Component of a putative tumor-recognition complex. Involved in the function of NK cells.
1xD	**Rho GDP-dissociation inhibitor 1 (ARHGDIA):** Regulates the GDP/GTP exchange reaction of the Rho proteins by inhibiting the dissociation of GDP from them, and the subsequent binding of GTP to them.
1xE	**Mortality factor 4-like protein 1 (hMRG15):** Component of the NuA4 histone acetyltransferase (HAT) complex, which is involved in transcriptional activation of selected genes principally by acetylation of nucleosomal histone H4 and H2A. This modification may both alter nucleosome - DNA interactions and promote interaction of the modified histones with other proteins which positively regulate transcription.
3xN/A	**Cat eye syndrome chromosome region, candidate 1 (hCECR1):** May be involved in growth regulation through an adenosine deaminase activity.
1xN/A	**Triggering receptor expressed on myeloid cells-like 2 (TREML2):** hypothetical protein

hClp1p interacted with a wide range of proteins involved in different processes. It is a component of CF II_m and of the tRNA splicing endonuclease complex (see Chapter 5) and it might be a subunit of even more complexes, which would explain, why it interacted with such a broad number of proteins involved in different pathways.

Like ARF1 hClp1 was indicated to bind single nucleotides. A future project could involve to investigate if the interaction between the two has any significant relevance.

Bait: SF3b49

Numbers of hits and score	Interacting proteins
145xA	**Splicing factor 3b 145 kDa (hSF3b145):** Subunit of the Splicing factor 3b and the U2 snRNP. Cross-links to pre-mRNA at a 20-nucleotide region, the anchoring site, upstream of the branch site. Interacts with SF3b145.
15xA	**Zinc finger CCHC domain-containing protein 8 (hZCCHC8)**
11xA	**NEDD8 ultimate buster-1 (hNYREN18):** Specific down-regulator of the NEDD8 conjugation system. Recruits NEDD8 and its conjugates to the proteasome for degradation.
10xA	**NEDD4-like E3 ubiquitin-protein ligase (hWWP1):** E3 ubiquitin-protein ligase, which accepts ubiquitin from an E2 ubiquitin-conjugating enzyme in the form of a thioester and then directly transfers the ubiquitin to targeted substrates.
2xC	**NEDD4-like E3 ubiquitin-protein ligase (hWWP2):** WWP2 contains 4 tandem WW domains, a complete HECT (homologous to the E6-associated protein carboxyl terminus) domain, associated with ubiquitin-protein ligase activity, and a C2 (calcium-dependent phospholipid-binding)-like domain characteristic of a large family of proteins including protein kinase C.
2xD	**Hypothetical protein hDKFZp77901**
2xD	**Zinc finger protein 553 hFLJ31751**
2xD	**Hypothetical protein hLOC339324**
1xD	**Coatomer protein complex, subunit epsilon (hCOPE):** Subunit of the heptameric coatomer protein complex, or COPI, which mediates transport between the Golgi complex and endoplasmic reticulum.
1xD	**Golgin subfamily B member 1 (hGOLGB1):** May participate in forming intercisternal cross-bridges of the Golgi complex.

As has been shown before by (Champion-Arnaud and Reed, 1994) we found that SF3b49 interacted with SF3b145. The identified proteins were not very relevant in regards to the coupling of 3' end processing and splicing. The same was the case for the proteins found to interact with SF3b155.

Bait: SF3b155

Numbers of hits and score	Interacting proteins
4xA	**Smad nuclear-interacting protein 1 (hSNIP1):** SNIP1 interacts with several SMADs. SNIP1 may regulate access of transcription factors to the coactivators p300 and CBP in select cells and regulate the pattern of gene expression in embryogenesis.
1xD	**Nuclear inhibitor of protein phosphatase 1 (hPPP1R8):** Inhibitor subunit of the major nuclear protein phosphatase-1 (PP-1). It has RNA-binding activity but does not cleave RNA and may target PP-1 to RNA-associated substrates. May also be involved in pre-mRNA splicing. Binds DNA and might act as a transcriptional repressor. Seems to be required for cell proliferation.
1xD	**Cytochrome P450 19A1 (hCYP19):** Catalyzes the formation of aromatic C18 estrogens from C19 androgens.
1xD	**Keratin, type I cytoskeletal 23 (hKRT23)**
1xD	**Putative helicase MOV-10 (hMOV10):** May be a helicase with an important function in development and/or control of cell proliferation.
1xE	**Myosin-9 (hMYH9):** Cellular myosin that appears to play a role in cytokinesis, cell shape, and specialized functions such as secretion and capping.
2xE	**26S proteasome non-ATPase regulatory subunit 11 (hPSMD11):** Acts as a regulatory subunit of the 26S proteasome which is involved in the ATP-dependent degradation of ubiquitinated proteins.

2.6.2. Experimental procedures

2.6.2.1. Mutant AAUAAA RNA substrates

Mutations in the AAUAAA hexanucleotide were generated with site directed mutagenic PCR, followed by DpnI digestion of the parent plasmid. The PCR products were transformed into *E.coli* for ligation and amplification. The sequence was verified by sequenzing.
The coupled assay was performed as described in 2.5.8.

2.6.2.2. Oligodeoxynucleotide-directed RNase H cleavage

The assay was performed as described previously except that the U1 DNA oligodeoxynucleotide was used in a final concentration of 10 μM. The sequences of the U1 and U4 oligos are described in (Frendewey et al., 1987).

2.6.2.3. Plasmids

CPSF73 and CPSF100 were cut out from the plasmids Htp-73 and Htp-100 and ligated into the pB27 vector (provided by Hybrigenics) using the restriction enzyme NotI. The sequences of hClp1, SF3b49, SF3b130 and SF3b155 were PCR amplified and ligated into pB27 using the restiction enzymes SpeI and PacI. hClp1 was PCR amplified with the primers hClp1-PacI (atattaattaagctacttcagatccatg) and hClp1-Spe1 (ccggttcatggatctgaagtag). SF3b49 was amplified with the primers SF3b49-SpeI (aagcgcactagtgatggctgccgggccgat) and SF3b49-PacI (caattcttaattaattactgagggagagggcct). The primers used to amplify SF3b130 were SF3b130-SpeI (cgcactagtgatgtttctgtacaac) and SF3b130-PacI (atattaattaagtcagaaggcgtagcg). SF3b155 was amplified with the primers SF3b155-speI (cgcactagtgatggcgaagatcgcc) and SF3b155- PacI (atattaattaagttataagatatagtc). The sequence of the inserts was verified by sequencing.

Chapter 3: Knock down of CPSF100, SF3b130 and SF3b155 by RNAi

3.1 Introduction **78**

3.2 Results **80**

3.2.1 Growth analysis of HeLa cells after knock down of CPSF100, SF3b130 and SF3b155 80

3.2.2 Western blot of total protein extracted from siRNA treated cells 82

3.2.3 The siRNA treatment lead to reduction of the targeted mRNAs 83

3.2.4 Knock down of SF3b subunits effected splicing *in vivo* 84

3.2.5 Kock down of CPSF100, SF3b130 and SF3b155 had no detectable effect on 3' end processing 85

3.3 Discussion **86**

3.4 Experimental procedures **88**

3.4.1 RNAi 88

3.4.2 Isolation of RNA and RT-PCR 89

3.4.3 Western blot analysis 90

3.1 Introduction

The 12S U2 snRNP consists of the seven Sm proteins common to the spliceosomal snRNPs and the U2-specific proteins U2-A' and U2-B''. In addition the splicing factors 3a and 3b (SF3a / SF3b) associate with the particle to form the fully assembled 17S U2 snRNP (Brosi et al., 1993; Krämer et al., 1999; Will and Lührmann, 2001). Relatively little is known about how these functionally important complexes are tethered to the U2 snRNP. It has been suggested that the U2 snRNP consists of two globular structures of which one contains SF3b and a 5' portion of the U2 snRNA, whereas the other contains the Sm core, U2-A, U2-B'', SF3a and the 3' part of the U2 snRNA (Behrens et al., 1993; Krämer et al., 1999). SF3a consists of the three proteins SF3a120, SF3a66 and SF3a60. SF3b contains the seven proteins SF3b155, SF3b145, SF3b130, SF3b49, SF3b14a/p14, SF3b14b and SF3b10 (Gozani et al., 1996; Das et al., 1999; Krämer et al., 1999; Will et al., 2001; Will et al., 2002). Subunits of SF3a and SF3b interact with pre-mRNA and are believed to play an essential role in splice site recognition and tethering of the U2 snRNP to the pre-mRNA (Gozani et al., 1996; Query et al., 1996; Gozani et al., 1998; Will et al., 2001). SF3b49, SF3b145, SF3b155 and all three SF3a subunits can be cross-linked to a 20 nucleotide region just 5' of the branch site in pre-spliceosomes and spliceosomes (Gozani et al., 1996). SF3b155 also binds to a region 5 nucleotides downstream of the branch point (BP; Gozani et al., 1998). SF3b155 interacts with SF3b14a/p14, which directly binds to the BP adenosine (Will et al., 2001). Additional known protein-protein interactions between SF3a and SF3b subunits are between SF3b155 and SF3b130 (Das et al., 1999), SF3b145 and SF3b49 (Champion-Arnaud and Reed, 1994), SF3a120 and SF3a60 (Chiara et al., 1994), and SF3a120 and SF3a66 (Krämer et al., 1999).

SF3b155 also interacts with U2AF65 and this interaction replaces the U2AF65-SF1 interaction and contributes to the recruitment of the U2 snRNP to the BP (Gozani et al., 1998). The C-terminus of SF3b155 contains 22 nonidentical, tandem repeats, which are similar to those found in the regulatory subunit A of the phosphatase PP2A. However, no phosphatase activity has been reported for SF3b155 so far. Interestingly, the repeats must serve an important function given that the *S. cerevisiae* homologue of SF3b155 consists almost entirely of them. The N-terminus is

less well conserved and contains multiple TPGH motifs and several RWDETP motifs (Wang et al., 1998). SF3b155 is phosphorylated during or just after step 1 but before step 2 of the splicing reaction. And this phosphorylation might be important for activity of the spliceosome (Wang et al., 1998). Cyclin E-cdk2 was shown to phosphorylate SF3b155 and antibodies directed against cyclin E immunoprecipitated the U2 snRNA and subunits of SF3b. Cyclin E-cdk2 is a critical regulator of cell cycle progression from G1 into S phase in mammalian cells. These results suggested that pre-mRNA splicing might be linked to the cell cycle machinery through phosphorylation of SF3b155 (Seghezzi et al., 1998).

SF3b130 was identified as component of SF3b (Das et al., 1999). However, very little is known about the function of the protein. The yeast homologues of SF3b155 and SF3b130 were shown to be essential proteins (Das et al., 1999), indicating that both proteins play an important role in mammals also. Interestingly, SF3b130 shows sequence homology with CPSF160 (Das et al., 1999). A domain assigned to bind RNA in CPSF160 is conserved in SF3b130, suggesting a role of the protein in nucleic acid binding (Dichtl et al., 2002b). However, so far SF3b130 has not been shown to interact with RNA.

CPSF100 shares homology with CPSF73 (Jenny et al., 1996). Like CPSF73 the 100 kDa subunits is a member of the metallo-β-lactamase / β-CASP family but unlike CPSF73, CPSF100 is not believed to posses endonuclease activity as it lacks part of all the conserved amino acids that are thought to be involved in the enzymatic function (Callebaut et al., 2002). The function of CPSF100 is largely unknown. It can be UV cross-linked to RNA (Edwalds-Gilbert and Milcarek, 1995), however RNA-binding experiments did not confirm a function in RNA binding (Bernhard Dichtl unpublished data).

Homologues of CPSF100 and CPSF73 have recently been identified in mammalian cells. The N-terminus of RC-74 is about 20% identical to CPSF100. Like CPSF100 the protein contains mutations in the amino acids thought to be involved in endonuclase activity. RC-68 is 40% identical to CPSF73 throughout the first 450 amino acids. It contains the metallo-β-lactamase / β-CASP motifs which have an intact active site. RC-68 localizes to the cytoplasm and the nucleus whereas RC-74 is found exclusively in the nucleus. RC-74 interacts with RC-68 but not with CPSF73 and none of the proteins was found to interact with CPSF160. Depletion of RC-68 by RNAi arrested HeLa cells eary in G_1 phase, but surprisingly the arrested cells continued to grow and reached the size typical for G_2 cells (Dominski et al., 2005b).

We were able to show *in vitro* that the U2 snRNP and CPSF couple splicing and 3' end processing and wanted to test whether we can detect this effect *in vivo* also. Therefore, we depleted SF3b155, SF3b130 and CPSF100 by RNAi in HeLa cells and tested for deficiency in splicing and 3' end processing in these cells. We were able to show that both SF3b subunits are required for splicing *in vivo*. However, neither SF3b155, SF3b130 nor CPSF100 knock down resulted in an observable influence on 3' end processing. This lack of effect can be explained by different scenarios as described in the discussion of this chapter.

Interestingly, knock down of both SF3b subunits resulted in decrease of cell viability whereas depletion of CPSF100 did not significantly affect cell viability. This could indicate that CPSF100 shares redundant functions with another protein in the cell.

3.2 Results

3.2.1 Growth analysis of HeLa cells after knock down of CPSF100, SF3b130 and SF3b155

In order to test if CPSF100, SF3b130 and SF3b155 are essential proteins and to investigate their roles in pre-mRNA processing we took advantage of the RNAi method. We treated HeLa cells independently with two different siRNAs complementary to the mRNAs of CPSF100, SF3b130 and SF3b155.

We determined the percentage of dead cells caused by the treatment with the siRNAs by staining the cells with trypan blue. Figure 3.1 shows the percentage of dead cells two and three days after transfection of the individual siRNAs as indicated at the bottom. Knock down of CPSF100 with the siRNA CPSF100-1 resulted in cell death of only 33% with a standard deviation (SD) of 6 three days after transfection. Knock down of the same protein with the CPSF100-2 siRNA reslted in a cell death of 29% (SD=2) three days after transfection. In comparison, knock down of the unessential gene lamin resulted in 20% (SD=5) cell death. Treatment of HeLa cells with siRNAs directed against SF3b130 resulted in cell death of 44% (SD=9) with the siRNA SF3b130-1 and 41% (SD=6) with the SF3b130-2 oligo. Knock down of

SF3b155 lead to 49% (SD=5) cell death three days after tranfection with the siRNA SF3b155-1 and 45% (SD=8) with the SF3b155-2 RNA oligo.

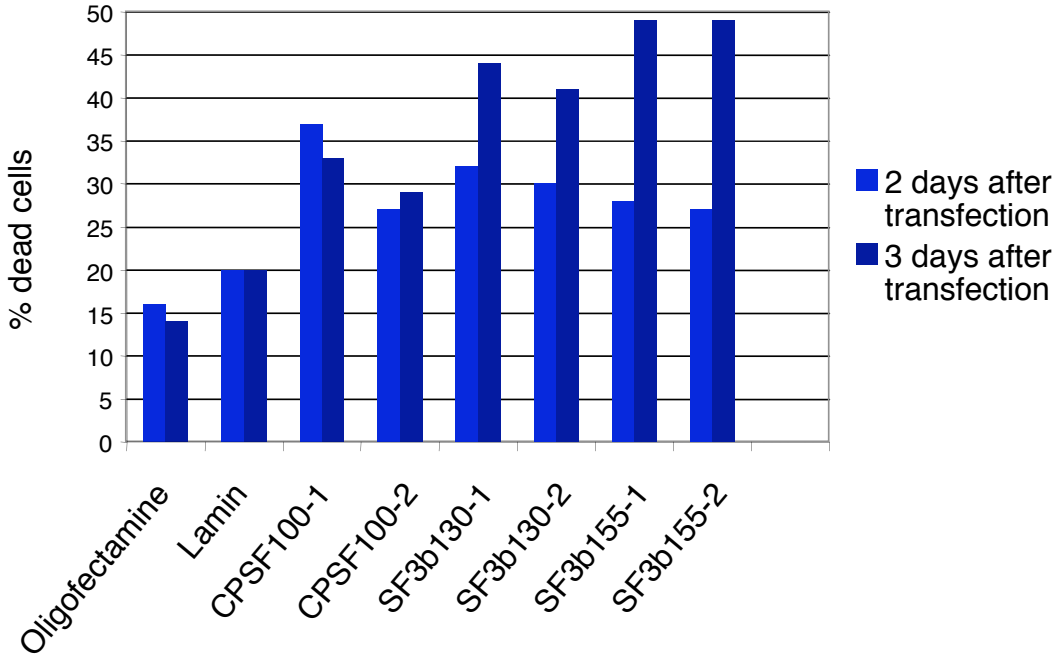


Figure 3.1 Percentage of dead cells after knock down of CPSF100, SF3b130 and SF3b155

The graph represents the percentage of dead cells 2 and 3 days after treatment with the individual siRNAs (indicated at the bottom). The graph shows the average of four independent experiments.

3.2.2 Western blot of total protein extracted from siRNA treated cells

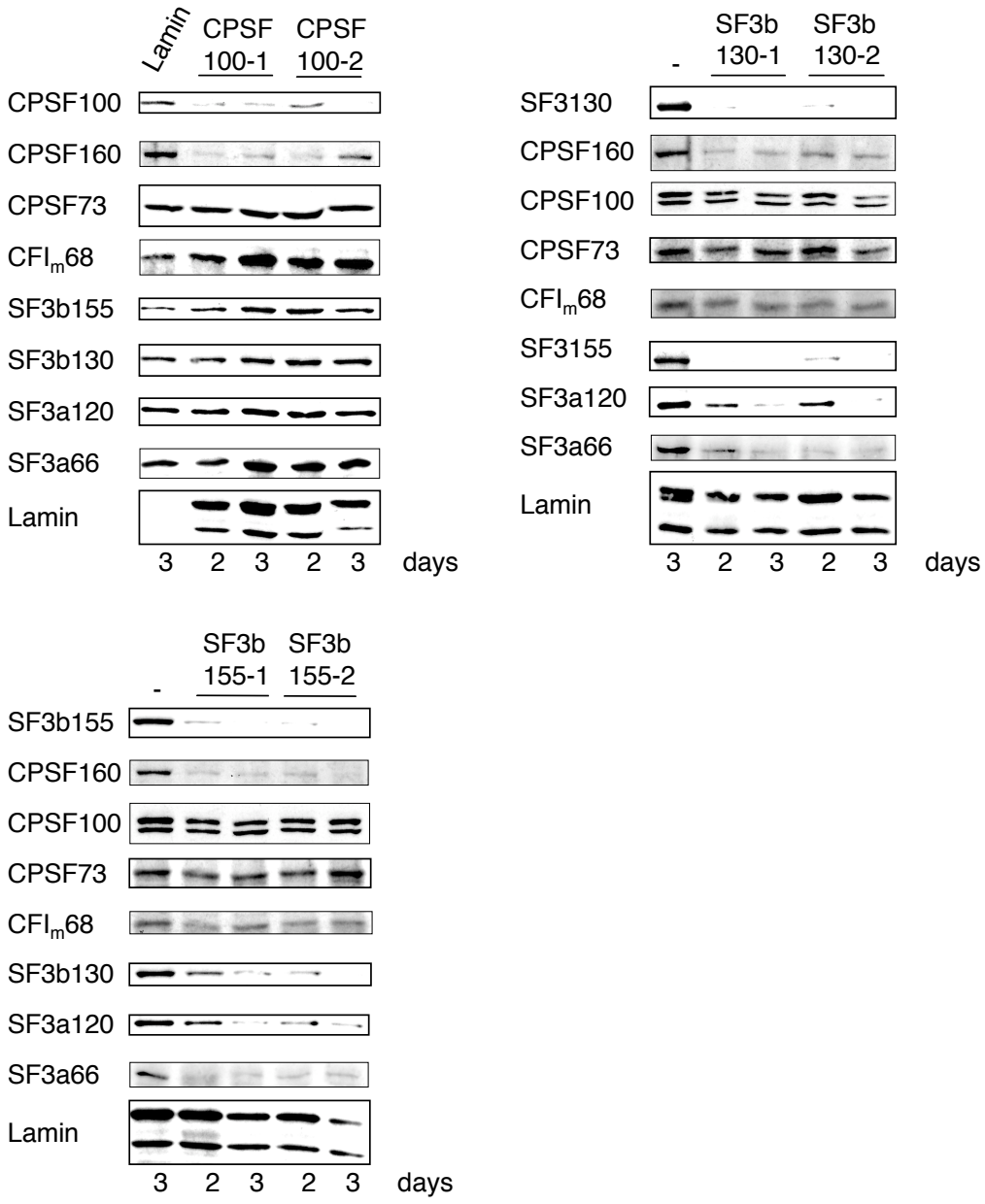


Figure 3.2 Western blot analysis of the siRNA treated cells

Total proteins were extracted from cells harvested 2 or 3 days (indicated at the bottom) after transfection with the siRNAs indicated at the top. Equal amounts of totoal proteins were loaded in each lane. Lanes with a (–) represent total proteins extracted from cells that were mock transfected. The antibodies used for the Western blotting are indicated at the left. The levels of the Lamin protein served as loading control.

Cells treated with siRNAs directed against CPSF100 showed significantly lower amounts of CPSF100 already two days after transfection with both siRNAs used (Figure 3.2). The protein level of CPSF160 was reduced in these cells but not the protein levels of CPSF73, CF I_m68, SF3b155, SF3b130, SF3a120, SF3a66 and of lamin. Treatment of HeLa cells with siRNAs directed against SF3b130 or SF3b155 resulted in decrease of the targeted proteins as well as reduced levels of CPSF160, CF I_m68, SF3b155 repectively SF3b130, SF3a120 and SF3a66. CPSF100, CPSF73 and lamin protein levels were not affected.

3.2.3 The siRNA treatment lead to reduction of the targeted mRNAs

In order to test if the targeted mRNAs were indeed degraded in the siRNA transfected cells, we performed RT PCR experiments. The reverse transcription was performed with random primers in order to obtain the polyadenylated and non-polyadenylated RNA species or with oligo dT primers which lead to reverse transcription of the polyadenylated RNAs only. We did quantitative PCR on such produced DNA with primers complementary to the mRNAs of the targeted proteins. In all three cases, transfection of the HeLa cells with the respective siRNAs resulted in specific degradation of the targeted mRNA (Figure 3.3). As loading control served the intron-less H3F1 RNA. We obtained the same result for DNA produced with the random or with the oligo dT primers.

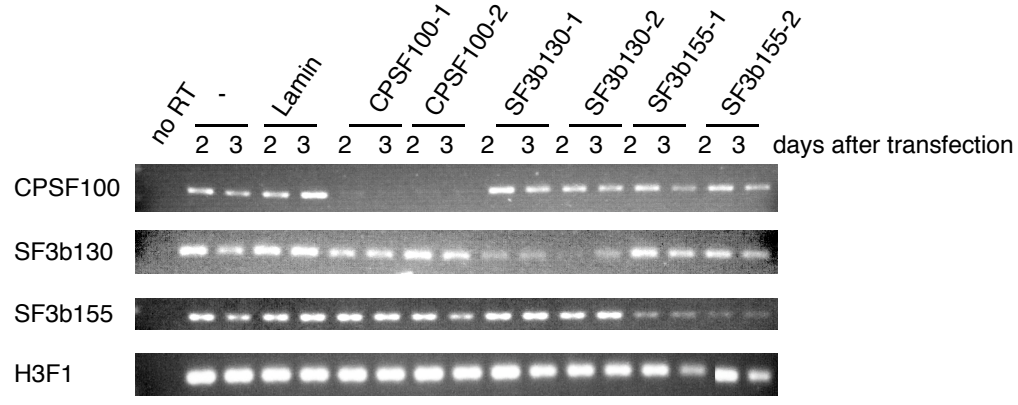


Figure 3.3 The siRNAs specifically reduce the targeted mRNA levels in HeLa cells

Total RNA was isolated from HeLa cells after 2 and 3 days incubation with the individual siRNAs (indicated on top). (-) represents reactions that were performed with RNA isolated from mock transfected cells. As control that no endogenous DNA was present in our preparations we used RNA isolated from HeLa cells on which no reverse transcription had been performed (no RT). The RNA was reverse transcribed with random primers and quantitative PCR was performed with primers complementary to the mRNAs indicated at the left. The PCR was carried out with 32 cycles. H3F1 served as an internal loading control.

3.2.4 Knock down of SF3b subunits affected splicing *in vivo*

To test whether the knocked down proteins are involved in splicing *in vivo* we did quantitative PCR on DNA obtained from reverse transcription with the oligo dT or the random primers, and primers complementary to two neighbouring exons (Figure 3.4). The last and an internal intron of the CPSF160 pre-mRNA and an internal intron of SF3a120 were not spliced out efficiently in SF3b130 and SF3b155 depleted cells, given that we detected the unspliced pre-mRNA species in these cells. We sequenced the PCR products and found that in the case of both CPSF160 introns the intron was not spliced out at all. The upper product of the SF3a120 PCR reaction was identified as partially spliced pre-mRNA that contained both exon borders and part of the intron. We were not able to detect unspliced pre-mRNA in CPSF100 knock down cells. These data show that both SF3b subunits are required for splicing *in vivo*.

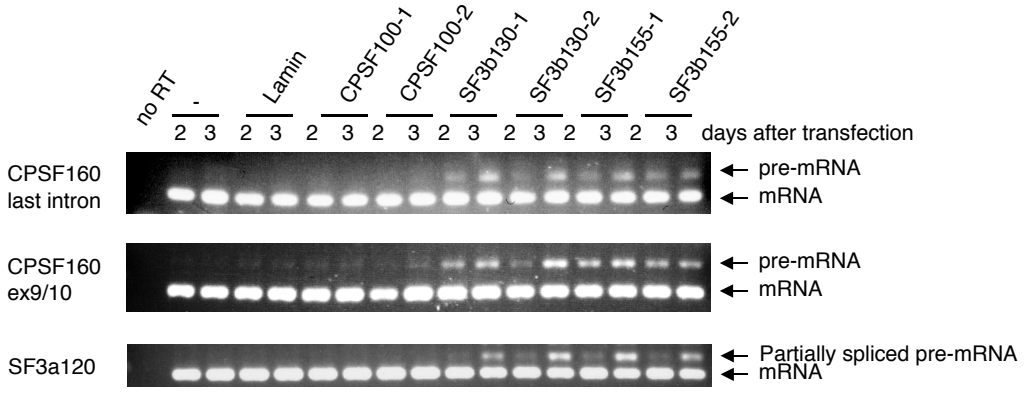


Figure 3.4 SF3b130 and SF3b155 are essential for splicing *in vivo*

Total RNA was isolated from HeLa cells after 2 and 3 days incubation with the individual siRNAs (indicated on top). (-) represents reactions that were performed with RNA isolated from mock

transfected cells. As control that no endogenous DNA was present in our preparations we used RNA isolated from HeLa cells on which no reverse transcription had been performed (no RT). The RNA was reverse transcribed with random primers and quantitative PCR was performed with primers complementary to adjacent exons of the pre-mRNA as indicated at the left. The PCR was carried out with 35 cycles.

3.2.5 Kock down of CPSF100, SF3b130 and SF3b155 had no detectable effect on 3' end processing

Deficiency in 3' end processing should likely lead to decreased levels of polyadenalytated mRNA. In order to test whether depletion of CPSF100, SF3b130 or SF3b155 resulted in deficient 3' end processing, we reverse transcribed the RNAs isolated from the siRNA treated cells with oligo dT primers. We did quantitative PCR on this DNA with primers annealing to the mRNAs of SF3b14 and CPSF73. In both cases we detected equal levels of mRNA in all cells (Figure3.5), indicating that knock down of CPSF100, SF3b130 or SF3b155 had no effect on the levels of polyadenylated mRNA. When we used 1/5 of the amount of DNA used in the PCR reactions we found that the PCR product was clearly reduced, indicating that the reaction was in linear range.

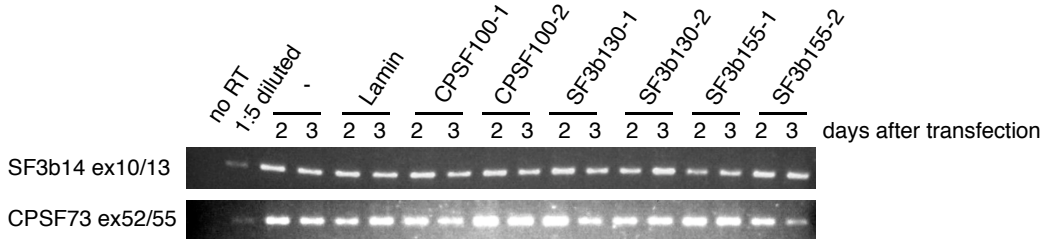


Figure 3.5 Depletion of CPSF100, SF3b130 and SF3b155 did not result in lower levels of polyadenylated mRNAs

Total RNA was isolated from HeLa cells after 2 and 3 days incubation with the individual siRNAs (indicated on top). (-) represents reactions that were performed with RNA isolated from mock transfected cells. As control that no endogenous DNA was present in our preparations we used RNA isolated from HeLa cells on which no reverse transcription had been performed (no RT). The RNA was reverse transcribed with oligo dT primers and quantitative PCR was performed with primers

complementary to adjacent exons of the pre-mRNA as indicated at the left. The PCR was carried out with 30 cycles..

3.3 Discussion

We have found that only approximately 30% of all cells died upon knock down of CPSF100. Depletion of the nonessential protein lamin resulted in cell death of 20% also, indicating that the transfection procedure caused a certain percentage of cell death independently of the siRNAs. We were surprised to find that knock down of CPSF100 did not result in a higher lethality given that Ydh1p (the yeast homologue of CPSF100) is an essential protein and 3' end formation is generally regarded as a vital process. One explanation for this observation could be that even though we knocked down CPSF100 quite efficiently as determined by Western blotting, the remaining levels of the protein could have been sufficient to allow survival of most cells. On the other hand it is possible that a protein exists within the cells, which shares redundant functions with CPSF100. Recently, RC-74 a homologue of CPSF100 has been identified. The function of this protein is so far unknown but it was shown to interact with the CPSF73 homologue RC-68 (Dominski et al., 2005b). It would be very interesting in the future to determine whether a double knock out of CPSF100 and RC-74 would cause lethality of the cells. However, we cannot exclude either that the function which CPSF100 carries out in a cell is indeed not very crucial for survival. Kock down of SF3b130 and SF3b155 lead to cell death of 40% respectively 45% three days after transfection with the individual siRNAs, clearly above the lethality level of the cells treated with the siRNAs directed against lamin. We therefore believe that both SF3b130 and SF3b155 are essential genes. The levels of the targeted proteins were clearly reduced as determined by Western blotting. However, a certain amount of the proteins could still have remained in the cells, allowing for the survival of a percentage of the cell population. Furthermore, not every single cell takes up siRNAs upon transfection. Therefore, there is always a pool of un-depleted cells present within the siRNA transfected cells.

The protein levels of CPSF160 were reduced in the CPSF100 depleted cells. Possibly this could mean that CPSF100 is involved of the processing of the CPSF160 pre-mRNA (or in its transcription/translation). However, given that we did not detect any

splicing or 3' end processing deficiency in these cells we think that this is rather unlikely. The low levels of CPSF100 present in the cells could however result in incompletely assembled and perhaps less stable CPSF complexes. This in turn could make the CPSF subunits more accessible to proteases. The yeast homologues of CPSF100 and CPSF160 have been shown to interact with each other. Maybe this interaction takes place in mammalian cells also and could be crucial for recruitment of CPSF160 to CPSF and therefore indirectly for the stability of the protein.

We have found that depletion of SF3b130 and SF3b155 resulted in deficient splicing. To our knowledge, this is the first time that it was shown *in vivo* that the two proteins are indeed involved in splicing. We wanted to test, if there are indications that CPSF and the U2 snRNP couple splicing and 3 end formation *in vivo*. If this was the case knock down of either CPSF or U2 snRNP subunits should result in inhibition of splicing as well as 3' end processing. However, depletion of CPSF100 did not result in deficient splicing. This could indicate that CPSF100 is not involved in splicing. However, it is also possible that knock down of CPSF100 was not efficient enough to cause a detectable defect in splicing. This would explain also, why a very low number of cells died upon depletion of CPSF100. On the other hand, it is possible that CPSF100 shares redundant functions with another protein, which could take over the role of CPSF100 in coupling splicing and 3' end processing when CPSF100 is depleted.

Depletion of neither CPSF100, SF3b130 nor SF3b155 resulted in an observable decrease of polyadenylated mRNA levels. As for CPSF100 this could be explained by the reasons mentioned above. It is possible that SF3b130 and SF3b155 are not involved in the 3' end processing of pre-mRNAs *in vivo*. On the other hand their effect could be to weak as to be detected by the method we used. It is also possible that the coupling of 3' end formation and splicing mediated by CPSF and the U2 snRNP is gene specific.

Taken together we were not able to show that the coupling between 3' end processing and splicing mediated by CPSF and the U2 snRNP takes place *in vivo* also. However, there are many reasons that could explain this lack of an effect. To address this question in the future, it would be necessary to look at the phenotypes caused by depletion of other subunits of CPSF and of U2 snRNP. It would also be advisable to

try more sensitive detection methods such as real time PCR and RNase protection assays.

3.4 Experimental procedures

3.4.1 RNAi

HeLa cells were seeded in 6-well plates at a concentration of 2.5*10^4 cells per well in MEM +Earls +GlutamMAX^{TM}I, 10% foetal bovine serum (all Gibco). The next day they were transfected with 225 pmol chemically synthesized, annealed siRNA (Dharmacon) in the presence of oligofectamine (Invitrogen) according to the manufacturers manual. The next day the medium was changed and the attached and detached cells harvested two and three days after transfection. For harvesting the following procedure was used: the medium was taken off and collected, the cells were washed with 1 ml of PBS and the wash was discarded. To detach the surviving cells from the plates they were incubated for 5 min. with 250 µl Trypsin and the cells combined with the medium from the first step. 50 µl of this cell suspension was given into a separate tube to determine the cell density. The rest of the suspension was centrifuged for 10 min. at 4°C and 3000 rpm. the supernatant was removed and the cells washed once in 1 ml cold PBS. The cell pellet was frozen in liquid nitrogen and stored at -80°C.

The cell viability was determined by staining the cells with trypan blue.

The sense sequences of the siRNA used in this study were as follows:

Lamin: CTGGACTTCCAGAAGAACAdTdT

CPSF 100-1: CTATGAAGGACGCTCTGATdTdT

CPSF 100-2: TATGATAGGTGGAACAATAdTdT

SF3b 130-1: AGCTCTTCCGAGTCCGAATdTdT

SF3b 130-2: AGGGCGAGCCGTTATGATTdTdT

SF3b 155-1: GATCGCCAAGACTCACGAAdTdT

SF3b 155-2: GCATAGGCGGACCATGATAdTdT

3.4.2 Isolation of RNA and RT-PCR

The RNA was isolated 2 and 3 days after transfection of the siRNAs with the RNeasy Mini kit (Qiagen). The DNA was digested on the column (Qiagen) following the manufacturers protocol.

Reverse transcription was performed with 0.2 μg total RNA, 200 ng random primer (Promega) or 5.6 μM oligo dT with the Super Script II reverse transcriptase (Invitrogen) in the presence of 20 U RNaseOUT (Invitrogen) at 42°C following the manufacturers protocol.

PCR reactions were carried out with the Expand high-fidelity PCR system (Roche). We used 1 µl of a 1:30 dilution of the DNA obtained in the reverse transcription. 10 µl [10 mM] dNTPs, 30 µl [10 mM] primer 1, 30 µL [10 mM] primer 2, 490 µl H_2O, 100 µl Buffer 3, 10 µl Enzyme and 60 µl [25 mM] $MgCl_2$ was combined and 24 µl added to the DNA. The PCR was carried out with the following program:

Step1: 95°C, 2 min.

Step 2: 95°C, 40 sec.

Step 3: 50°C, 1 min.

Step 4: 68°C, 4 min.

Step 5: 68°C, 10 min

For how many times the steps 2 to 4 were repeated is indicated in the legends to the different figures.

The primer pairs we used in this study were:

CPSF160-ex9 / CPSF160-ex10: aagctcgagcttgctgcc / agctcaggctcctcaaag

CPSF160-ex60 / CPSF160-ex61: acctgtacctgagcacca / gttttgtacaaaaaggggtg

CPSF100-ex24 / CPSF100-ex25: cctggacatcagtcagtttt / gtaagtttttctggcagattc

CPSF73-ex52 / CPSF73-ex55: catatttggagaagactgtg / ggtaacaaaagtagagtcaag

H3F1-primer1 / H3F1-primer2: gtgaaatccgccgctatcag / gtttaagttggcgcacaccc

SF3a120-ex16 / SF3a120-ex18: cttgattccacccaaaggtt / cattgggttggaagtccact

SF3b155-ex49 / SF3b155-ex51: cccaatgtatttgagacatc / gtcactactggcatttactg

SFb130-ex53 / SF3b130-ex55: gtgatcatgaactaccatgt / gtaggagcgaaagctgaggt

SF3b14-ex10 / SF3b14-ex13: gaacacacctgaaactagag / gtttcaagcaggtcattttg

3.4.3 Western blot analysis

The frozen cell pellets were thawed on ice and resuspended in 50 µl RIPA buffer (50 mM Tris-HCl pH8.0, 150 mM NaCl, 1% NP40, 0.5% sodiumdeoxycholate, 0.1% SDS, 0.2 mM Na_3VO_4, 10 mM NaF, 0.4 mM EDTA, 10% glycerol, before use the following solutions were add to a final concentration of 0.2 mM Vanadate, 0.5 mM DTT, 0.5 mM PMSF, 2 µg/ml Pepstatin, 2 µg/ml Leupeptin. The cells were vortexed quickly and put on a shaker at room temperature for 5 to 10 min. The suspension was centrifuged for 5 min at 4°C and 2000 rpm and the supernatant transferred to a fresh tube. For Western blot analysis $3x10^4$ cells were run on an 8% SDS-gel.

Chapter 4: Rse1p and its potential role in coupling splicing and 3' end processing in yeast

4.1 Introduction **92**

4.2 Results **93**

4.2.1 Does Rse1p directly interact with subunits of CPF? 93

4.2.2 The *rse1-1* strain is sensitive to cordecypin 94

4.2.3 The *rse1-1* strain is not deficient in 3' end cleavage *in vitro* 95

4.2.4 The mRNA levels of intron containing genes are lowered in the *rse1-1* mutant strain at restrictive temperature but not the levels of pre-mRNAs without introns 97

4.2.5 Poly(A) tails shorten in the *rse1-1* strain after six hours at 37°C 99

4.2.6 *RSE1* genetically interacts with subunits of the 3' end processing machinery 101

4.2.7 Northern analysis of the single and double mutant strains 104

4.2.8 Poly(A) site selection assay 110

4.3 Discussion **111**

4.4 Experimental procedures **114**

4.4.1 Droplet test 114

4.4.2 Growth curve 114

4.4.3 RNA analyses 114

4.4.4 Protein-protein interactions 114

4.4.5 Extract preparation and *in vitro* cleavage and polyadenylation assays 115

4.4.6 3' end labeling of total RNA 115

4.4.7 Plasmids and primers 115

4.4.8 Yeast strains 116

4.1 Introduction

In yeast only very few genes contain introns and most of the times they are located close to the 5' end of the pre-mRNA. Therefore, there has been almost no research done on the coupling between splicing and 3' end processing in yeast. However, most proteins of the 3' end processing machinery and the U2 snRNP are conserved from mammalians to yeast and we wanted to investigate whether the coupling mechanism mediated by CPSF and the U2 snRNP we observed in humans takes place in *S. cerevisiae* also.

Below is a table of the mammalin U2 snRNP subunits and their homologues in yeast.

Mammalian protein	Yeast homologue
SF3b155	Hsh155
SF3b145	Cus1p
SF3b130	Rse1p
SF3b49	Hsh49p
SF3b14a	Snu17p
SF3b14b	Rds3p
SF3b10	Rcp10
SF3a120	Prp21p
SF3a66	Prp11p
SF3a60	Prp9p
U2-A'	Lea1p
U2-B''	Msl1p
SmB/B'	SmB1
SmD3	SmD3
SmD2	SmD2
SmD1	SmD1
SmE	SmE
SmF	SmF
SmG	SmG

Table 4.1 Mammalin subunits of the U2 snRNP and their yeast homologues

To address the question if the U2 snRNP and CPF couple 3' end formation and splicing, we characterized the temperature sensitive *rse1-1* yeast strain. Rse1p (named for RNA splicing and ER-toGolgi transport) was originally identified in a screen for mutants deficient in ER-to-Golgi transport (Chen et al., 1998). The secretion defect of the *rse1-1* mutant strain could be suppressed by either increasing the amount of *SRA1* or deleting the intron from the chromosomal *SAR1* locus. Sar1p is a small GTP-binding protein required for transport from the endoplasmic reticulum to the Golgi apparatus (Nakano and Muramatsu, 1989). This suggested that the secretion deficiency of the Rse1p mutant strain was an indirect effect. In agreement with this, Rse1p was shown to be required for splicing and for spliceosome assembly and to be part of the U2 snRNP (Chen et al., 1998; Caspary et al., 1999). It interacts only transiently or weakly with the U2 snRNA. Like its mammalian homologue SF3b130, its sequence is similar to CPSF160 and XPE proteins (Caspary et al., 1999). Rse1p interacted in a two-hybrid screen with Prp9 (Fromont-Racine et al., 1997).

We found that the *rse1-1* mutant strain was sensitive to cordycepin. In addition *rse1-1* genetically interacted with different subunits of CPF and Northern blot experiments indicated that some double mutants had increased levels of un-spliced pre-mRNA at restrictive temperature compared to the single mutants. This suggested that the coupling between splicing and 3' end processing takes place in yeast also and that this coupling might be mediated by the U2 snRNP and CPF.

4.2 Results

4.2.1 Does Rse1p directly interact with subunits of CPF?

Given that many proteins involved in 3' end processing and splicing are conserved in mammals and yeast we decided to test whether the coupling mechanism of the two processes mediated by CPSF and the U2 snRNP takes place in yeast as well.

Rse1p is the yeast homologue of SF3b130 which we found to interact with CPSF160, CPSF100 and CPSF30. To test whether Rse1p interacts with subunits of the yeast 3' end processing machinery, we expressed the C- and N-terminal half of the protein in *E. coli* (it was not possible to obtain full length recombinant protein). GST-pull down experiments were performed with numerous subunits of the 3' end

processing machinery; however, both recombinant fragments of Rse1p bound to all tested proteins (results not shown). Given that it is highly unlikely that Rse1p binds to such a large number of polypeptides, we believe that both recombinant proteins are sticky and bind to any partner protein in GST-pull down experiments.

4.2.2 The *rse1-1* strain is sensitive to cordecypin

The *rse1-1* strain is temperature sensitive (Chen et al., 1998) and grows slowly at 30°C. The mutation is lethal at 33°C (Figure 4.1).

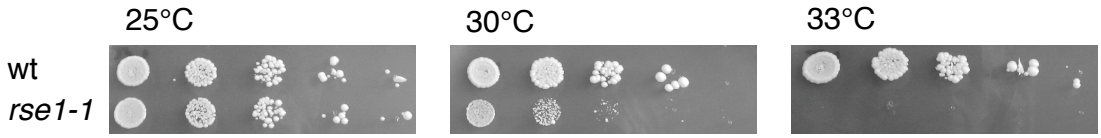


Figure 4.1 The *rse1-1* strain is temperature sensitive

Ten-fold serial dilutions of the yeast strains indicated at the left were put on SD-complete medium plates and incubated at the temperatures indicated at the top.

We tested the *rse1-1* temperature sensitive mutant strain for its sensitivity to cordycepin. Figure 4.2 shows that the *rse1-1* strain grew slower on cordycepin-containing medium than the wild-type strain at 30°C. Cordycepin is a drug that prevents the addition of poly(A) tails (Zeevi et al., 1982). Sensitivity to cordycepin is therefore an indication that the protein could be involved in 3' end processing.

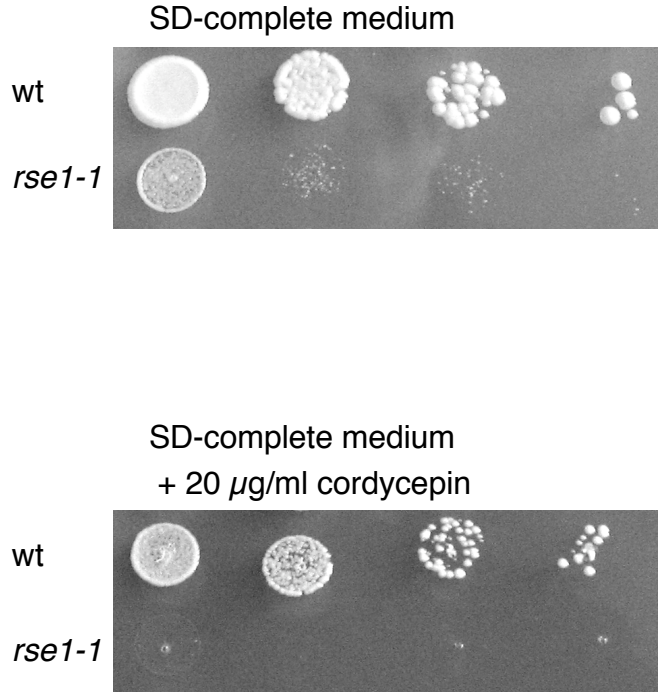


Figure 4.2 The *rse1-1* mutant is sensitive to cordycepin

Ten-fold serial dilutions of the yeast strains indicated at the left were put on SD-complete medium with or with out cordycepin as indicated at the top and incubated at 30°C.

4.2.3 The *rse1-1* strain is not deficient in 3' end cleavage *in vitro*

We analyzed whether the mutation in Rse1p affected cleavage and polyadenylation *in vitro* and tested extracts from wild-type and *rse1-1* cells for their ability to cleave a synthetic *CYC1* pre-mRNA (Figure 4.3). Extracts from isogenic wild-type cells (lanes 2 and 4) and *rse1-1* cells (lanes 3 and 5) efficiently cleaved the RNA precursor at 30°C and 34°C. This indicated that the *rse1-1* strain is not deficient in 3' end cleavage.

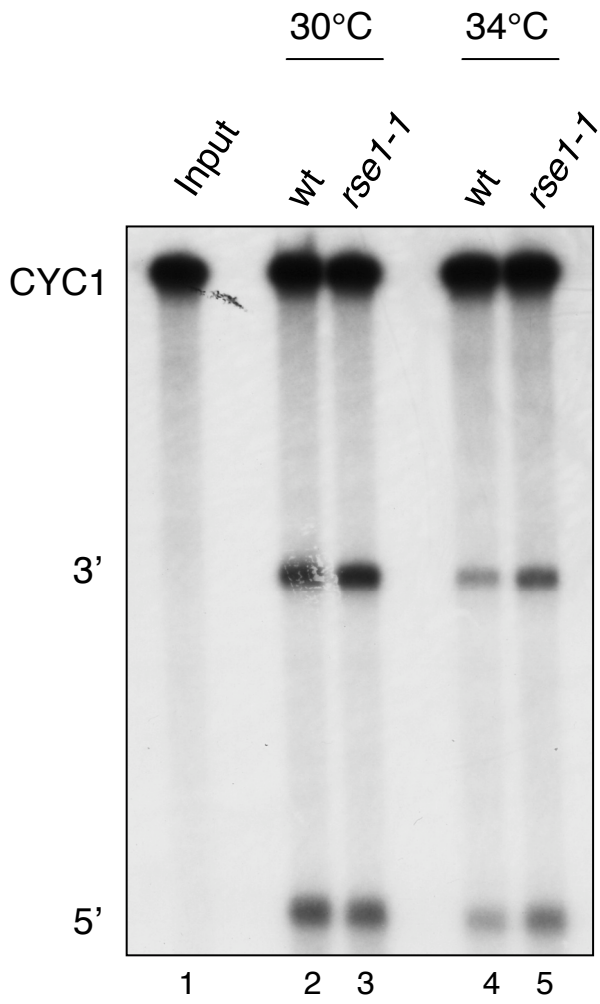


Figure 4.3 The *rse1-1* strain is not deficient in 3' end cleavage *in vitro*

3' end cleavage assay with extract obtained from the *rse1-1* and its isogenig wild-type strain. Input represent mock treated reactions. Equal amounts of total proteins were used in each lane. As substrate served the *CYC1* pre-mRNA. The migration position of the full length RNA and its 3' end cleavage (3') and 5' end cleavage product (5') are indicated at the left. The assay was performed at 30°C and at 34°C as indicated at the top.

We also tested the extracts for their polyadenylation activity with pre-cleaved *CYC1* RNA precursor, but the extracts from wild-type and *rse1-1* cells were both inactive in polyadenylation (results not shown). We generated numerous extracts but were in no case successful to obtain extract that was active in polyadenylation with the above mentioned strains. Possibly the procedure to generate extracts with strains carrying

this specific genetic background requires as different protocol than the one generally used.

4.2.4 The mRNA levels of intron-containing genes are lowered in the *rse1-1* mutant strain at restrictive temperature but not the levels of pre-mRNAs without introns

Figure 4.4 shows a growth curve of the *rse1-1* and its isogenic wild-type strain following shift from 25°C to 37°C. The *rse1-1* cells started to grow significantly more slowly than the wild-type cells after two to three hours incubation at 37°C.

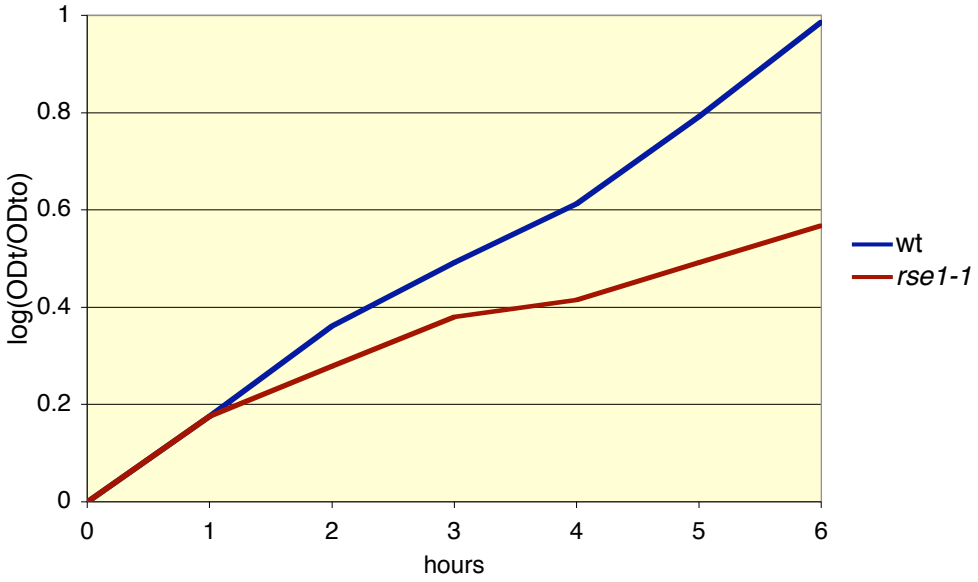


Figure 4.4 Growth analysis of the *rse1-1* mutant strain at 37°C

The graph represents the growth behavior of the *rse1-1* and its isogenic wild-type strain after shift to 37°C. The strains were grown to exponential growth at 25°C and then shifted to a 37°C water bath. The OD was measured after every hour and the growth determined as the log(ODt/ODt0).

Low levels of mRNAs can serve as an indication that a mutant strain might be deficient in pre-mRNA splicing or in 3' end formation. We carried out Nothern blot analysis of the *rse1-1* and isogenic wild-type strains. We found that the *rse1-1* cells contained lower levels of the *CYH2* and *ASC1* mRNAs after three hours shift to 37°C (Figure 4.5), in accordance with the time point after which the cells ceased to grow at

restrictive temperature. Unspliced *CYH2* pre-mRNA is present in the wild-type and the *rse1-1* cells; interestingly, the levels of unspliced *CYH2* pre-mRNA were not elevated in the *rse1-1* cells compared to the wild-type cells. In contrast a faint band of probably unspliced *ASC1* pre-mRNA is visible only in the *rse1-1* but not in the wild-type strain after three hours incubation at restrictive temperature, indicating that Rse1p is required for efficient splicing of this substrate. The levels of the *ACT1* mRNA were marginally lower compared to wild-type cells after six hours shift to restrictive temperature. The levels of the *ADH1* and *NRD1* pre-mRNA were not significantly affected by the *rse1-1* mutation; *18S* rRNA served as loading control. *CYH2*, *ASC1* and *ACT1* pre-mRNAs contain introns, whereas the other pre-mRNAs tested do not. These results indicate that Rse1p is required for the processing of intron-containing pre-mRNAs but not for the processing of intronless pre-mRNAs. At present we do not know whether the low levels of mRNAs from intron-containing genes were due solely to a splicing deficiency of the *rse1-1* strain or if these pre-mRNAs were also 3' end processed with lower effciency.

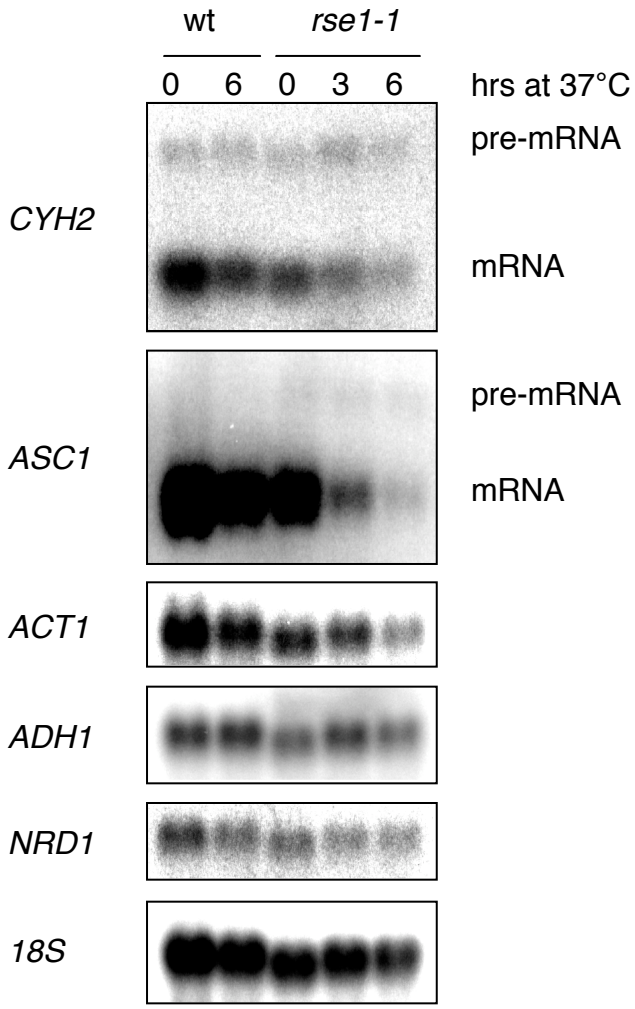


Figure 4.5 The mRNA levels of intron-containing genes are reduced in the *rse1-1* strain upon shift to restrictive temperature

Northern analysis of total RNA extracted form wild-type and mutant *rse1-1* cells grown at 25°C, or after shift to 37°C for 3 and 6 hours as indicated on the top of each lane. The filters were developed with random-primed labeled probes directed against the RNA species indicated at the left of each panel.

4.2.5 Poly(A) tails shorten in the *rse1-1* strain after six hours at 37°C

Next, we analyzed the poly(A) tail length distribution of the *rse1-1* mutant strain before and after shift to restrictive temperature. Up to three hours after shift to 37°C the poly(A) tail length distribution of the *rse1-1* cells were comparable to wild-type cells; however, after six hours at 37°C the tails were considerably shorter (Figure 4.6).

Given that the *rse1-1* strain ceased growth significantly after three hours at 37°C but showed a normal distribution of poly(A) tail length at this time it is likely that the effect we observed after six hours shift could be due to secondary effects caused by the *rse1-1* mutant protein and not due to a direct effect of the mutant protein.

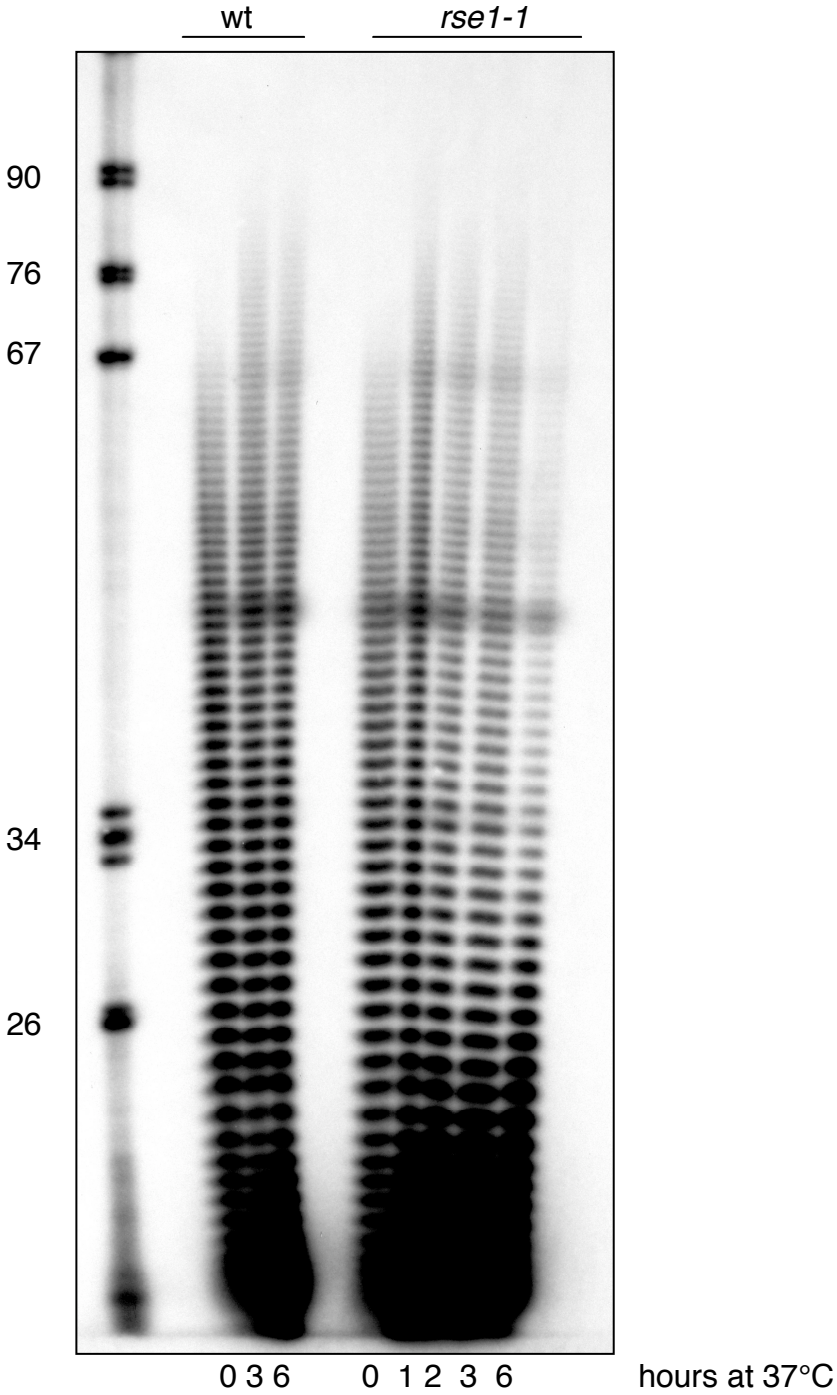
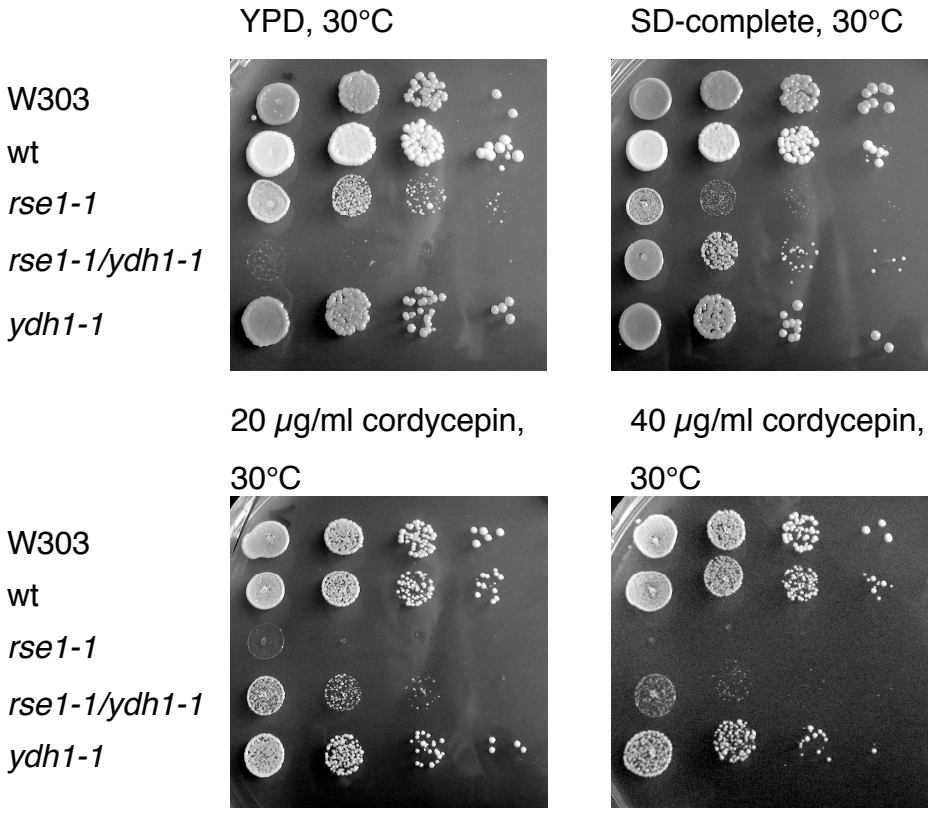


Figure 4.6 Poly(A) tails shorten in the rse1-1 mutant strain six hours after shift to restrictive temperature

Poly(A) tail length distribution assay of *rse1-1* and its isogenic wild-type strain after growth at 25°C, or following shift to 37°C for 1, 2, 3 and 6 hours. The position and size (in number of nucleotides) of the marker bands are indicated on the left.

4.2.6 *RSE1* genetically interacts with subunits of the 3' end processing machinery

We showed here that CPSF100 interacts with SF3b130. Therefore, we tested if Rse1p interacts genetically with Ydh1p, the yeast homologue of CPSF100. The mutant strains *ydh1-1* and *ydh1-3* but not *ydh1-2* were deficient in 3' end cleavage. *ydh1-1* was also shown to be deficient in the polyadenylation step of the 3' end processing reaction. All three mutant strains did not recognize the *ACT1* poly(A) site efficiently (Chapter 6; Kyburz et al., 2003). Figure 4.7 shows that the *ydh1-1* mutation was synergistically lethal in combination with the *rse1-1* mutation at 30°C on YPD but not on SD-complete medium. Furthermore, the cells carrying both mutations were less sensitive to corycepin than the *rse1-1* single mutant cells (Figure 4.7). *rse1-1/ydh1-2* and *rse1-1/ydh1-3* double mutants both grew better on SD-complete medium with or



without cordycepin than the *rse1-1* strain (results not shown).

Figure 4.7 *rse1-1* and *ydh1-1* are synergistically lethal at 30°C on YPD-medium

Ten-fold serial dilutions of the yeast strains indicated at the left were put on YPD and SD-complete plates or SD-complete plates containing 20 μg/ml or 40 μg/ml cordycepin as indicated at the top. W303 is the isogenic wild-type strain of *ydh1-1* and wt stands for the isogenic wild-type strain of *rse1–1*.

SF3b130 interacted directly in GST-pull downs with CPSF160 (the homologue of Yhh1). Furthermore, SF3b130, CPSF160 and Yhh1p share homology (Dichtl et al., 2002b). Therefore, we generated *rse1-1/yhh1-3* and *rse1-1/yhh1-6* double mutants. The mutant strains *yhh1-3* and *yhh1-6* are both deficient in the poly(A) addition step of the 3' end processing reaction. Both strains were also shown to negatively affect poly(A) site selection and transcription termination. Y*hh1-3* but not *yhh1-6* was shown to be deficient in 3' end cleavage *in vitro* (Dichtl et al., 2002b). The *rse1-1/yhh1-6* strain grew as slowly on YPD at 30°C as the *rse1-1* single mutant but the double mutant grew better than the single mutant on SD-complete medium with or without cordycepin (Figure 4.8). The *rse1-1/yhh1-3* did also grow better on SD-complete medium with or without cordycepin than the *rse1-1* single mutant (results not shown). We do not know why the double mutants of *rse1-1/ydh1-1* and *rse1-1/yhh1-6* grow better on SD-complete than on YPD medium.

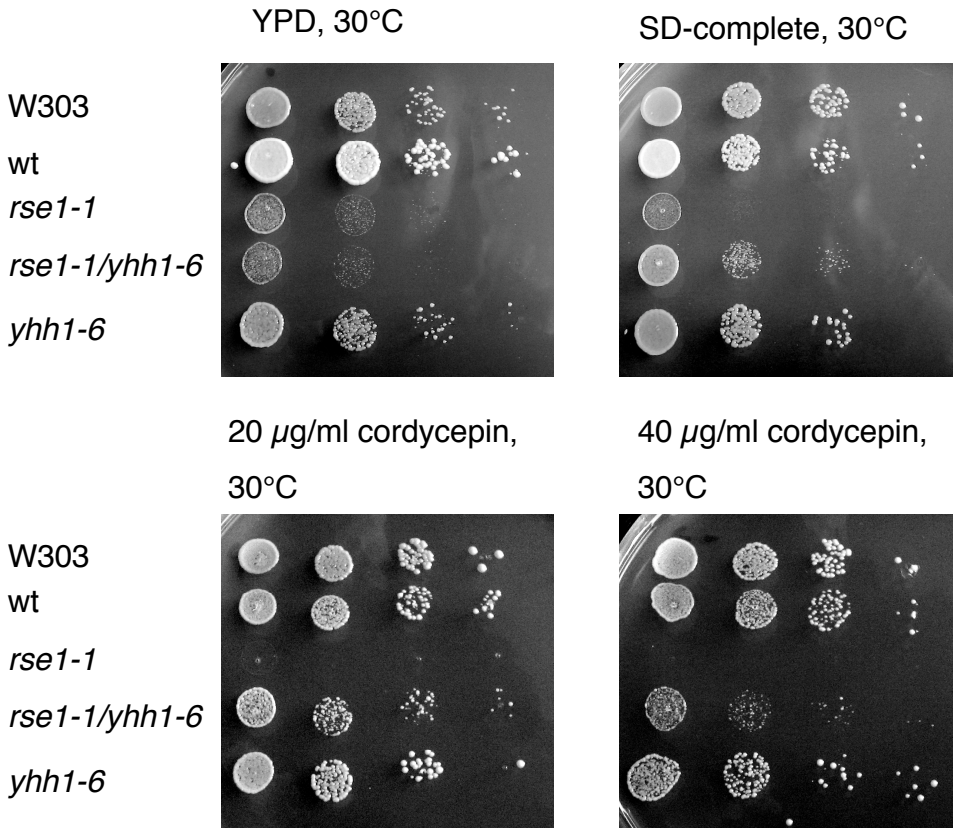


Figure 4.8 *rse1-1* and *yhh1-6* genetically interact

Ten-fold serial dilutions of the yeast strains indicated at the left were put on YPD and SD-complete plates or SD-complete plates containing 20 µg/ml or 40 µg/ml cordycepin as indicated at the top. W303 is the isogenic wild-type strain of *yhh1-6* and wt stands for the isogenic wild-type strain of *rse1–1*.

The cordycein sensitivity of *rse1-1* indicated that this strain might be deficient in polyadenylation. Therefore, we generated double mutants of *rse1-1* and mutants in PAP. *pap1-3* was shown to be deficient in polyadenylation but not in 3' end cleavage (Minvielle-Sebastia et al., 1994; P. Preker and W. Keller unpublished results). Indeed the double mutant with *rse1-1* showed synergistic lethatlity at 30°C on YPD and SD-complete medium. The double mutant was as sensitive to cordycepin as the *rse1-1* mutant strain (Figure 4.9).

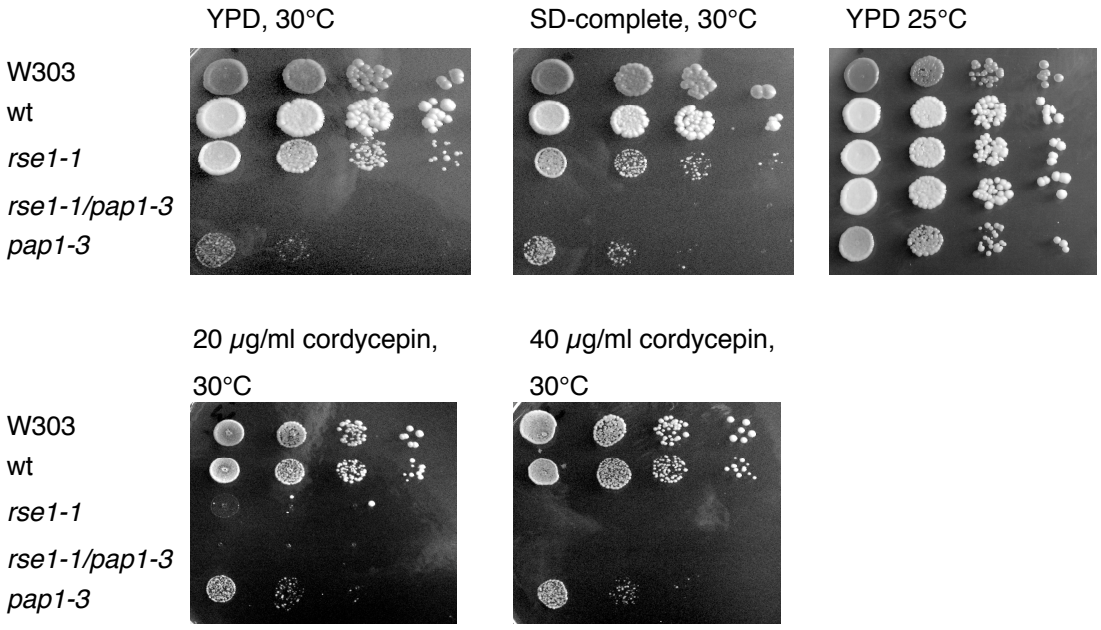


Figure 4.9.*rse1-1* and *pap1-3* are synergistically lethal at 30°C

Ten-fold serial dilutions of the yeast strains indicated at the left were put on YPD and SD-complete plates or SD-complete plates containing 20 µg/ml or 40 µg/ml cordycepin as indicated at the top. W303 is the isogenic wild-type strain of *pap1-3* and wt stands for the isogenic wild-type strain of *rse1–1*.

Furthermore we generated *rse1-1* / *pap1-9* double mutants. *Pap1-9* is deficient in polyadenylation (P. Preker and W. Keller unpublished data). This double mutant showed the same growth on YPD, SD-complete or cordycepin containing medium as the *rse1-1* single mutant. Taken together these results show that Rse1p genetically interacts with Ydh1p, Yhh1p and PAP.

4.2.7 Northern analysis of the single and double mutant strains

We measured the growth curves of the single and double mutants following shift from 25°C to 30°C or 33°C in YPD medium. At 30°C the *rse1-1* and *ydh1-1* single mutant strains grew as well as the wild-type strains. The *rse1-1*/*ydh1-1* double mutant grew slower than the single mutant strains but did not cease growth. At 33°C the *rse1-1* and *ydh1-1* single mutants started to grow slower than the wild-type cells after four respectively three hours. The *rse1-1*/*ydh1-1* double mutant cells ceased growth after approximately one hour shift to 33°C (Figure 4.10).

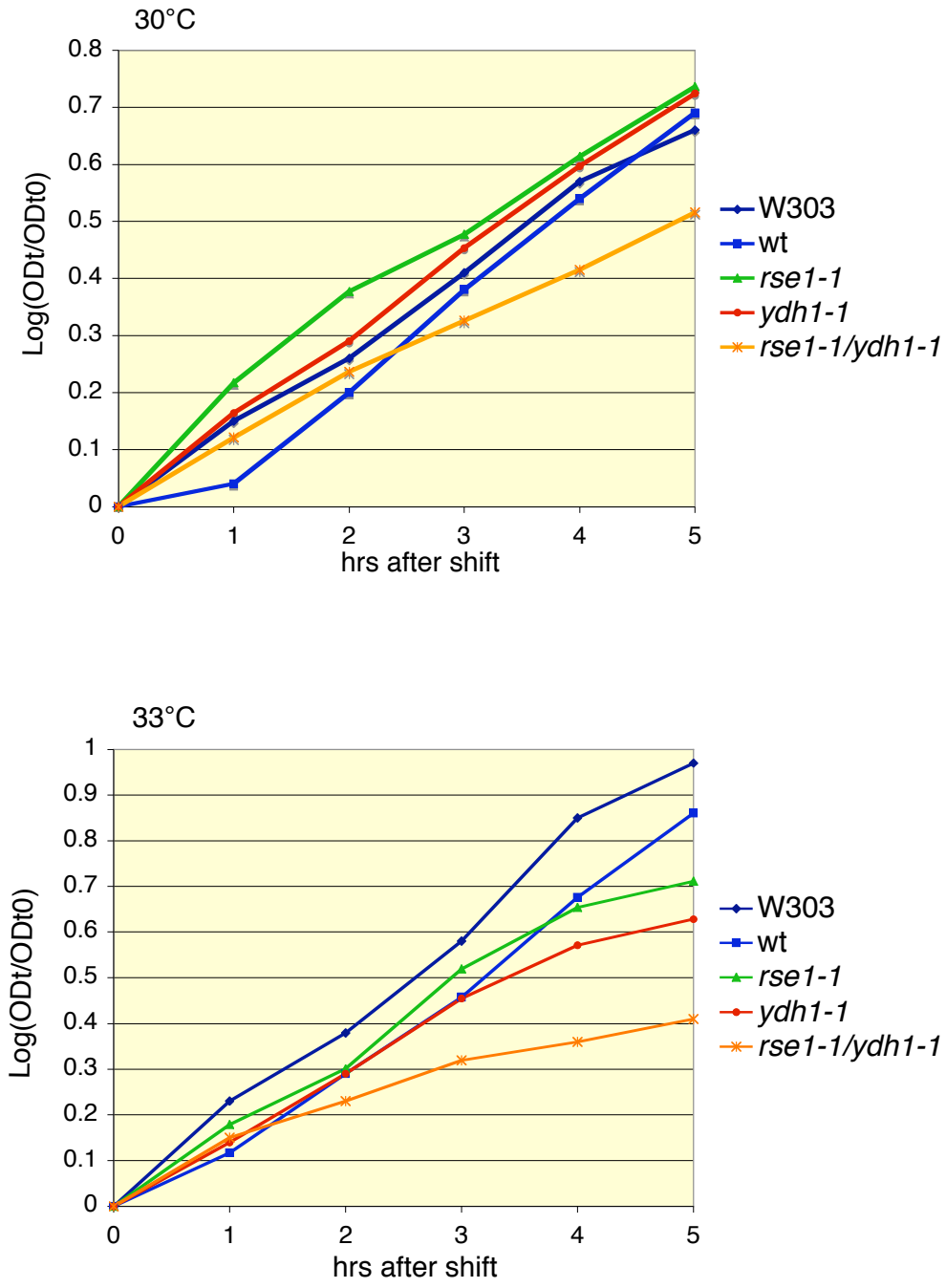


Figure 4.10 Growth analysis of the single and double mutant strains after shift to 30°C and 33°C

The graph represents the growth behavior of the single and double mutant strains and their isogenic wild-type strains after shift to 30°C or 33°C as indicated at the top. W303 is the isogenic wild-type strain of *ydh1-1* and wt is the isogenic wild-type strain of *rse1-1*. The strains were grown to exponential growth at 25°C and then shifted to a water bath with the higher temperature. The OD was measured after every hour and the growth determined as the log(ODt/ODt0).

The *yhh1-6* single mutant grew as well as the wild-type cells at 30°C and 33°C and the *rse1-1/yhh1-6* double mutant strain grew marginally less well than the respective single mutants at both temperatures (Figure 4.11).

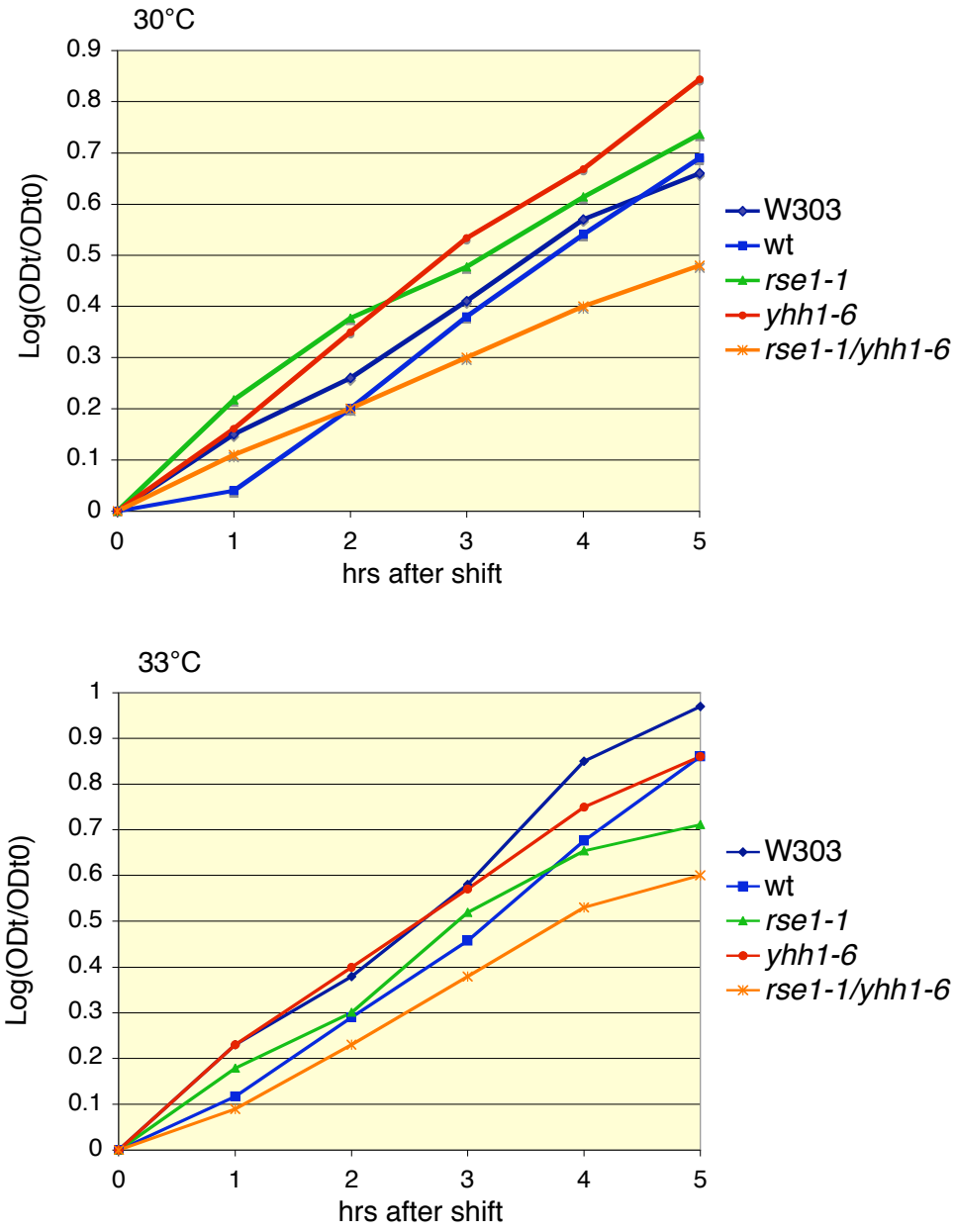


Figure 4.11 Growth analysis of the single and double mutant strains after shift to 30°C and 33°C

The graph represents the growth behavior of the single and double mutant strains and their isogenic wild-type strains after shift to 30°C or 33°C as indicated at the top. W303 is the isogenic wild-type strain of *yhh1-6* and wt is the isogenic wild-type strain of *rse1-1*. The strains were grown to

exponential growth at 25°C and then shifted to a water bath with the higher temperature. The OD was measured after every hour and the growth determined as the log(ODt/ODt0).

The *pap1-3* single mutant strain grew more slowly than the wild-type strain after one hour shift to 30°C and the *rse1-1/pap1-3* double mutant cells ceased growth after approximately one hour at 30°C. The *pap1-3* and *rse1-1/pap1-3* strains both ceased to grow one hour after shift to 33°C (Figure 4.12).

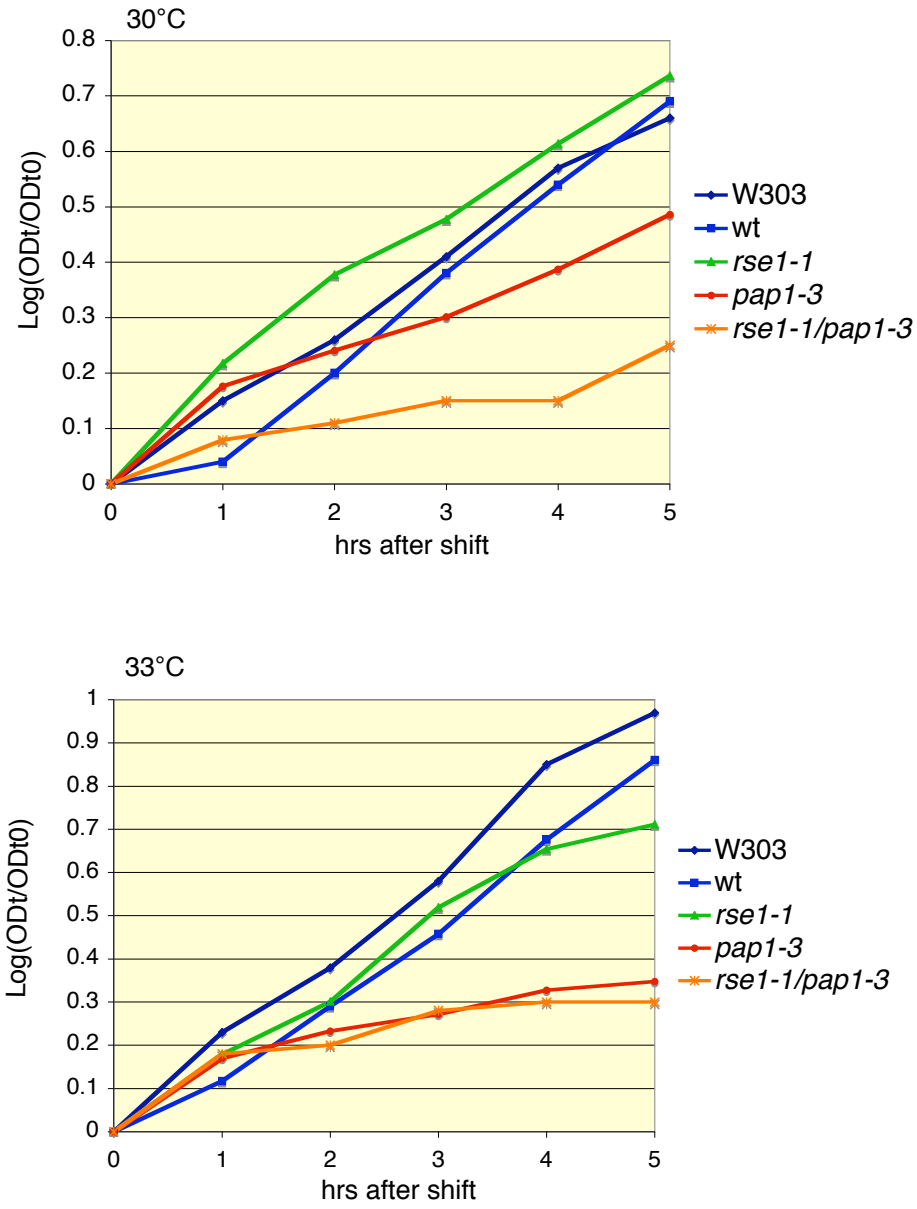


Figure 4.12 Growth analysis of the single and double mutant strains after shift to 30°C and 33°C

The graph represents the growth behavior of the single and double mutant strains and their isogenic wild-type strains after shift to 30°C or 33°C as indicated at the top. W303 is the isogenic wild-type strain of *pap1-3* and wt is the isogenic wild-type strain of *rse1-1*. The strains were grown to exponential growth at 25°C and then shifted to a water bath with the higher temperature. The OD was measured after every hour and the growth determined as the log(ODt/ODt0).

We did Northern analyzes of the single and double mutants after shift to 30°C and 33°C (Figure 4.13.). The levels of the *CYH2* and *ASC1* mRNAs were significantly reduced compared to the wild-type cells after 5 hours shift to 30°C and 33°C in the *rse1-1* cells. The unspliced *CYH2* pre-mRNA was visible in the Northern blot with the same intensity in the wild-type as in the *rse1-1* strain. The unspliced form of the *ASC1* pre-mRNA was visible after 5 hours shift to 30°C and 33°C in the *rse1-1* cells but not in the wild-type cells. The mRNA levels of *ACT1*, *ADH1*, *NRD1* and *18S* were not affected by the *rse1-1* mutation. The *ydh1-1* mutant cells showed reduced levels of the *CYH2*, *ASC1*, *ACT1* and *ADH1* mRNA after 5 hours growth at 30°C and 33°C, the *NRD1* mRNA levels did not change significantly upon shift to the higher temperatures. The same low levels of mRNA and elevated levels of pre-mRNA of the single mutants were found in the double mutant strain *rse1-1*/*ydh1-1*.

The levels of the *CYH2*, *ASC1* and *ADH1* mRNAs were reduced in the *yhh1-6* cells only after 5 hours shift to 33°C. *ACT1* mRNA levels did not seam to be effected and the *NRD1* mRNA levels were increased five hours after shift to 30°C and 33°C. The mRNA levels of the *rse1-1*/*yhh1-6* double mutant cells were comparable to the *yhh1-6* single mutant with the exception that the *NRD1* mRNA levels were not increased. Interestingly this double mutant had higher levels of unspliced *CYH2* pre-mRNA compared to the single mutants. The unspliced *ASC1* pre-mRNA was also found in higher levels compared to the *rse1-1* or *yhh1-6* single mutants even before shift to restrictive temperature. In the *rse1-1* mutant this pre-mRNA is only visible after five hours shift to the restrictive temperature and not at all in the *yhh1-6* mutant cells. However, in the double mutant cells unspliced *ASC1* pre-mRNA is present even before shift to restrictive temperature.

The *pap1-3* cells contained lower levels of the *CYH2*, *ASC1* and *ACT1* mRNAs after two and five hours shift to 30°C and 33°C. The *ADH1* mRNA levels diminished only at 33°C after two and five hours shift. The *NRD1* mRNA levels were increased five hours after shift to 33°C. Interestingly, cells carrying the *pap1-3* mutation contained unspliced *ASC1* pre-mRNA after five hours shift to 30°C and 33°C. The mRNA levels of the *rse1-1*/*pap1-3* double mutant strain were comparable to the *pap1-3* single mutant cells. The amounts of unspliced *ASC1* pre-mRNA were however significantly higher in the double mutant compared to the single mutants.

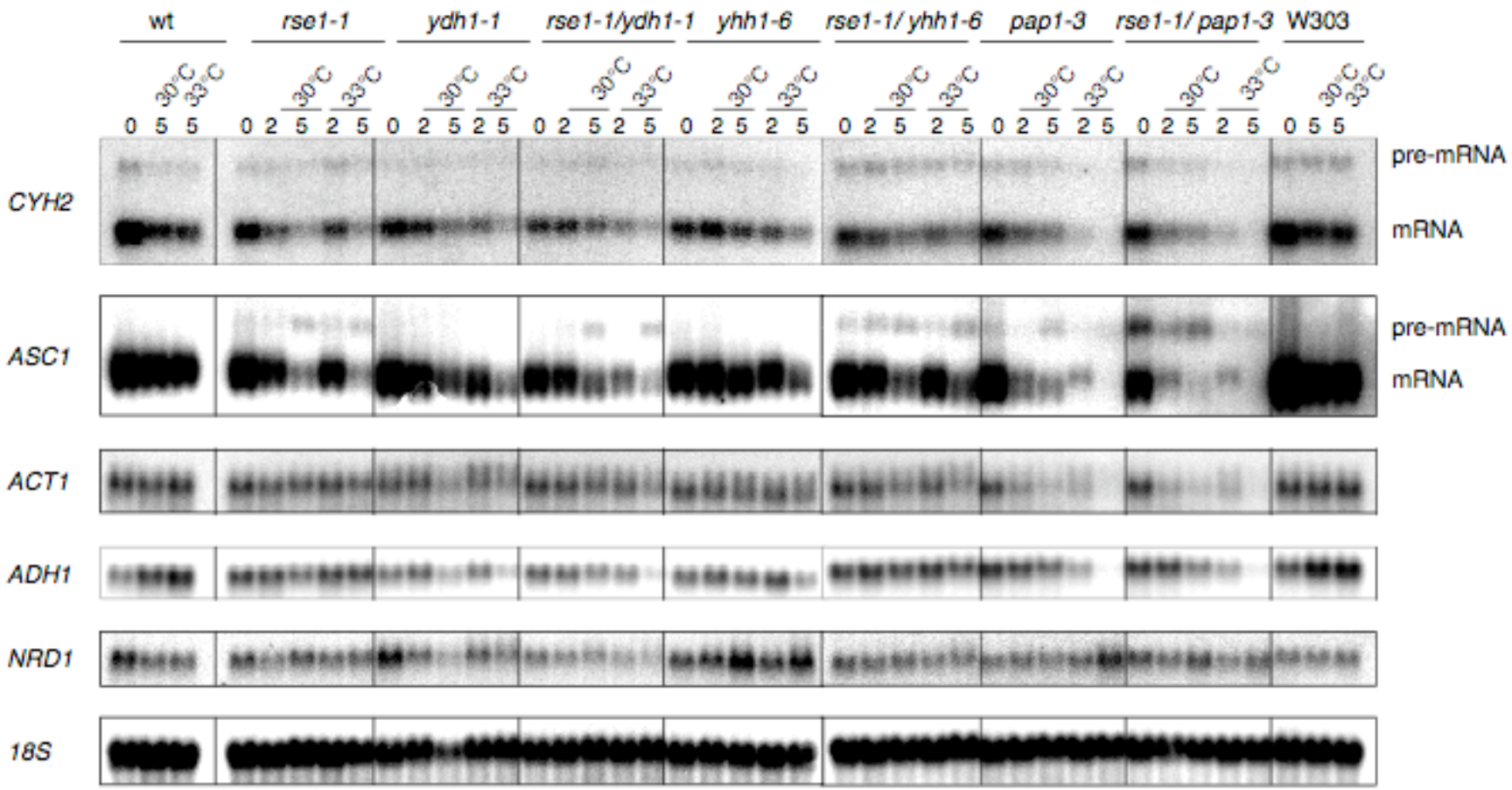


Figure 4.13 Northern analysis of the single and double mutant strains.

Northern analysis of total RNA extracted form wild-type (wt / W303) and single and double mutant strains (as indicated on top) grown at 25°C, or after shift to 30°C or 33°C for 2 and 5 hours as indicated on the top of each lane. The filters were developed with random-primed labeled probes directed against the RNA species indicated at the left of each panel.

4.2.8 Poly(A) site selection assay

Given that the U2 snRNP plays a vital role in splice site selection, we decided to test if Rse1p influences poly(A) site selection. We analyzed poly(A) site usage of the *ACT1* pre-mRNA in wild-type and mutant cells. The *ACT1* 3' untranslated region (UTR) contains at least four polyadenylation sites, of which the most proximal one (site I) is used with highest frequency in wild-type cells. We found that the *rse1-1* cells cleaved site I over site IV with the same ration as the wild-type cells even after 5 hours shift to restrictive temperature (Figure 4.14). As has been shown before for *ydh1-1* (Kyburz et al., 2003) and *yhh1-6* (Dichtl et al., 2002b) the ratio of site I over site IV was significantly lower in the *ydh1-1*and *yhh1-6* strains before and after five hours shift to restrictive temperature the same was the case for the respective double mutants. Also *pap1-3* single mutant and the *rse1-1/pap1-3* double mutant strains were deficient in *ACT1* poly(A) site selection after shift to 33°C.

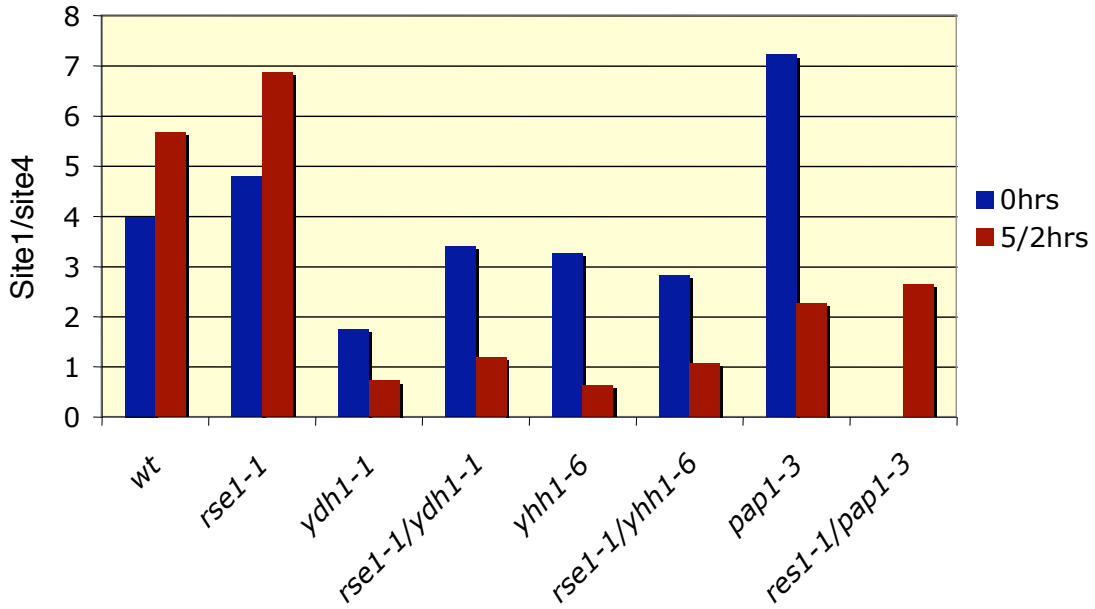


Figure 4.14 *ACT1* poly(A) site selection assay

The graph represent the evaluation of three independent experiments of *ACT1* poly(A) site selection of single and double mutant strains before and after 5 respectively 2 hours shift to 33°C. All strains were analyzed 5 hours after shift to 37°C except *pap1-3* and *rse1-1/pap1-3* which were analyzed after 2 hours shift to 37°C. The y-axis represents the ration of site1/site4 usage.

4.3 Discussion

The coupling between splicing and 3' end processing in mammalians is quite well established. However, there is very little information concerning the question whether this coupling is found in yeast as well. We showed here that the U2 snRNP and CPSF are involved in coupling the two processes in mammalian cells and wanted to find out whether this mechanism is conserved in yeast.

If the yeast U2 snRNP is indeed involved in coupling splicing and 3' end processing one would expect that mutation in its subunits would not only affect splicing but 3' end processing as well. Therefore, we investigated if Rse1p, a subunit of the U2 snRNP, is involved in 3' end formation. We found that the *rse1-1* strain is not deficient in 3' end cleavage *in vitro* or in poly(A) site selection *in vivo*. The length of the poly(A) tails were reduced in this strain after six hours shift to restrictive temperature, which could be the result of a direct involvement of Rse1p in poly(A) tail synthesis. However, this phenotype only appeared after six hours shift to restrictive temperature, whereas the cells ceased to grow already after three hours. Therefore, it is more likely that this was not a direct but rather an indirect effect of the *rse1-1* mutation. Furthermore, Northern analysis showed that mRNA levels of genes that do not contain an intron were not reduced in the *rse1-1* strain upon shift to restrictive temperature. Given that un-polyadenylated mRNAs are prone to rapid degradation, this indicated that the pre-mRNAs of intron-less genes were efficiently 3' end processed. The combined results suggest that Rse1p is not involved in the 3' end processing reaction of intron-less pre-mRNAs. Given that the coupling of splicing and 3' end processing can only take place on intron-containing mRNA this finding is not surprising. Interestingly, Northern analysis showed that the levels of intron-containing pre-mRNAs are reduced in this strain upon shift to restrictive temperature. We do not know at present whether this effect is due solely to the splicing deficiency of the *rse1-1* mutant or to an additional 3' end processing deficiency.

However, we showed that *rse1-1* is sensitive to cordycepin, a drug which prevents the addition of poly(A) tails. In addition, the *rse1-1* mutation genetically interacted with mutations in CPF subunits. Northern Blot analysis of the different single and double mutants indicated that the *rse1-1/yhh1-6* and *rse1-1/pap1-3* double

mutants had higher levels of unspliced pre-mRNAs than the single mutants. Given the genetic interaction and the observation that double mutants in Rse1p and CPF subunits showed an increased splicing deficiency indicated that splicing and 3' end processing are coupled in yeast and that the coupling mechanism mediated by CPSF and the U2 snRNP we observed in mammals could take place also in yeast cells.

Furthermore, the single mutant *pap1-3* had elevated levels of unspliced *ASC1* pre-mRNA after five hours shift to 30°C and 33°C. This indicated that this mutation might disturb the coupling of splicing and 3' end processing. On the other hand, Pap could be even more directly involved in the splicing mechanism than only through the coupling mechanism. However, given that we observe this effect only after five hours shift to the non-permissive temperature one would have to look at earlier time points, in order to be able to rule out that this was merely a secondary effect. Interestingly, Ysh1p the putative endonuclease was originally identified as a splicing factor (Noble and Guthrie, 1996). In this case, it is not clear yet either if the protein is directly involved in the splicing mechanism or enables the coupling between 3' end processing and splicing. It will be very interesting in the future to further investigate the role that the 3' end processing factors play in splicing.

The levels of the Nrd1 transcript were higher in the *yhh1-6* but not in the *rse1-1/yhh1-6* double mutant cells compared to the wild-type strains. Nrd1 was shown to be involved in the transcription termination of snoRNAs. Elevated levels of this transcript could be an indication for a deficiency in this pathway (Steinmetz et al., 2001). To our knowledge, Yhh1p has so far not been implied to play a role in snoRNA termination and it would be interesting to investigate this possibility in the future. Furthermore, it is intriguing that the double mutant does not show higher levels of the *NRD1* message. This suggests that the *rse1-1* mutation could potentially counteract the effect of *yhh1-6* on snoRNA termination or *NRD1* transcription regulation.

The *rse1-1/pap1-3* double mutant caused synergistic lethality on YPD and on SD-complete medium. Interestingly, we found that the *rse1-1/ydh1-1* and *rse1-1/yhh1-6* double mutants grew better on SD-complete than on YPD containing medium. In both cases the double mutant cells grew better than the *rse1-1* single mutant strain on the SD-complete medium. At present we do not know what caused this peculiar growing behavior. One possible explanation could be the following:

Cells generally grow slower on SD-complete compared to YPD medium, given that the YPD is a richer medium than the SD-complete medium. Assuming that the double mutants are deficient in pre-mRNA processing, this would lead to unprocessed and maybe incorrectly processed mRNAs. When the cells are growing quickly, a large pool of deficient mRNAs would be generated in the cells. However, when the cells grow slower, it is possible that they have enough time to repair the deficient mRNAs and therefore be able to survive.

We found that the *rse1-1*/*ydh1-1* and *rse1-1*/*yhh1-6* double mutants grew better on cordycepin-containing medium than the *rse1-1* single mutant. This could indicate that the effect of mutations in Ydh1p and Yhh1p counteract the effect of the *rse1-1* mutation, which causes sensitivity to cordycepin. However, the double mutants generally grew better than the *rse1-1* strain on SD-complete medium plates also – therefore, the observation that the *rse1-1*/*ydh1-1* and *rse1-1*/*yhh1-6* double mutants grew better on cordycepin-containing medium than the *rse1-1* single mutant does not necessarily need to be an indication of opposite effects of the single mutants on the cordycepin sensitivitiy but could also be explained by different growth behavior of the strains on the SD-medium.

In summary the combined results suggest that the coupling between splicing and 3' end formation is indeed taking place in *S. cerevisiae* as well. The genetic interactions indicate that this coupling could be mediated by subunits of CPF and the U2 snRNP. However, proof that the two complexes indeed physically interact is still missing. Future work should investigate the interaction network of the 3' end processing machinery and proteins of the spliceosome. Furthermore, it would be highly interesting to find out how the 3' end processing proteins act on the splicing mechanism.

4.4 Experimental procedures

4.4.1 Droplet test

The strains were grown over night and diluted to an OD_{600} of 0.1, 0.01, 0.001, 0.0001, 0.00001. 5 μl of each dilution were spotted on three different YPD plates and incubated at 15, 23 and 37˚C.

4.4.2 Growth curve

The yeast strains were incubated at 25˚C over night diluted in the morning to an OD_{600} of 0.1 (100 ml cultures) and incubated again at 25˚C until they reached an OD_{600} of 0.2 - 0.4. The strains were transferred to a water bath at 37˚C (t0); the OD_{600} was measured every hour and the cells were diluted with pre-warmed medium when they reached an OD_{600} higher than 0.8. After 0, 2, 4, 6 and 8 hours 40 ml of the cultures were harvested. Approximately 10 ml ice were added and centrifuged at 4˚C for 5 min at 5 k. The cells were washed with cold water and centrifuged again. The pellets were frozen in liquid Nitrogen and stored at –70˚C.

4.4.3 RNA analyses

Northern analysis and Rnase H experiments were carried out as described in (Dichtl et al., 2002b).

4.4.4 Protein-protein interactions

GST-pull down experiments were performed as described in chapter 2.5.5. We increased the salt concentration of the binding and the washing buffers in order to get rid of the unspecific binding but without success.

4.4.5 Extract preparation and *in vitro* cleavage and polyadenylation assays

The experiments were performed as described in 6.5.5.

4.4.6 3'-end labeling of total RNA

Procedure: To 2 μg total RNA, was dissolved in a total volume of 162 µl 5 mM Tris HCl, pH 7.0, 12.5 mM KCl, 175 µM MnCl2, 2.5% glycerol, 50 mg BSA, 1 μl poly(A) polymerase (yeast) and 1 μl cordycepintriphosphate was added and the mixture incubated at 37˚C for 30 min. The poly(A) polymerase was deactivated at 95˚C for 3 min and 188 μl H_2O added to the solution. The radioactivity of two times 10 μl of each sample was measured in the scintillation counter.

The amount of solution equivalent to 1200000 cpm was digested by addition of 80 μl the digestion-mix (10 mM Tris HCl, pH 8.0, 300 mM NaCl, 1 µg RNase A, 25 U RNase T1). The reaction was carried out at 37˚C for two hours, stopped by addition of 20 μl stop-mix (50 mM EDTA, 5% SDS, 20 µg Glykogen, 1 µg Proteinase K) and incubation for 30 min at 42˚C. 2.5 M NaOAc (f.c. 0.08 M) and 300 μl ethanol were added and the RNA precipitated over night at –20˚C. The mixture was centrifuged at 14 k, 4˚C for 15 min and the pellet washed with 70% ethanol, dried in the speedvac for 5 min at low temperature and resuspended in 9 μl FA (95% formamide, 20 mM EDTA, 0.05% bromphenolblue, 0.05% xylenecyanol).

The poly(A) tails were separated on a 10% denaturing sequencing polyacrylamide gel. 1x TBE served as running buffer and the gel was run at 1200 V until the bromphenolblue reached the bottom of the gel.

4.4.7 Plasmids and primers

The sequence of Rse1 was PCR amplified from genomic DNA with the primers: 3'-Rse1H6-EagI (CCATCATCGGCCGCTAGTGATGGTGATGGTGATGCATGTAATTTGTTCTGACTTCGT) and 5'-Rse1-XmaI: (ATGGTTAGCCCCGGGGATGTGGGGAGGTGGCAAAATGGC-3'). The PCR

product was ligated into the GDV vector. The N-terminal part of the protein was PCR amplified with the primers Rse1-Not1-short: (ATCAAGCTGCGGCGCTTAGTGATGGTGATGGTAGCGAGGACGTGGCACA TACAATACGAATTC) and 5'-Rse1-XmaI: (ATGGTTAGCCCCGGGGATGTGGGGAGGTGGCAAAATGGC-3'). The C-terminal part of the protein was PCR amplified with the primer pair Rse1-Xma1-short (ATGGTTAGCCCCGGGGCTGACTTGGGTGCCACCAGCGGGAATTC) and Rse1H6-Not1
(CAACGAAGTCAGAACAAATTACATGCATCACCATCACCATCACTAAGCG GCCGCAGCTTGAT). The N- and C-terminal fragments were both cloned into the GDV vector.

4.4.8 Yeast strains

Genotypes of yeast strains used in this study were: CYK568: MATa; *rse1-1; ura3-52; leu2-3,112;* CKY294: MATa; *ura3-52; leu2-3,112;* YPP106: MATα; *ura3-1; ade2-1; leu2-3,112; his3-11,15;trp1Δ; ydh1::TRP1*[pIA115; *CEN4-URA3-YDH1*]; YBD40: MATα; urs3-1; ade2-1; leu2-3,112; his3-11.15; *;trp1Δ; yhh::TRP1* [pBD57; *CEN4-URA3-YHH*]. PJP14: MATα; *ura3-1; trp1-1; ade2-1; leu2-3,112; his3-11.15; pap::LEU2* [pHCp50; *CEN4-URA3-PAP*];

To generated the double mutants the CYK568 strain was crossed with the YPP106, PJP14 or the YPD40 strain. The spores were dissected and the resulting strains tested for temperature sensitivity at 37°C. The sensitive strains were transformed with plamids carrying the mutant 3' end processing subunits. pAK21 (*CEN-LEU2-ydh1-1*), pAK22 (*CEN-LEU2-ydh1-2*)and pAK22 (*CEN-LEU2-ydh1-3)* were transformed into the double mutants generated with the YPP106 strain. The double mutants produced with the strain YBD40 were transformed with the plasmide: pBD243 (*CEN-LEU2-yhh1-3*) or pBD244 (*CEN-LEU2-yhh1-6*). The double mutants generated with the strains PJP14 were transformed with pBD13 (*CEN-LEU2-pap1-3*) or pBD14 (*CEN-LEU2-pap1-9*). The yeast strains were plated onto 5'FOA plates in order to shuffle out the wild-type plasmid. The genotypes of the double mutants were verified by PCR and sequencing.

Chapter 5: Characterization of the tRNA endonuclease complex and its involvement in 3' end processing

5.1 Introduction **118**

5.2 Results **119**

5.2.1 CFII purification 119

5.2.2 tRNA splicing activity co-elutes with CF II_m. 121

5.2.3 Purified CF II_m is active in tRNA splicing 123

5.2.4 His-Flag-hClp1 purified complexes are active in pre-mRNA 3' end cleavage but not His-Flag-hSen purified complexes 126

5.3 Discussion **129**

5.4 Experimental procedures **132**

5.4.1 Purification of CF II_m 132

5.4.2 3' end cleavage reaction 132

5.4.3 Purification of tRNA endonuclease complexes and tRNA endonuclease assay 133

5.4.4 Western blot analysis 133

5.1 Introduction

Only recently, the homologues of the yeast tRNA splicing endonuclease in humans have been identified by searching of the data bases for the mammalian homologues and isolation of the cDNA of the putative h*SEN2* and h*SEN34* genes (Paushkin et al., 2004). Two different splice variants of the hSEN2 gene were isolated, one of them lacking the entire exon 8 (hSen2ΔEx8). The authors generated HEK293 cell lines stably expressing His-Flag-tagged hSen2, hSen34 as well as hSen2ΔEx8. Purification of the complexes revealed that hSen2 and hSen34 were active in tRNA splicing. However, complexes purified with His-Flag-tagged hSen2ΔEx8 retained the ability to cleave precursor tRNA but the fidelity and accuracy of cleavage was severely reduced. Mass spectrometry analysis of the different complexes revealed that the homologue of the yeast Sen54p protein was present in all three complexes, but the homologue of the yeast Sen15p was not found in the complex isolated with His-Flag-hSenΔEx8.

Interestingly, mass spectrometry analysis of a band running at approximately 45 kDa identified hClp1, a subunit of Cleavage Factor II_m, in all the different complexes. This finding was very surprising given that no link between tRNA splicing and 3' end formation had previously been observed. Furthermore, the presence of the 3' end processing subunits symplekin and CstF64 were probed and they were seen in all the differently purified complexes. Therefore stable cell lines expressing His-Flag-hClp1 were generated and the complex was isolated (Paushkin et al., 2004). The purified complex displayed a protein pattern that was almost identical to that of complexes purified by the tagged versions of hSen2, hSen34 and hSen15. Furthermore, it posessed tRNA endonuclease activity. These results suggested that the tRNA endonuclease complex is tightly associated with 3' end processing factors.

Paushkin and coauthors wanted to test the hypothesis that the tRNA endonuclease subunits hSen2 is involved in 3' end processing. Therefore the levels of the hSen2 protein were depleted by RNAi. As expected this depletion resulted in increased levels or pre-tRNA consistent with a role of the protein in tRNA splicing. Furthermore, it was shown that depletion of the hSen2 protein also resulted in increased levels of pre-mRNA, which had not been cleaved at the 3' end. This

suggested that the tRNA splicing endonulease complex could be involved in 3' end formation.

We wanted to further clarify the role of the tRNA endonuclease subunits in 3' end processing. We therefore performed 3' end cleavage experiments with nuclear extract depleted of hPcf11 or hClp1 and tried to reconstitute the activity with the different complexes that had been purified by Paushkin and collaborators (2004). However, only addition of the complex purified with the His-Flag-hClp1 protein was able to reconstitute cleavage activity. Furthermore Western Blot analysis showed that hPcf11 was present only in the His-Flag-hClp purified complex but not in the complexes purified with any of the hSen proteins.

S. Paushkin and C. Trotta assayed different fractions of purified CPSF, CstF, CF I_m and CF II_m for tRNA endonuclease activity. They found that all CF II_m fractions were active.

Taken together these results suggest that hClp1 is a subunit of the tRNA endonuclease complex as well as a subunit of CF II_m. However, we do not believe that the tRNA endonuclease complex is indeed involved in 3' end processing for reasons described in the discussion.

5.2. Results

5.2.1 CFII purification

Given that CF II_m has never been purified to homogenity, we wanted to try to obtain pure factor. We therefore looked in our CPSF purification on which column CPSF and CF II_m subunits separated from each other. Western blot experiments revealed that subunits of CF II_m separated form CPSF subunits on the Mono Q column (results not shown). We pooled the fractions of the MQ column that contained CF II_m and further purified the factor over Mono S followed by Resource Q and a Superose S200 gel filtration column (Figure 5.1).

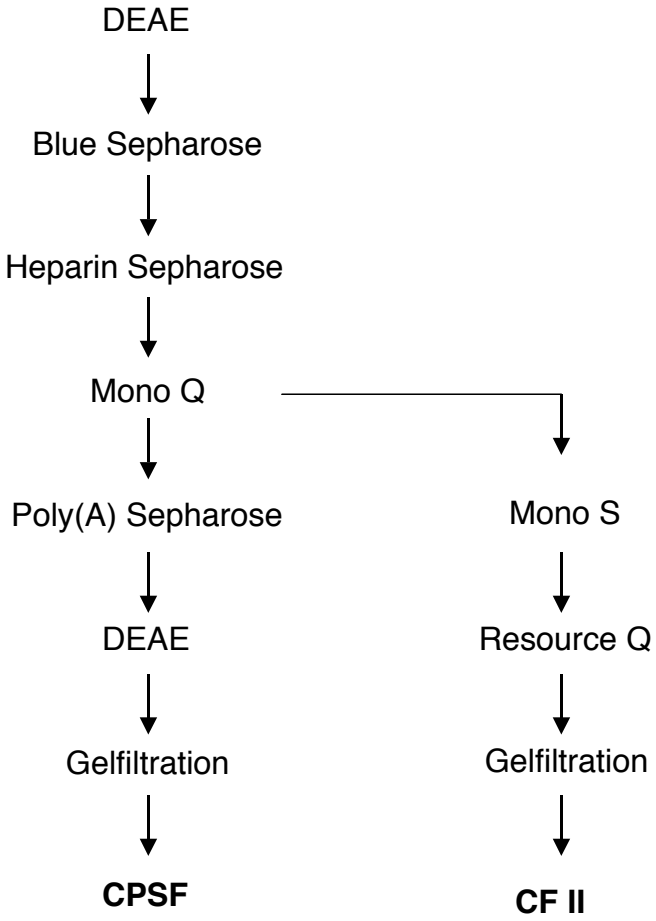


Figure 5.1 Purification scheme of CPSF and CF II_m

hClp1 and hPcf11 eluted on the gel filtration column mainly between the fractions 55 to 58 as determined by Western blotting (Figure 5.2 left panel). These fractions also contained the 59, 68 and 72 kDa subunits of CF I_m. The polypeptide composition of the gel filtration column fractions is shown in Figure 5.2, right panel. Unfortunately, this purification procedure did not result in very pure CF II_m, given that many proteins co-purified with CF II_m.

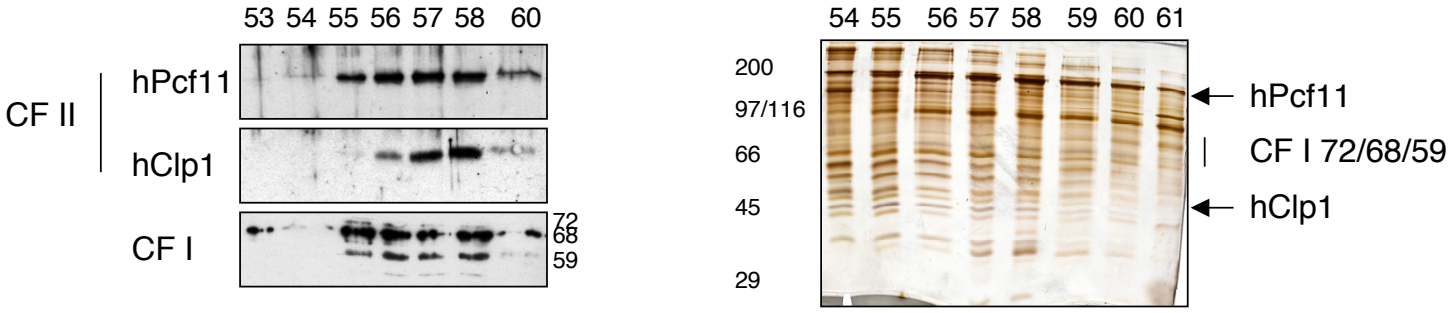


Figure 5.2 Protein composition of CF II_m after purification on a gel-filtration column

The left panel shows a Western blot of the gel-filtration fractions (indicated at the top) with antibodies directed against the polypeptides indicated at the left. The right panel shows a silver stained gel of the gel-filtration fractions 54-61. The migration position of the molecular weight marker is indicated at the left. The approximate migration position of hPcf11, CF I_m 72 kDa, CF I_m 68 kDa, CF I_m 59 kDa and hClp1 is indicated at the right.

5.2.2 tRNA splicing activity co-elutes with CF II_m

In order to test if the pre-tRNA endonuclease activity is associated with the different 3' end processing factors, S. Paushkin and C. Trotta assayed purified CPSF, CstF, CF I_m and CF II_m in tRNA splicing assays. As can be seen in Figure 5.3 neither pure CstF (lane 7), different purification grades of CPSF (lanes 10 and 12) nor CF I_m (lane 11) were active in tRNA splicing. In contrast, all the protein fractions that contained CF II_m were active in tRNA splicing (lanes 3, 6, 8, 9, 13, 14, 15).

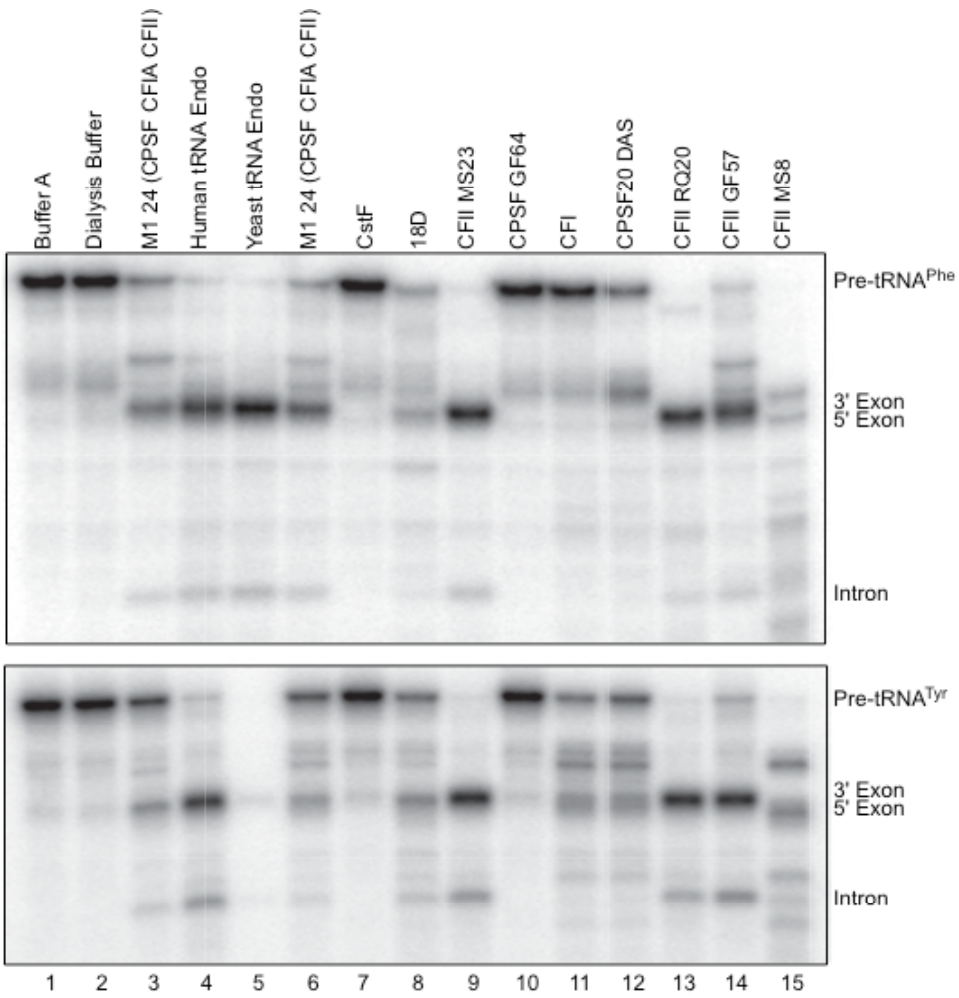


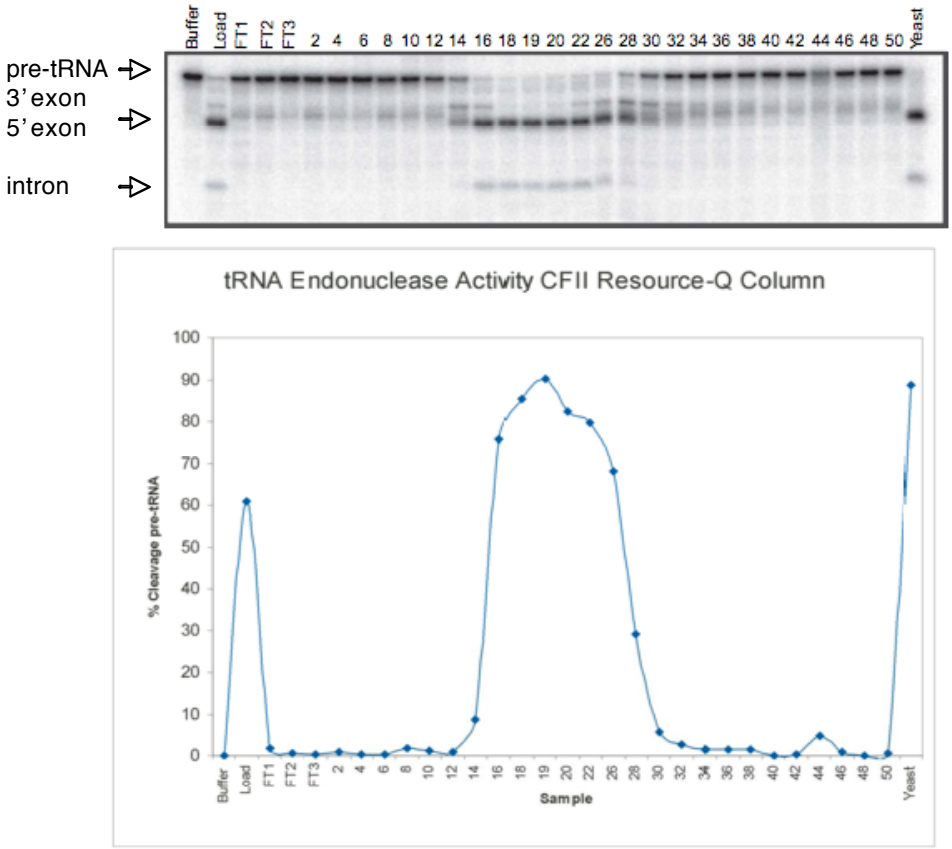
Figure 5.3 pre-tRNA splicing assay (performed by S. Paushkin and C. Trotta)

The purified fractions indicated at the top were assayed for their ability to cleave the pre-tRNA Phenylalaline (upper panel) or the pre-tRNA Tyrosin (lower panel). The migration position of the free 3' and 5' exons and the intron are indicated at the right. Buffer A was the buffer used to purify the human and yeast tRNA endonuclease complexes (Human / Yeast tRNA Endo). Dialysis Buffer indicates reactions that were only incubated with buffer that was used for the dialysis of the different fractions used in the lanes 6 – 15. The lane labeled with M1 24 contained partially purified CPSF that contained CF I_m and CF II_m. Lane 7 contained highly purified CstF. The reaction in lane 8 was performed with partially purified CF II_m after purification on a Resource Q (H. deVries). The lane labeled with CF II_m MS23 contained proteins obtained after purification on the MonoS column fraction 23. CPSF GF64 was highly purified CPSF that was run over a gelfiltration column. Lane 11 contained highly pure CF I_m (U. Rüegsegger) and lane 12 pure CPSF obtained after the poly(A) sepharose

column. CF II_m RQ20 stands for the fraction 20 of the Resource Q column described in this capter. Lane 14 contained CF II_m that was purified over a gelfiltration column. Lane 15 contained CF IIm that was purified by H. deVries over a Mono S column, fraction 8 of this column was used.

5.2.3 Purified CF II_m is active in tRNA splicing

In order to confirm the hypothesis that CF II_m and the tRNA endonuclease proteins form a single complex T. Trotta and S. Paushkin carried out tRNA splicing assays with different fractions of purified CF II_m. The fractions 16 to 26 of the Resource Q column were active in splicing tRNA (Figure 5.4A) in accordance with the elution profile of hClp1 and hPcf11 as determined by Western Blotting (Figure 5.4B).



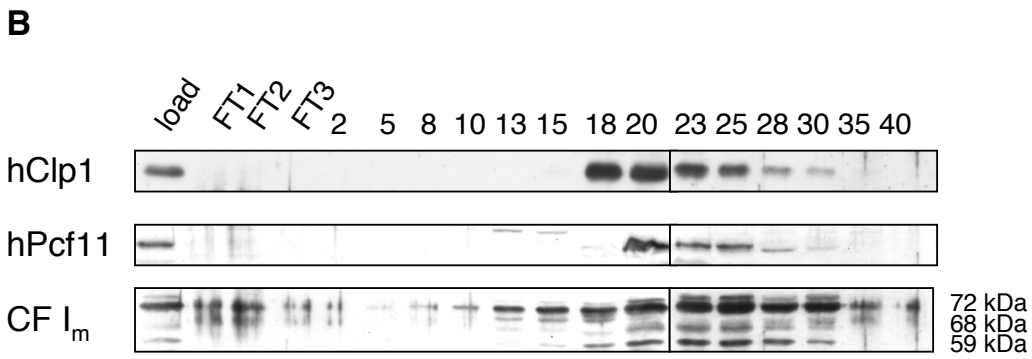


Figure 5.4 tRNA endonuclease activity co-purifies with CF II on the Resource Q column

A) Upper panel shows a tRNA endonuclease assay (performed by S. Paushkin and C. Trotta) with fractions of the Resource Q column indicated at the top. FT1-3 stands for reactions incubated with the flow-through fractions number 1-3 of the RQ column. The migration positions of the free 5' and 3' exons, the intron and pre-tRNA are indicated at the left. The lower panel shows the quantitative evaluation of the cleavage activity of the Resource Q fractions.

B) Western blot analysis of Resource Q column fractions. The numbers of the fractions are indicated at the top. FT1-3 stands for the flow through fractions of the column. The antibodies used are indicted at the left.

The gel filtaration fractions number 54 to 64 were also active in tRNA splicing (Figure 5.5). hClp1 and hPcf11 eluted on this column mainly between the fractions 55 to 58 (Figure 5.2). Therefore, the elution profile of CF II_m subunits and the tRNA endonuclase activity did not fully overlap suggesting that the two complexes might not form a single complex. On the other hand, some tRNA endonuclease complexes might contain CF II_m whereas others might not. This could explain why CF II_m and tRNA endonuclease activity was found in the earlier but not in later fractions.

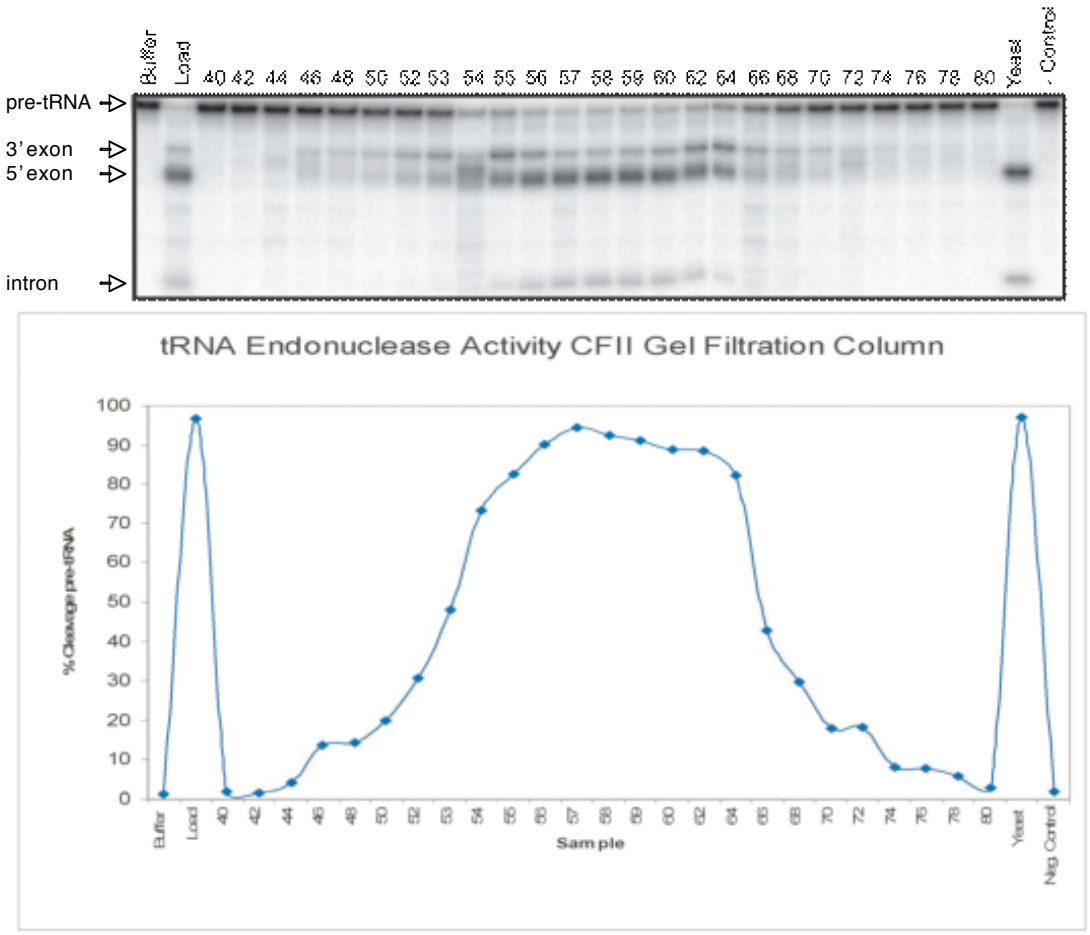


Figure 5.5 tRNA endonuclease activity co-purifies with CF II on the gel filtration column (performed by S. Paushkin and C. Trotta)

The upper panel shows a tRNA endonuclease assay with fractions of the gel filtration column indicated at the top. The free 5' and 3' exons, the intron and pre-tRNA are indicated at the left. The lower panel shows the quantitative evaluation of the cleavage activity of the gel filtration column.

As expected, Western blot assays with antibodies directed against hSen2 confirmed, that this protein was present in the purified CF II_m fractions that were active in tRNA cleavage (Figure5.6)

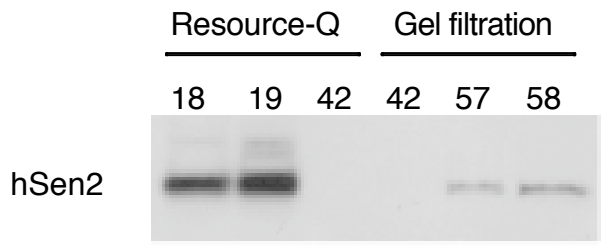


Figure 5.6 hSen2 protein is present in purified CF II (performed by S. Pauskin and C. Trotta)
Western blot of different fractions of purified CF II after a Resource Q or following a gel filtration column as indicated on top. The numbers represent the different fraction of the corresponding column. An antibody directed against hSen2 was used.

5.2.4. His-Flag-hClp1-purified complexes but not His-Flag-hSen purified complexes are active in pre-mRNA 3' end cleavage

If the model that tRNA splicing endonuclease proteins and CF II_m are present in a common complex, then complexes purified with the His-Flag-tagged tRNA endonuclease subunits should be active in 3' end cleavage. To test this hypothesis, we depleted nuclear extract of hPcf11 and tried to reconstitute the cleavage activity with complexes purified with His-Flag-hClp1 or complexes purified with His-Flag-hSen2 that lacked exon 8 (His-Flag-hSen2Δ8), His-Flag-hSen2, or a mutant hSen2 protein that contained a mutation in the active site, His-Flag-hSen34 or a mutant hSen34 protein that contained a mutation in the active site of the protein or His-Flag-hSen54 (Figure 5.7). However, only the addition of the His-Flag-hClp1-purified complex could reconstitute the cleavage activity but not the addition of the other complexes. Mock purifications of HEK293 cells which did not contain any tagged proteins served as negative controls. We performed the reconstitution assay also with hClp1-depleted HeLa nuclear extract and got the same results as with the extract depleted of hPcf11 (results not shown).

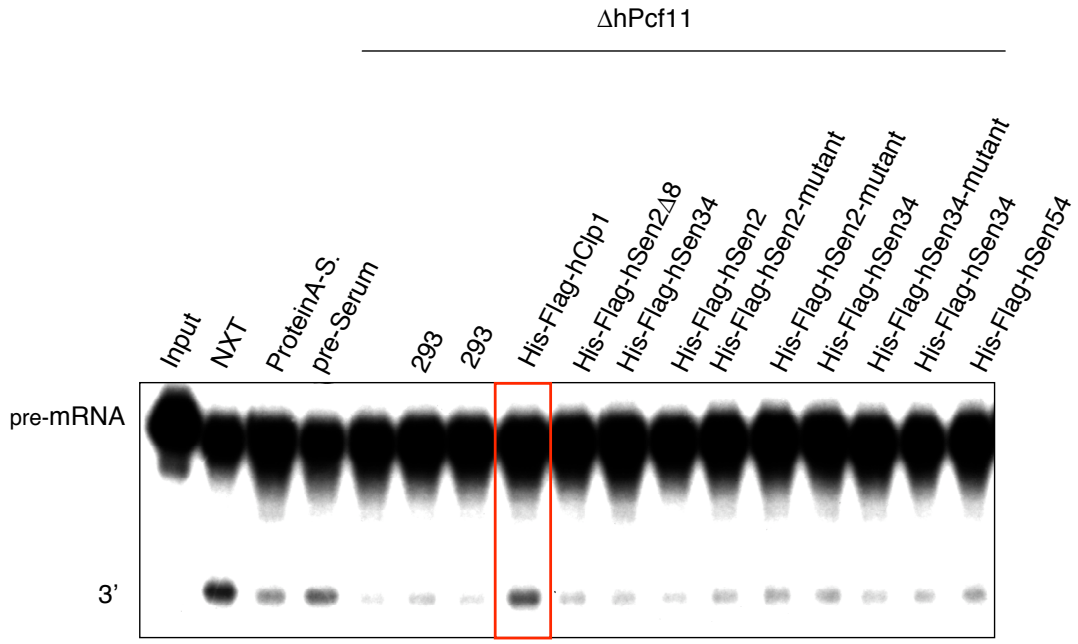


Figure 5.7 Complexes purified with His-Flag-tagged hClp1 but not other purified tRNA endonuclease complexes can reconstitute 3' end cleavage activity.

3' end cleavage assay with hPcf11 depleted HeLa nuclear extract and the differently purified complexes (indicated on top). In the lane labeled with NXT undepleted extract was used in the assay. ProteinA-S stands for nuclear extract that was mock depleted, and pre-Serum for extract that was depleted with pre-immune serum. The migration position of the unprocessed pre-mRNA (pre-mRNA) or the 3' end cleavage product (3') is indicated at the left.

Furthermore, Western blot analysis showed that the complexes purified with His-Flag-hClp1 but not with the His-Flag-hSen proteins contained hPcf11 (Figure 5.8, upper panel) even though all of them contained hClp1 (Figure 5.8, lower panel). We did not detect CF I_m 68 kDa, CF I_m 59kDa, CPSF160, CPSF73 nor CPSF100 in any of the complexes (data not shown).

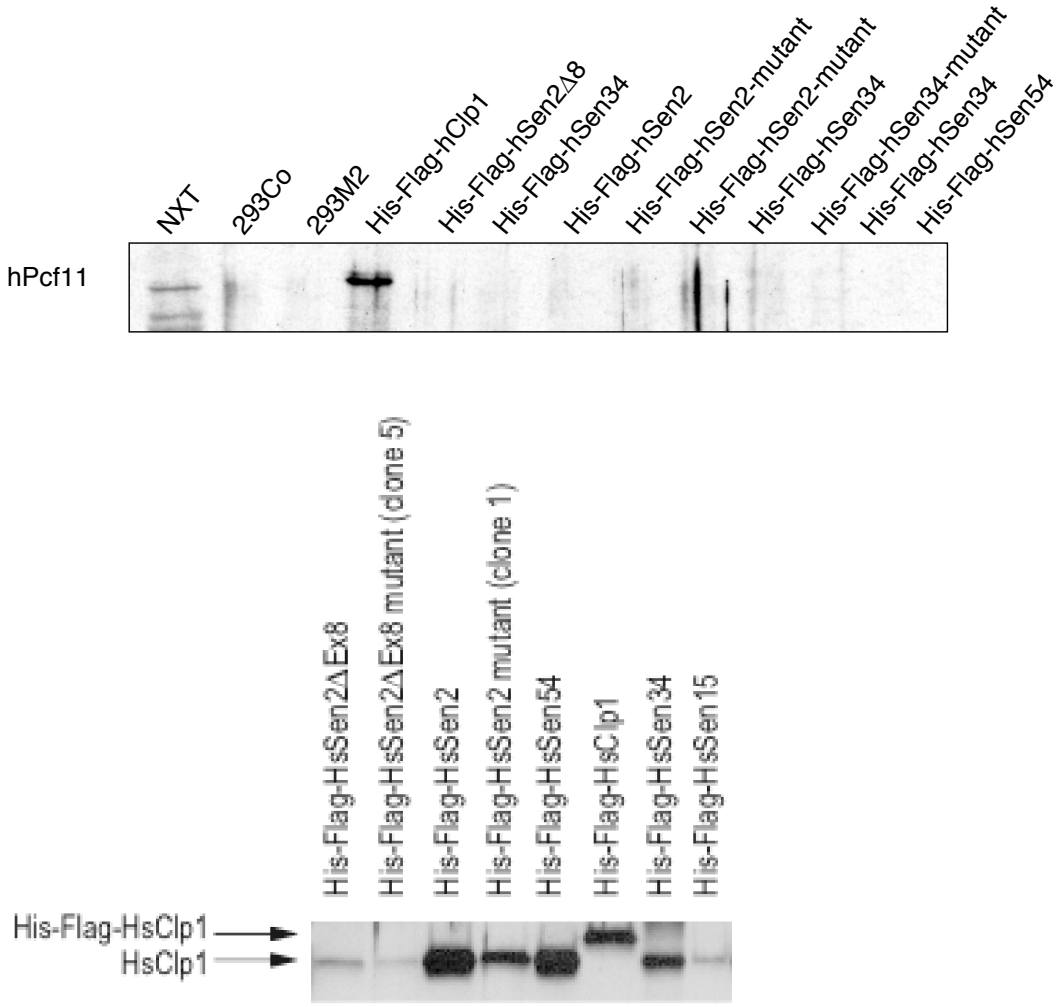


Figure 5.8. His-Flag-hClp-purified complex contains hPcf11

Western Blotting with the differently purified complexes as indicated at the top with antibodies directed against hPcf11 (upper panel) or hClp1 (lower panel; performed by S. Paushkin and C. Trotta). 293Co and 293M2 correspond to purifications performed with HEK cells that did not contain any tagged proteins.

In summary, the results indicate that the tRNA endonuclease complex and CF II_m are not one but rather two distinct complexes. However, it appears that hClp1 is a genuine component of both the tRNA splicing complex and of CF II_m.

5.3 Discussion

Paushkin and coworkers proposed a model, suggesting that tRNA splicing and pre-mRNA 3' end formation are catalyzed by the same components of an endonuclease complex in mammalian cells (Pauskin et al., 2004). These authors hypothesized that this endonuclease complex functions in the formation of mRNA, tRNA and potentially other RNA substrates. They based this model on the finding that hClp1, CstF64 and symplekin co-purified with the tRNA splicing complex. Furthermore, tagged hClp1 pulled down the proteins involved in tRNA splicing and knock down of hSen2, a subunit of the tRNA endonuclease complex resulted in decrease of tRNA splicing as well as 3' end processing activity.

We found that tRNA splicing activity is associated with purified CF II_m but not with CstF, CPSF or PAP. This is in agreement with the finding that tRNA splicing proteins were pulled down with tagged hClp1. Furthermore, the elution profile of hClp1 and hPcf11 after purification of CF II_m on Resource Q overlapped with tRNA endonuclease acitvity. However, the elution profile of the CF II_m subunits on the gel filtration column did not fully overlap with the tRNA endonuclease activity. The proteins started to elute in the same fractions but the endonuclease activity eluted over more fractions than hClp1 or hPcf11. Given that gel filtration columns separate complexes according to their size, this could indicate that only a part of the endonuclease complexes present in a cell are associated with CF II_m. On the other hand this could also mean that CF II_m and the tRNA endonuclease complex are two independent complexes of similar sizes. In agreement with the second possibility, Western blot analysis showed that hPcf11, the second subunit of CF II_m, was only present in complexes purified with His-Flag-tagged hClp1 but not in the complexes purified with tagged hSen proteins.

To test if the tRNA endonuclease complexes are active in 3' end formation we tested differently purified complexes for their ability to reconstitute CF II_m activity. Only the complex that was purified with tagged hClp1 but not the complexes purified with tagged hSen proteins were active in 3' end cleavage. Taken together, these results do not support the model in which tRNA splicing endonuclease subunits and proteins involved in 3' end processing form a single complex capable of different processing reactions. However, the data support the model that hClp1 is part of the

3' end processing complex as well as a subunit of the tRNA endonuclease complex. hClp1 contains the Walker A and B motifs, which have been implicated in ATP/GTP binding (Walker et al., 1982; Saraste et al., 1990). Unpublished UV cross-linking experiments by Henk deVries suggested that hClp1 indeed binds GTP. It is possible that hClp1 plays similar roles in the 3' end formation as well as in the tRNA splicing complex. Given that both complexes are involved in RNA processing, it is likely that they could share a subunit that carries out a similar function in both complexes – potentially its role could be the binding of GTP. This would not be the first example of a subunit of the 3' end processing machinery that is found in a different complex as well. Swd2p was shown to be part of CPF and is also a subunit of the SET-complex (Dichtl et al., RNA 2004).

CstF64 and symplekin were found to be associated with the tRNA enodnuclease complex (Paushkin et al., 2004). We did not detect other CPSF or CF I_m subunits in these complexes. At present it is not known whether CstF64 and symplekin are integral parts of the endonuclease complex or if they are only losely associated. This question could be addressed by His-Flag tagging and purification of the two protein.

It was also shown that knock down of hSen2 caused defects in tRNA splicing and 3' end processing, implicating that this protein is involved in 3' end formation (Paushkin et al., 2004). It is possible that the deficient 3' end processing observed upon hSen2 depletion could in fact be caused by a secondary effect. Depletion of hSen2 did not cause lethality of the cells and therefore stable cell lines in which hSen2 was permanently knocked down were generated. The effect of this knock down was analyzed thirty days after initial silencing of the gene – therefore after a relatively long period of time, during which many secondary effects of the knock down could have accumulated. Furthermore, the 50% knock down of the hSen2 protein was not lethal. Given that 3' end processing in yeast is essential one would expect that knock down of a protein involved in 3' end formation in human cells would also lead to cell death. It is a possibility that 50% of the hSen2 protein level normally found in cells could be sufficient for survival. However, taking all the results into consideration we think that it is unlikely that hSen2 is directly involved in 3' end processing.

We suggest that hClp1 is part of the tRNA splicing endonuclease complex as well as a subunit of CF II_m. We cannot exclude the possibility that the tRNA

endonuclease complex functions in the formation of mRNA but we think the results supporting this hypothesis are rather weak.

5.4 Experimental procedures

5.4.1 Purification of CF II_m

The CF II_m containing fractions of the Mono Q column were pooled and dialyzed against Buffer A containing: 25 mM $(NH_4)_2SO_4$, 20 mM Hepes-KOH pH. 7.9, 0.4 mM EDTA pH. 8.0, 0.02% NP40, 10% Glycerol, 3 mM $MgCl_2$, 0,5 mM DTT, 1 mM PMSF, 2 μg/ml Pepestatin, 2 μg/ml Leupeptin. A 8 ml Mono S HP10/10 column was washed with 2 M KCl, Buffer B (same as Buffer A but containing 380 mM $(NH_4)_2SO_4$) followed by Buffer A. The pooled fractions were loaded, the column washed with 40 ml Buffer A and the proteins eluted by a gradient that started with Buffer A and ended after 160 ml with Buffer B. The fractions containing CF II_m were identified by western blotting, pooled and dialysed against Buffer C containing: 20 mM $(NH_4)_2SO_4$, 50 mM Tris-HCl pH 7.9, 10% Glycerol, 0.02% NP40, 0.5 mM EDTA, 3 mM Mg, 0.5 mM DTT, 1 mM PMSF, 2 μg/ml Pepestatin, 2 μg/ml Leupeptin.

A 6 ml Resource Q column (Pharmacia) was washed with 12 mM Buffer C, followed by 12 ml Buffer D containing the same ingredients as Buffer C but 300 mM $(NH_4)_2SO_4$ instead of 20 mM $(NH_4)_2SO_4$, and finally the column was washed with 30 ml Buffer C again. The pooled and dialysed Mono S fractions were loaded, the column washed with 30 mM Buffer C and the proteins eluted with a gradient of 120 ml ranging from Buffer C to Buffer D.

A Superdex 200 16/60 (Pharmacia) was washed with 2 ml 2 M KCl followed by overnight washing with Buffer A containing 30 mM $(NH_4)_2SO_4$ instead of 25 mM $(NH_4)_2SO_4$. 1 ml of fracition 19 of the Resource Q column was loaded onto the gel filtration column and 50 μl fractions collected.

5.4.2 3' end cleavage reaction

15 fmol RNA substrate was incubated with 3 μl HeLa cell nuclear extract, and in case of the reconstitution reactions with 6 μl of the different complexes, in a final concentration of: 2 mM DTT, 0.2 U RNA guard, 2.6% Polyvinylalcohol, 0.01%

NP40, 0.01 µg/µl Cratin Kinase, 20 mM Creatin Phosphate, 0.5 mM cordycepin, 1 mM $MgCl_2$ and 0.1 µg/µl tRNA in a final volume of 25 µl. The reaction was incubated at 30°C for one hour. 1,5 ml Proteinase K buffer (200 mM Tris HCl pH 7.9, 300 mM NaCl, 25 mM EDTA pH 8.0, 2 % SDS), 60 µl Proteinase K (10 mg/ml), 7.5 µl Glykogen (20 mg/ml; Boehringer) and 682.5 µl H_2O was combined and 75 µl given to the reaction. The mixture was incubated at 42°C in order to digest the proteins. The RNA was precipitated by addition of 250 µl 100% ethanol and zetrifugation at 14 k for 30 min. The RNA was washed with 70% ethanol, dried, resuspended in 6 µl RNA lading dye (FA; 95% formamide, 20 mM EDTA, 0.05% bromphenolblue, 0.05%xylencyanol) and run on a 10% polyacrylamide / 8.3 M urea gel.

5.4.3 Purification of tRNA endonuclease complexes and tRNA endonuclease assay

Purification of the tRNA enodonuclease complexes and the tRNA endonuclease assays were performed as described in (Paushkin et al., 2004).

5.4.4 Western blot analysis

25 µl of the different complexes were run on a 10% SDS gel, the proteins transferred to nitrocellulose and analyzed by the antibodies indicated in each figure.

Chapter 6: The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3' end formation

(This chapter has been published in Nucleic Acids Res. 31, 3936-3945, 2003).

6.1 Abstract **135**

6.2 Introduction **135**

6.3 Results **138**

6.3.1 Ydh1p is required for cleavage and polyadenylation *in vitro* 138

6.3.2 mRNAs are unstable in *ydh1* mutant strains at restrictive temperature 142

6.3.3 Ydh1p is required for poly(A) site selection of *ACT1* pre-mRNA 142

6.3.4 Ydh1p interacts with the C-terminal domain of RNA polymerase II 145

6.3.5 Ydh1p interacts with other subunits of CPF and with Pcf11p, a subunit of CF IA 147

6.4 Discussion **149**

6.5 Experimental procedures **152**

6.5.1 Yeast strains 152

6.5.2 Plasmids and primers 153

6.5.3 Expression of recombinant proteins in *E. coli* 154

6.5.4 Protein-protein interactions 154

6.5.5 Extract preparation and *in vitro* cleavage and polyadenylation assays 155

6.5.6 RNA analyses 155

6.5.7 Protein extraction for Western blotting 155

6.1 Abstract

Cleavage and polyadenylation factor (CPF) is a multi-protein complex that functions in pre-mRNA 3' end formation and in the RNA polymerase II (RNAP II) transcription cycle. Ydh1p/Cft2p is an essential component of CPF but its precise role in 3' end processing remained unclear. We found that mutations in *YDH1* inhibited both the cleavage and the polyadenylation steps of the 3' end formation reaction *in vitro*. Recently, we demonstrated that an important function of CPF lies in the recognition of poly(A) site sequences and RNA binding analyses suggested that Ydh1p/Cft2p interacts with the poly(A) site region. Here we show that mutant *ydh1* strains are deficient in the recognition of the *ACT1* cleavage site *in vivo*. The C-terminal domain (CTD) of RNAP II plays a major role in coupling 3' end processing and transcription. We provide evidence that Ydh1p/Cft2p interacts with the CTD of RNAP II, several other subunits of CPF and with Pcf11p, a component of CF IA. We propose that Ydh1p/Cft2p contributes to the formation of important interaction surfaces that mediate the dynamic association of CPF with RNAP II, the recognition of poly(A) site sequences and the assembly of the polyadenylation machinery on the RNA substrate.

6.2 Introduction

All eukaryotic mRNA precursors (pre-mRNA) are extensively modified before they can serve as templates for protein synthesis. Pre-mRNA 3' end processing is initiated by endonucleolytic cleavage at the poly(A) site. Subsequently, the upstream cleavage product is polyadenylated whereas the downstream fragment is rapidly degraded (for review see (Zhao et al., 1999a). The yeast 3' end processing reaction can be reconstituted *in vitro* with the cleavage and polyadenylation factor IA (CF IA), cleavage and polyadenylation factor IB (CF IB), cleavage and polyadenylation factor (CPF) and the poly(A) binding protein (Pab1p; Chen and Moore, 1992; Ohnacker et al., 2000). Interactions between their subunits allow bridging of the different factors and ensure their coordinated action on the substrate. So far the CPF components

Fip1p and Pfs2p were shown to bridge CPF with CF IA by interacting with Rna14p, a subunit of CF IA (Preker et al., 1995; Ohnacker et al., 2000). A table of the factors involved in 3' end processing and their subunits is provided as supplementary material.

The polyadenylation signals that guide the processing machinery in yeast are redundant and more degenerate compared to the well-conserved sequences in higher eukaryotes. Still, conserved elements can be found, which are the efficiency element (EE), the positioning element (PE), the poly(A) site and U-rich sequences. The EE is located at variable distances upstream of the cleavage site (Graber et al., 1999). The PE is often found approximately 20 nucleotides upstream of the cleavage site and consists of an A-rich sequence (Russo et al., 1991). U-rich regions, located directly upstream and downstream of the cleavage site, and the poly(A) site itself act in concert to produce multiple recognition sites (Guo and Sherman, 1996; Graber et al., 1999; van Helden et al., 2000; Dichtl and Keller, 2001).

RNase H protection mapping experiments with a CYC1 pre-mRNA suggested that CPF is involved in the recognition of the poly(A) site by specific interactions of its subunits Yhh1p/Cft1p, Ydh1p/Cft2p and Yth1p with sequences surrounding the poly(A) site (Barabino et al., 2000; Dichtl and Keller, 2001). Nab4p/Hrp1p (CF IB) binds to the EE (Kessler et al., 1997; Chen et al., 1998) and through interaction with Rna14p, possibly enables CF IA to bind to the PE (Gross and Moore, 2001b). However, it was also reported that Nab4p/Hrp1p (Minvielle-Sebastia et al., 1998) as well as the PE and EE (Dichtl and Keller, 2001) are not essential for cleavage *in vitro*, underscoring the importance of the poly(A) site region itself.

The yeast and mammalian 3' end processing factors are highly homologous (reviewed in Shatkin and Manley, 2000). In mammals cleavage and polyadenylation specificity factor (CPSF) contributes to poly(A) site selection by binding to the well conserved AAUAAA element upstream of the cleavage site (Keller et al., 1991); CPSF 160 and possibly additional CPSF subunits are thought to mediate specific interactions to the RNA (Murthy and Manley, 1995). The binding of purified CPSF is weak, but is strongly enhanced by a cooperative interaction with cleavage stimulation factor (CstF) bound to the downstream elements (Wilusz and Shenk, 1990; Gilmartin and Nevins, 1991; MacDonald et al., 1994).

Transcription by RNA polymerase II (RNAP II) and pre-mRNA processing reactions are coupled events (reviewed in (Bentley, 2002; Proudfoot et al., 2002; Howe, 2002; Proudfoot and O'Sullivan, 2002). The C-terminal domain (CTD) of RNAP II plays a central role in linking the processing reactions to transcription. Current models suggest that the CTD is hypo-phosphorylated during transcription initiation and that escape of RNAP II into the elongation phase is accompanied by phosphorylation of the CTD. It has been proposed that the change in charge upon phosphorylation enables proteins involved in pre-mRNA processing to bind to the CTD (Komarnitsky et al., 2000; Cho, 2001; Licatalosi et al., 2002). This includes proteins involved in capping (McCracken et al., 1997; Cho et al., 1997; Ho et al., 1998), splicing (Yuryev et al., 1996) and 3'end formation (McCracken et al., 1997; Rodriguez et al., 2000; Barilla et al., 2001; Fong and Bentley, 2001; Dichtl et al., 2002b; Licatalosi et al., 2002). It is assumed that the assembled proteins subsequently travel with the elongating RNAP II during transcription and act on the nascent RNA transcript.

In mammals the CTD was suggested to play a direct role in 3'end cleavage *in vivo* (McCracken et al., 1997) and *in vitro* (Hirose and Manley, 1998; Ryan et al., 2002). Experiments in yeast provided evidence that transcription in the absence of the CTD was accompanied by a reduction of cleavage efficiency and the resulting mRNAs had shorter poly(A) tails (Licatalosi et al., 2002; McNeil et al., 1998). Correct transcription termination requires a functional poly(A) signal in all organisms (Tran et al., 2001; Orozco et al., 2002), reviewed in (Proudfoot, 1989) and yeast strains carrying mutations in CPF and CF IA were shown to be deficient in correct transcription termination, indicating the coupling between 3'end processing and transcription termination (Birse et al., 1998; Barilla et al., 2001; Dichtl et al., 2002a; Dichtl et al., 2002b; Hammell et al., 2002; Sadowski et al., 2003). Furthermore, correct transcriptional termination was found to require the transcription factors Sub1p (Calvo and Manley, 2001) and Res2p (Aranda and Proudfoot, 2001) and chromatin remodeling factors (Alen et al., 2002). The Nrd1p complex was shown to be required for correct termination at snoRNA genes (Steinmetz et al., 2001).

Ydh1p/Cft2p (which will be referred to as Ydh1p in the remainder of this paper) is the 105 kDa subunit of CPF. It shares 24.4% identity and 43% similarity with the mammalian CPSF-100 protein. It is also significantly related to Ysh1p and to

CPSF-73 (Jenny et al., 1996). Ydh1p is essential for cell viability (Preker et al., 1997) and was shown to bind RNA (Zhao et al., 1997). RNase H protection experiments suggested that the protein interacts with the poly(A) site region (Dichtl and Keller, 2001). Here, we show that Ydh1p is required for cleavage and polyadenylation *in vitro* and for poly(A) site recognition *in vivo*. Furthermore, we provide evidence that Ydh1p interacts with several subunits of CPF, with Pcf11p, a subunit of CF IA, and with the CTD of RNAP II. The results suggest that Ydh1p is part of an interaction surface that mediates important contacts with CF IA, the CTD of RNAP II and the pre-mRNA substrate.

6.3 Results

6.3.1 Ydh1p is required for cleavage and polyadenylation *in vitro*

In order to investigate the role of Ydh1p in 3'end processing we generated temperature sensitive *ydh1* alleles (see Materials and Methods). Figure 6.1A shows growth curves of *ydh1* mutant and isogenic wild-type cells following shift from 25°C to 37°C. The *ydh1-1* strain displayed growth arrest at 37°C after 3 hours whereas the strains *ydh1-2* and *ydh1-3* ceased growth after approximately 5 hours. A drop-test revealed that the mutant strains did not form colonies at the restrictive temperature (Figure 6.1B).

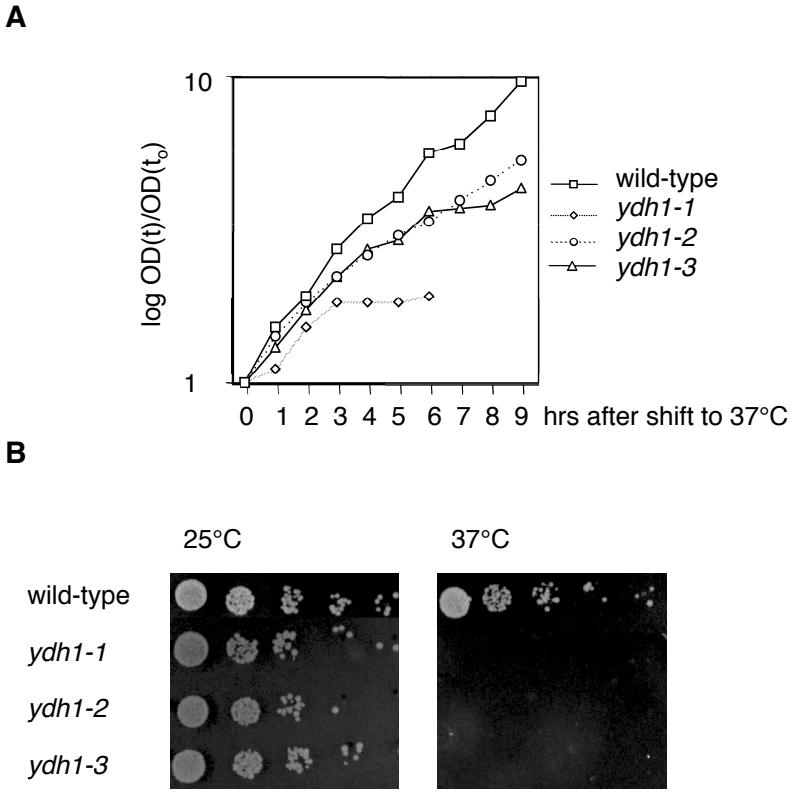


Figure 6.1 *YDH1* mutant strains display a temperature sensitive phenotype

A) Growth curves of wild-type and mutant *ydh1* strains at 37°C.

B) Ten-fold serial dilution of cultures spotted on YPD plates followed by incubation at the indicated temperature for two days.

Next, we analyzed whether mutations in *YDH1* affect cleavage and polyadenylation *in vitro* and tested extracts from wild-type, *ydh1-1*, *ydh1-2* and *ydh1-3* cells for their ability to cleave a synthetic *CYC1* pre-mRNA (Figure 6.2A). Extracts from wild-type (lanes 2 and 6), *ydh1-2* (lanes 4 and 8) and *ydh1-3* cells (lanes 5 and 9) accurately cleaved the substrate RNA at 30°C and 34°C. Notably, *ydh1-3* extract (lane 9) showed a reduced efficiency of cleavage at 34°C compared to wild-type (lane 6). In contrast, extract of *ydh1-1* cells was deficient in cleavage at either temperature (lanes 3 and 7).

To show that the deficiency in cleaving the RNA substrate is due to inactive Ydh1p or CPF, respectively, we carried out reconstitution assays either with extract which is mutant in a CF IA subunit (*rna15-1*) or with purified factors at 30°C (Figure 6.2B). The cleavage activity of *ydh1-1* extract was restored upon addition of *rna15-1*

extract (lane 4) or purified CPF (lane 6) but not by addition of purified CF IA (lane 7). As expected, the *rna15-1* extract on its own lacked cleavage activity (lane 5; (Minvielle-Sebastia et al., 1994) and purified CPF and CF IA alone were not able to cleave the substrate (lanes 8 and 9); cleavage occurred at specific and at cryptic sites when both factors were added to the reaction (lane 10); site-specific cleavage was obtained by including CF IB in the reaction (lane 11; Minvielle-Sebastia et al., 1998).
Next we analyzed whether mutations in *YDH1* affect the polyadenylation reaction of pre-cleaved *CYC1* substrate (CYC1-Pre) *in vitro* (Figure 6.2.C). No polyadenylation activity could be observed in *ydh1-1* extract at 30°C and 34°C (lanes 3 and 7). In contrast, the *ydh1-2* and *ydh1-3* extracts specifically and efficiently polyadenylated the substrate at both temperatures (lanes 4, 5, 8 and 9) comparable to wild-type (lanes 2 and 6). Specific polyadenylation activity was restored in *ydh1-1* extract by addition of purified CPF (Figure 6.2D, lane 4) but not by addition of purified CF IA (lane 5). As shown before, CPF by itself unspecifically polyadenylated the substrate (lane 6; (Ohnacker et al., 2000), whereas CF IA alone displayed no polyadenylation activity (lane 7). Specific polyadenylation activity was restored upon combination of CPF, CF IA, CF IB and Pab1p (lane 8).

Notably, addition of 100 to 250 ng recombinantly expressed Ydh1p failed to reconstitute the cleavage and polyadenylation activities in *ydh1-1* extract, possibly because the recombinant protein was not able to replace its mutant counterpart in the CPF complex (results not shown).

The above results showed that the *ydh1-1* extract was strongly reduced in cleavage and polyadenylation activities *in vitro* and that both activities could be reconstituted by addition of purified CPF. Furthermore, the cleavage activity of the *ydh1-3* mutant was lower at 34°C compared to 30°C. The results indicate that Ydh1p is involved in both steps of the 3' end processing reaction.

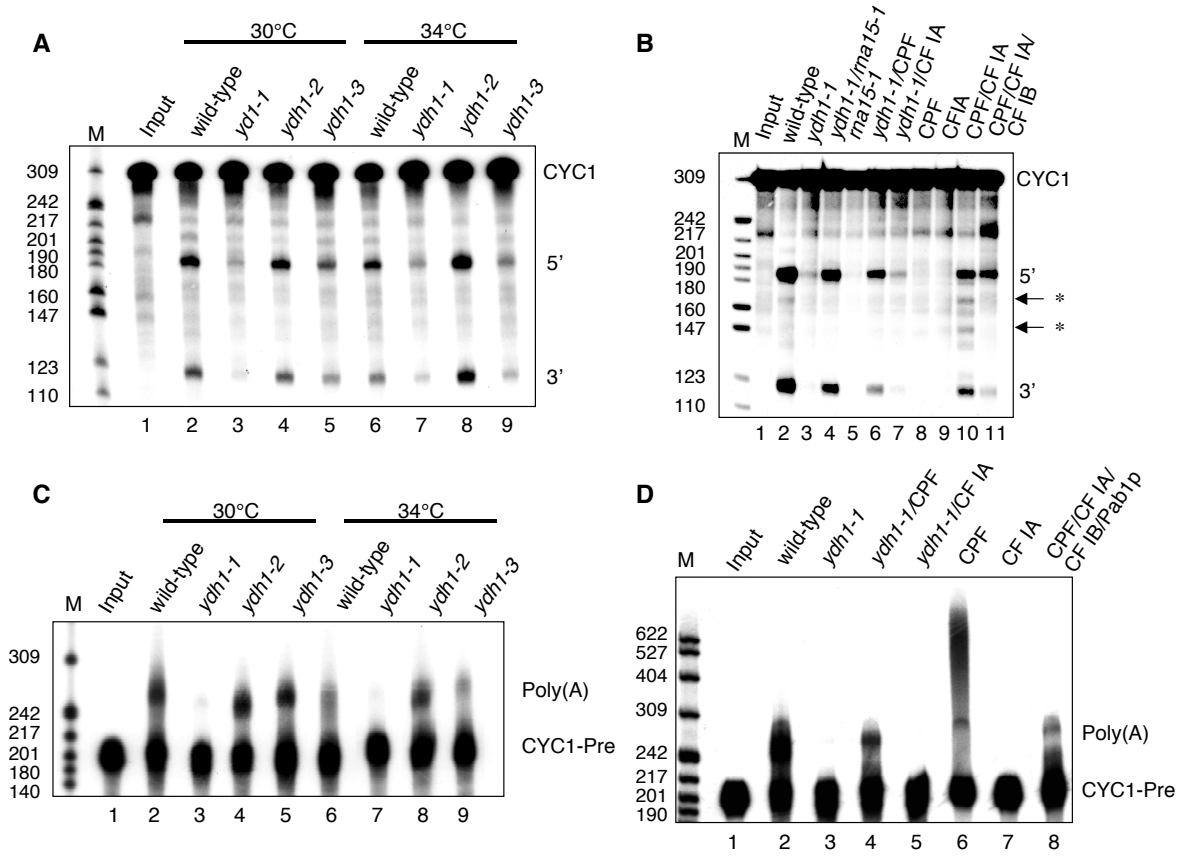


Figure 6.2 Ydh1p is required for cleavage and polyadenylation *in vitro.*

In vitro analysis of cleavage (A and B) and polyadenylation activities (C and D) of *ydh1* mutant extracts. Input lanes (1) represent mock-treated reactions. The migration positions of the *CYC1* and *CYC1*-precleaved (CYC1-Pre) RNA substrate, specific (5' and 3') and cryptic (asterisks) cleavage products and the polyadenylation products [Poly(A)] are indicated on the right of each panel. The position and size (in number of nucleotides) of the marker bands are indicated on the left.

A) Extracts prepared from yeast strains as indicated on top were monitored for their ability to cleave internally ^{32}P-labelled *CYC1* RNA substrate at 30°C and 34°C.

B) Reconstitution of cleavage activity in *ydh1-1* extract at 30°C. As indicated on top, *ydh1-1* extract was combined with equal amounts of *rna15-1* extract (lane 4), purified CPF (lane 6) or CF IA (lane 7).

C) Specific polyadenylation was analysed at 30° and 34°C with internally ^{32}P-labelled precleaved *CYC1* RNA substrate that ends at the natural poly(A) site.

D) Reconstitution of specific polyadenylation activity in *ydh1-1* extract at 30°C. *ydh1-1* extract was combined with purified CPF (lane 4) or CF IA (lane 5).

6.3.2 mRNAs are unstable in *ydh1* mutant strains at restrictive temperature

mRNAs without poly(A) tails are rapidly degraded in living cells. Yeast strains with a 3' end processing deficiency are therefore expected to under-accumulate mRNAs after shift to the restrictive temperature. For this reason, we performed Northern blot analyses on total RNA extracted from strains grown at 25°C and after shift to 37°C (Figure 6.3A). The amount of 18S rRNA served as control for the loading of the RNA (panel VI). *ADH1*, *ACT1*, *CYH2* and *CYC1* mRNA levels diminished in the *ydh1-1* mutant cells after two hours shift to 37°C (panels I-IV, lanes 3-6). The level of the stable *PGK1* mRNA ($t_{1/2}$ = 45 min) dropped significantly only after the cells were shifted to 37°C for 6 hours (panel V, lanes 3-6). In comparison the *ADH1*, *ACT1*, *CYH2* and *CYC1* mRNA levels of the *ydh1-2* mutant strain did only drop slightly after 4 hours and the *PGK1* mRNA level remained stable even after shift to 37°C for 6 hours (panels I-V, lanes 7-10). The mutant strain *ydh1-3* under-accumulated *ACT1* mRNA already after 2 hours at 37°C; *ADH1*, *CYH2* and *CYC1* mRNA levels were reduced after 4 hours (panels I-IV, lanes 11-14). The *PGK1* mRNA level showed a slight decrease after 4 hours and more so after 6 hours (panel V, lanes 11-14). The RNA polymerase II transcribed U24, U14 and snR13 snoRNAs were stable in all mutants tested, even after shift to 37°C for 6 hours (panels VII to IX), indicating that snoRNA processing was not affected in *ydh1* mutant strains. These results showed that *ydh1* mutant strains under-accumulated mRNAs at restrictive temperature. The *ydh1* mutant phenotypes are possibly due to a deficiency of the strains in poly(A) site recognition (see below) which might impair the efficiency in 3' end processing.

6.3.3 Ydh1p is required for poly(A) site selection of *ACT1* pre-mRNA

The previous observation that Ydh1p binds RNA around the poly(A) site (Dichtl and Keller, 2001), raised the possibility that the protein is required for poly(A) site recognition. To test this, we analyzed poly(A) site usage of *ACT1* pre-mRNA in wild-type and mutant cells (Mandart and Parker, 1995). The *ACT1* 3' untranslated region (UTR) contains at least four polyadenylation sites, of which the most proximal one (site I) is used with highest frequency in wild-type cells (Figure 6.3B; lanes 1 and 2).

All tested *ydh1* mutants used site I three times more often then site IV at 25°C (lanes 3, 5 and 7). At restrictive temperature, however, processing shifted from site I towards site IV in a ratio of 1:1 in *ydh1-1* and *ydh1-3* cells (lanes 4 and 8); the *ydh1-2* mutant showed a ratio of 2:1 (lane 6). These results indicated that mutations in *ydh1-1*, *ydh1-2* and *ydh1-3* impair poly(A) site recognition *in vivo*. To verify that the *ACT1* mRNAs were polyadenylated we incubated the reactions with RNase H and the oligos ACT1-RnaseH and dT to digest the poly(A) tails. This resulted in a slight downward shift and sharpening of the RNA bands. This showed that the *ACT1* mRNAs were polyadenylated (results not shown).

Western blot analyses were carried out to test whether the observed phenotypes were the result of an under-accumulation of Ydh1p or other subunits of the 3' end processing machinery (Figure 6.3C). As control, purified CPF or CF IA were analyzed in parallel (lane 1). In wild-type and *ydh1-2* cells Ydh1p was stable even after 6 hours at 37˚C (panel I, lane 2, 3, 7-9), whereas it was reduced in *ydh1-1* and *ydh1-3* cells after 3 hours at at 37˚C (lanes 5, 6, 11 and 12). Ysh1p/Brr5p, Pfs2p and Ssu72p (panels II, IV and VII) under-accumulated in *ydh1-1* and *ydh1-3* cells after 6 hours at 37˚C (lanes 6 and 12). The levels of Pap1p, Fip1p and Yth1p were constant even after 6 hours at 37˚C (panels III, V, VI). In addition, mutations in *YDH1* did not affect the levels of the CF IA subunits Rna14p and Rna15p (panels VIII, IX).

These results indicated that the defects observed in the *ydh1* mutant strains might be caused by a destabilization of the Ydh1p protein. Since the destabilization at restrictive temperature of Ysh1p/Brr5p, Pfs2p and Ssu72p was detectable at later time-points compared to the appearance of phenotypes (see Northern blot and *ACT1* poly(A) site selection), we consider it unlikely that lower levels of factors other than Ydh1p are responsible for the deficiencies of the *ydh1* mutant cells.

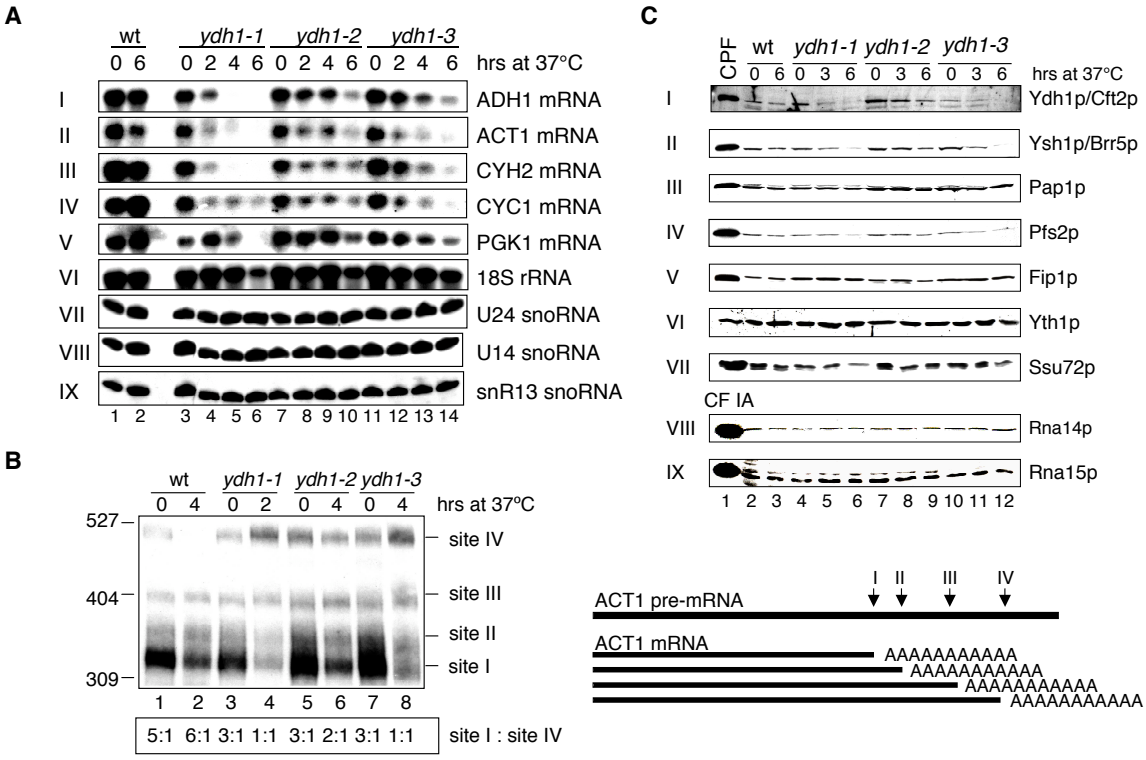


Figure 6.3 mRNAs are unstable in *ydh1* mutant strains at restrictive temperature and the mutant strains are deficient in the recognition of *ACT1* poly(A) site *in vivo*

A) Northern analysis of total RNA extracted from wild-type and mutant *ydh1* cells grown at 23°C, or after shift to 37°C for 2, 4 and 6 hours as indicated on top of each lane. The RNAs were separated on formaldehyde/1.2% agarose gels (panels I-VI) or 8.3 M urea/8% polyacrylamide gels (panels VII-IX). The filters were developed with random-primed labeled probe (panel I-V) or end-labeled oligonucleotides (panels VI to IX) directed against the RNA species indicated at the right of each panel.

B) Analysis of *ACT1* poly(A) site usage in wild-type and mutant *ydh1* cells. The drawing on the right shows the relative positions of the different poly(A) sites. Total RNA extracted from wild-type and mutant *ydh1* cells after growth at 23°C, or following shift to 37°C as indicated on top of each lane, was treated with an oligonucleotide complementary to the 3' region of the ACT1-mRNA (ACT1-RnaseH) and RNaseH. The RNAs were separated on an 8 M urea/6% polyacrylamide gel and the filters were incubated with random primed labeled probe directed against the 3' end of the *ACT1* mRNA. RNA levels were quantified by PhosphorImager scanning (Molecular Dynamics); the ratios of poly(A) site I : site IV usage for each lane are indicated at the bottom.

C) Western analysis of wild-type and *ydh1* mutant extracts prepared from cells grown at 23°C, or after shift to 37°C for 3 and 6 hours. Equal amounts of total protein were loaded in each lane. Lane 1 shows purified CPF (panels I-VII) and CF IA (panels (VIII and IX). The filters were treated with antibodies directed against the proteins indicated at the right.

6.3.4 Ydh1p interacts with the C-terminal domain of RNA polymerase II

The CTD of RNAP II plays an important role in coupling transcription and 3' end formation (reviewed in (Proudfoot and O'Sullivan, 2002; Bentley, 2002). We were interested whether Ydh1p contributes to this process by physically interacting with the CTD. To test this, we carried out *GAL4* based yeast two hybrid tests and assayed for the activation of expression of the *HIS3* and *lacZ* reporter genes. Strain Y190 was co-transformed with plasmids carrying the *GAL4* DNA binding domain fused to the CTD (pBD-CTD) and the *GAL4* activation domain fused to test genes. As positive controls we analyzed the known CTD interactors *YHH1/BRR5* (pACT2-YHH1) and *PCF11* (pACT2-PCF11) (Dichtl et al., 2002b; Licatalosi et al., 2002; Sadowski et al., 2003) and pACT2-YSH1 as negative control. As shown in Figure 6.4A the plasmids carrying *YDH1, YHH1* and *PCF11* in combination with the plasmid pBD-CTD enabled cells to grow on medium lacking histidine under stringent conditions (in the presence of 35 mM 3'-aminotriazole). In contrast, no growth was observed with pACT2-YSH1 or when empty pBD and pACT2 plasmids were tested. These results were confirmed by an X-Gal filter lift assay, which monitors β-galactosidase expression (results not shown). Thus, *YDH1* showed a two-hybrid interaction with the CTD of RNAP II.

To confirm that Ydh1p interacts with the CTD, and to test whether this interaction is influenced by the phosphorylation state of the CTD we performed GST pull-down experiments with *in vitro* translated radioactively labeled proteins and recombinant GST-CTD fusion protein. Ydh1p interacted with the CTD (Figure 6.4B, lane 3) and this interaction was enhanced upon phosphorylation of the CTD (lane 4); no signal was observed with GST alone (lane 2). Ysh1p/Cft1p interacted with neither form of the GST-CTD protein. These results suggested that Ydh1p binds the CTD and that the interaction is enhanced upon phosphorylation of the CTD.

Correct transcription termination by RNAP II requires a functional poly(A) site on the nascent RNA and the CTD plays an important role in coupling transcription and pre-mRNA processing (reviewed in Proudfoot, 1989). The observations that Ydh1p is involved in poly(A) site recognition and that it interacted

with the CTD raised the possibility that Ydh1p might also be involved in transcriptional termination. Therefore, we carried out transcriptional run-on analysis with the well-characterized *GAL1/10* controlled *CYC1* gene *in vivo* (Birse et al., 1998). We did not observe a significantly increased RNAP II density downstream of the *CYC1* terminator in *ydh1-1* and *ydh1-3* cells grown at 25°C or following shift to 37°C compared to wild-type (results not shown). Similarly, we did not detect the accumulation of read-through products in *ydh1* strains in reverse transcription experiments with primers annealing downstream of the cleavage site of different snoRNA species (snR39b, snR45, snR3, snR50, snR71; (Steinmetz et al., 2001). This suggested that Ydh1p is not generally required for termination of RNAP II during transcription of pre-mRNAs and snoRNAs.

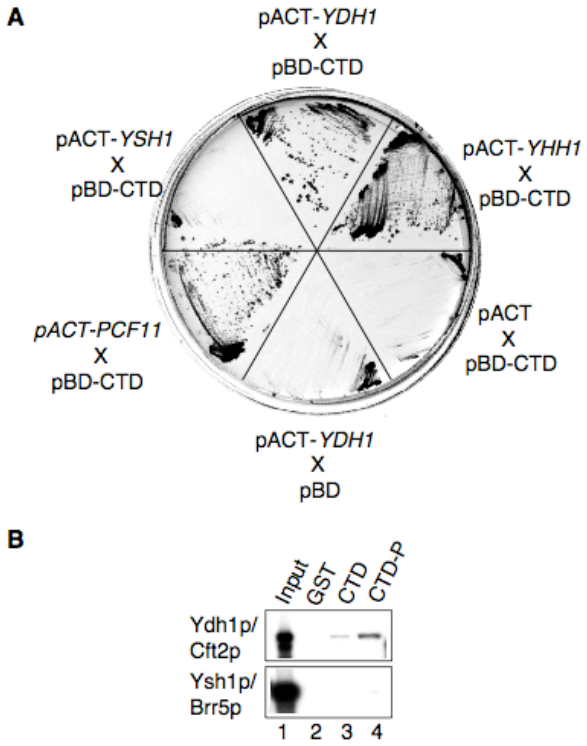


Figure 6.4 Ydh1p interacts with the CTD of RNAP II

A) *GAL4* based two hybrid analysis. Y190 cells were co-transformed with a plasmid encoding the *GAL4* DNA binding domain fused to the CTD (pBD-CTD) and plasmids (pACT2) encoding the *GAL4*

activation domain fused to the genes indicated. Transformants were tested for expression of the *HIS3* reporter gene on medium lacking histidine, and containing 35 mM 3-amino-1,2,4-triazole. As control, pBD CTD was co-transformed with the empty pACT2 plasmid respectively pACT2-YDH1 with the empty pBD plasmid.

B) Pull-down experiments with 1 µg GST (lane 2), GST-CTD (lane 3) and phosphorylated GST-CTD (lane 4) and *in vitro* translated, [^{35}S]methionine labeled proteins (indicated on the left). Lane 1 shows 10% of the *in vitro* translated reactions used in the assay. Bound proteins were separated by SDS-PAGE and visualized by autoradiography.

6.3.5 Ydh1p interacts with other subunits of CPF and with Pcf11p, a subunit of CF IA

As Ydh1p is part of a multi-protein complex we were interested to examine with which of the other CPF subunits the protein interacts. Therefore, we expressed a N-terminally GST-tagged and C-terminally His-tagged version of the Ydh1 protein (G-Ydh1-H) in *E. coli* and carried out GST pull-down experiments with radioactively labeled *in vitro* translated proteins. As shown in Figure 6.5A, G-Ydh1-H interacted with itself and with the CPF subunits Yhh1p/Cft1, Ysh1p/Brr5p, Pta1p, Pfs2p, Ssu72p, YDL094cp (panels I – VII) and the CF IA subunit Pcf11p (panel X). G-Ydh1-H did not pull-down the CPF subunits Fip1p and Yth1p (panels VII and IX) nor the CF IA subunits Rna14p, Clp1p and Rna15p (panels XI – XIII), Nab4p/Hrp1 (CF IB) or Pab1p (panels XIV and XV). Since we identified a large number of potential Ydh1p interaction partners, we decided to investigate whether these interactions were specific and therefore assignable to distinct regions of Ydh1p. For this purpose the pull-downs were repeated with C-terminally truncated Ydh1p proteins (Figure 6.5B). Figure 6.5A shows that the very C-terminus was required for interaction of Ydh1p with itself (panel II) as most of the interaction was lost upon deletion of the C-terminal 137 amino acids. The N-terminal 246 amino acids were sufficient for interaction with Pfs2p (panel V) and central protein sequences most likely mediated interactions with Yhh1p/Cft1p, Ysh1p/Brr5p, Pta1p, Ssu72, pYDL094c and Pcf11p (panels I, III, IV, VI, VII and X). These results suggested that the Ydh1p subunit might contribute to the stability and structural order of CPF and bridges the two factors CPF and CF IA.

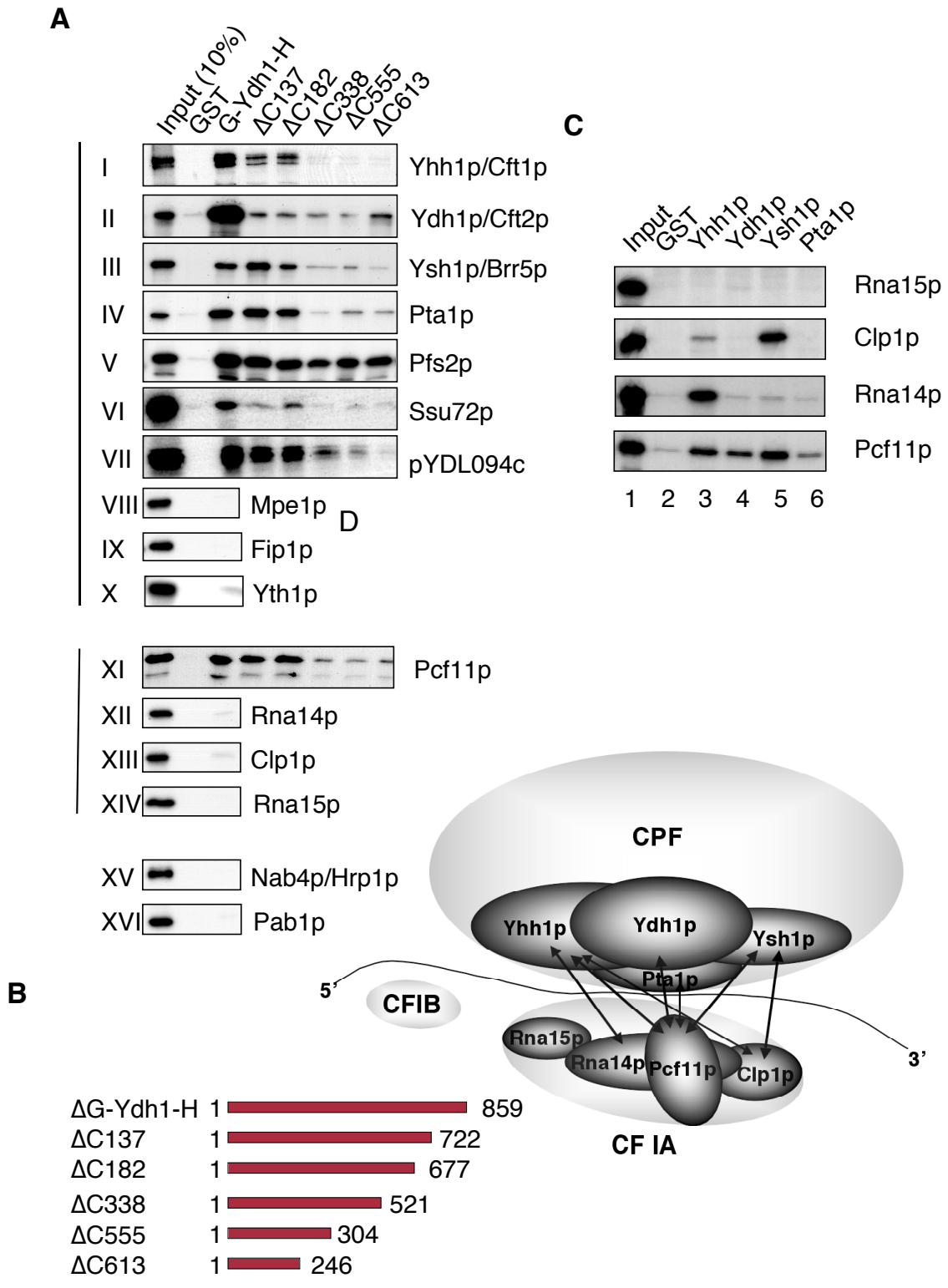


Figure 6.5.Ydh1p interacts with other subunits of CPF and with Pcf11p, a subunit of CF IA

A) GST pull-down experiments with 0.5 µg GST, 100 ng G-Ydh1-H (Ydh1p) or 100 ng of the different C-terminally truncated, GST-tagged Ydh1 fragments (as indicated on top of each lane) and *in vitro* translated, [^{35}S]methionine labeled proteins (indicated on the right). The first lane shows 10% of the *in vitro* translated protein used in the binding reactions.

B) Schematic drawing of the full length and C-terminal truncations of Ydh1p. The numbers indicate the length of the constructs in amino acids.
C) GST pull-down experiment with GST-tagged recombinant proteins (indicated on top of each lane) and *in vitro* translated, [^{35}S]methionine labeled proteins (indicated on the right).
D) Model of protein-protein interaction surface (not all CPF subunits and protein interactions are shown).

To obtain a better understanding of how CPF interacts with CF IA we decided to assay for further interactions between the two factors. Figure 6.5C shows a GST pull-down with *in vitro* translated protein and GST-tagged recombinant proteins. Of the analyzed CPF subunits we found that only Yhh1p/Cft1p interacted strongly with Rna14p. Yhh1p/Cft1p also bound to Pcf11p and more weakly to Clp1p (lane 3). Ydh1p pulled-down only Pcf11p (lane 4). Ysh1p/Brr5p interacted strongly with Clp1p and Pcf11p (lane 5). A weak signal was observed for the interaction between Pta1p and Pcf11p (lane 6).
These results suggested that Ydh1p, Yhh1p/Cft1p, Ysh1p/Brr5p and possibly also Pta1p are involved in forming a protein-protein interaction surface between CPF and CF IA. Our combined results indicate that Ydh1p plays an important role in determining the RNA binding specificity of CPF and in the assembly of the 3' end formation machinery and its tethering to RNAP II.

6.4 Discussion

The yeast and mammalian 3' end formation machineries display a surprisingly complex subunit composition. At present up to twenty polypeptides are known that constitute the factors that catalyze the yeast 3' end formation reaction *in vitro* (Gavin et al., 2002; Dichtl et al., 2002b). The ongoing characterization of the components suggests distinct and specialized tasks of the individual polypeptides in both steps of the 3' end processing reaction as well as in the coupling of 3' end formation to transcription. In this work we characterized the role of the yeast CPF subunit Ydh1p in these processes.

To analyze if Ydh1p is necessary for cleavage and polyadenylation we produced conditional *ydh1* mutant strains and carried out *in vitro* assays. The

experiments revealed a deficiency of the mutant extracts in both steps of the 3' end processing reaction. The activity could be rescued upon addition of purified CPF. mRNAs without a poly(A) tail are prone to rapid degradation *in vivo*. Northern analysis revealed that the levels of a number of different mRNAs were reduced at restrictive temperature in Ydh1p mutants. The results support a requirement for Ydh1p in both steps of 3' end processing. However, we do not expect Ydh1p to be directly involved in the catalysis of the 3' end processing reaction, because it took several hours at the restrictive temperature before the mRNA levels were reduced. Furthermore, the *ACT1* poly(A) site selection assay revealed that a substantial amount of pre-mRNA appeared to be cleaved and polyadenylated *in vivo* at the restrictive temperature in all mutants.

We reported previously that Ydh1p binds to a *CYC1* pre-mRNA around the poly(A) site; this suggested a role for the protein in poly(A) site selection (Dichtl and Keller, 2001). Here we show that Ydh1p is involved in poly(A) site selection of *ACT1* pre-mRNA *in vivo*, as all analyzed *ydh1* mutant strains displayed more frequent use of alternative poly(A) sites compared to the wild-type strain. These results underscore our previously postulated model that Ydh1p contributes to poly(A) site selection (Dichtl and Keller, 2001). In contrast to mammalian polyadenylation signals, the *cis* acting sequences in yeast are highly degenerate (reviewed in Zhao et al., 1999a; Guo and Sherman, 1995). We proposed that the RNA binding subunits of CPF act in concert to achieve specific recognition of the correct poly(A) site (Dichtl and Keller, 2001). The preferential use of cleavage sites located downstream of the major site in Ydh1 mutants may be caused by the coupling of transcription elongation and 3' end processing. Reduced RNA binding efficiency of CPF containing mutant Ydh1p could lead to skipping of the first cleavage site of the nascent pre-mRNA emerging from the elongating RNAP II. Ydh1p (Dichtl and Keller, 2001), Yhh1p/Cft1p (Dichtl et al., 2002b) and Yth1p (Barabino et al., 2000) were shown to bind preferentially to U-rich elements localized directly upsteam and downstream of the poly(A) site, thus colocalizing with the region to which the complete CPF factor binds (Dichtl and Keller, 2001). These interactions with the RNA substrate are thought to be crucial for poly(A) site selection and for cleavage activity. Thus, the observed *in vitro* cleavage deficiency of the *ydh1* mutant cells might result from insufficient poly(A) site selection. It remains to be determined how Ydh1p interacts with RNA. The primary

sequence of the protein does not display any clear similarities to known RNA binding domains. Preliminary RNA binding experiments with portions of Ydh1p indicated that the RNA binding activity is distributed throughout the entire length of the protein (results not shown). However, detailed analyses have not been done yet. Interestingly, RNA binding activity has not been demonstrated for the mammalian homologue of Ydh1p, CPSF 100 kDa. Considering the highly conserved *cis* acting elements in mammalian pre-mRNAs (reviewed in Zhao et al., 1999a) it seems possible that a smaller set of proteins is sufficient for specific poly(A) site recognition, whereas in yeast a cooperative interplay of more RNA binding proteins is required to recognize the more degenerate sequence elements in a specific fashion.

We have shown that Ydh1p interacted with the CTD of RNAP II both in a two-hybrid test and in a GST pull-down assay. This interaction was enhanced upon phosphorylation of the CTD. The only other CPF subunit found to interact with the phosphorylated CTD is Yhh1p/Cft1p (Dichtl et al., 2002b). We suggest that Ydh1p might be involved in tethering CPF to transcribing RNAP II and thereby contributes to the coupling of 3' end processing and transcription. However, the *ydh1* mutants did not reveal a defect in transcriptional termination at the *CYC1* terminator. As this was the only pre-mRNA that was tested by transcriptional run-on analysis we cannot exclude that Ydh1p might be involved in transcriptional termination of other genes. Moreover, reverse transcription analysis on snoRNAs in *ydh1* mutant strains did not reveal 3' extended transcripts, indicating that Ydh1p was not required for snoRNA termination either. We propose that Ydh1p helps to tether CPF to elongating RNAP II but that it is not needed for general transcription termination. Ydh1p and Yhh1p/Cft1p show a phenotypic similarity in that both proteins bind RNA, are involved in poly(A) site recognition and bind to the CTD (Dichtl et al., 2002b). However, the observation that mutants in *yhh1* are severely impaired in transcription termination at the *CYC1* gene (Dichtl et al., 2002b) whereas *ydh1* mutants have no general termination defect, indicates that the functional roles of the individual 3' end formation factor subunits can be substantially different.

It is not well understood how the subunits of the CPF and CF IA factors assemble to form functional 3' end processing complexes on the RNA substrate. CPF consists of up to fifteen polypeptides, which interact with each other and form a stable factor. We found that Ydh1p interacted with the CPF subunits Yhh1p/Cft1p,

Ysh1p/Brr5p, Pta1p, Pfs2p, Ssu72p, and pYDL094c. These many interactions indicate that the protein contributes substantially to the assembly and structural order of the CPF factor. Strikingly, Ydh1p also bound strongly to itself in the GST pull-down assays. This may indicate that there is more than one Ydh1p molecule present per 3' end processing unit. However, silver stain analysis of purified CPF factor suggested an apparent stoichiometric presence of Ydh1p compared to other CPF subunits (Dichtl et al., 2002b). Interestingly, we found that the CF IA subunit Pcf11p interacted with the CPF components Ydh1p, Yhh1p/Cft1p, Ysh1p/Brr5p and weakly with Pta1p. In addition, we showed that the CPF subunits Yhh1p/Cft1p and Ysh1p/Brr5p interacted with the CF IA subunits Rna14p and Clp1p, respectively. The latter interactions were also observed between the homologous subunits of the mammalian 3' end formation machinery. CPSF 160 kDa interacts with the CstF 77 kDa protein (Murthy and Manley, 1995) and CPSF 73 kDa interacts with hClp1p (de Vries et al., 2000). So far, only Pfs2p and Fip1p were shown to bridge CPF and CF IA (Preker et al., 1995; Ohnacker et al., 2000). We propose that Ydh1p, Yhh1p/Cft1p, Ysh1p/Brr5p, Pfs2p, Fip1p and possibly Pta1p contribute to a protein-protein interaction surface which acts in the assembly of the 3' end formation machinery and which appears to be conserved in evolution (Figure 5D).

Our analysis of Ydh1p suggests that the protein is an important constituent of protein interaction surfaces which act in the association of 3' end processing factors with RNAP II, the recognition of poly(A) signal sequences and the assembly of an active 3' end formation complex on the RNA substrate.

6.5 Experimental procedures

6.5.1 Yeast strains

For random mutagenesis of *YDH1*, mutagenic PCR was carried out with a low concentration of ATP (Ohnacker et al., 2000). The primers used for PCR were: Ydh1-N (5' CCCTTACGGATTGAAGTCATT 3') and Ydh1-C (5' TTGAACCTTTTATTTGTGCTG 3'). Plasmid pBD63 (YDH1-LEU2-CEN) was obtained by subcloning of the *YDH1* gene from pIA115 (YDH1-URA3-CEN; 53) into

the *BamHI* and *SacI* restriction sites of pRS415. The plasmid was digested with the restriction enzymes *NsiI* and *SpeI*. The fragment containing plasmid sequences and sequences flanking the *YDH1* gene was co-transformed together with the mutagenized PCR product into the yeast strain YPP106 (Preker et al., 1997). Transformants were selected on minimal medium lacking leucine and replica-plated onto 5-FOA plates in order to shuffle out the pIA115 plasmid. The colonies were then replica-plated onto YPD-plates and incubated at 25˚C, 33˚C or 37˚C respectively. Plasmids of candidate colonies that showed a ts phenotype were isolated by standard procedures. The conditional growth phenotype was then verified by retransformation of the isolated plasmids into YPP106, 5-FOA treatment and growth tests at elevated temperatures.

Genotypes of yeast strains used in this study were: YPP106: MAT∝; *ura3-1; ade2-1; leu2-3,112; his3-11,15; trp1Δ; ydh1::TRP1* [pIA 115; *CEN4 – URA3 – YDH1*] (Preker et al., 1997); YAK1: MAT∝; *ura3-1; ade2-1; leu2-3,112; his3-11,15; trp1Δ; ydh1::TRP1* [pAK21 *CEN - LEU2 – ydh1-1*]; YAK2: MAT∝; *ura3-1; ade2-1; leu2-3,112; his3-11,15; trp1Δ; ydh1::TRP1* [pAK22 *CEN - LEU2 – ydh1-2*]; YAK3: MAT∝; ura3-1; ade2-1; leu2-3,112; his3-11,15; trp1Δ; *ydh1*::TRP1 [pAK23 *CEN - LEU2 – ydh1-3*]; Y190: *ura3-52; trp1-901; ade2-101; leu2-3; 112 his3-200r; gal4D; gal 80D; URA3:GAL1-lacZ; LYS2::GAL1-HIS3*; cyhr; Clonetech.

6.5.2 Plasmids and primers

The plasmids encoding the C-terminally truncated Ydh1p fragments were obtained by digestion of plasmid pBD71 (Dichtl and Keller, 2001) with restriction enzymes *AccI* (pBD91; encoding the recombinant protein ΔC338), *BamHI/ XbaI* (pBD92; encoding ΔC182), *BamHI/ SpeI* (pAK5; encoding ΔC137), *BamHI/ AflII* (pAK6; encoding ΔC555), *BamHI/ AgeI* (pAK7; encoding ΔC613). GST-Ysh1p was encoded by pBD38, which was obtained by subcloning of Ysh1p into p20 (GST expression vector) using the restriction enzymes *NdeI* and *BamHI*. GST-Tev-Pta1p was encoded by pBD51, which was obtained by PCR amplification of Pta1p followed by digestion with *NdeI* and *BamHI* and ligation into p26 (GST-Tev expression vector). GST-Yhh1p-H_6 was encoded by pBD75 (Dichtl et al., 2002b).

The plasmid encoding the protein used to produce antibody directed against Ydh1p was constructed by digestion of pQE-9 (His 6 expression vector Qiagen) with

BamHI/ HindIII. The insert was constructed by PCR, amplifying the sequence between primer Ydh5' (5' ATCGCGGATCCATGACTTATAAATACAATTG 3') and Ydh3' (5' AGCCCAAGCTTATTTACTCAATTCGTTTGGT 3') of *YDH1*. For details about pBD-CTD, pACT2-*YHH1*, pACT2-*YSH1* and pACT2-*PCF11* see (Dichtl et al., 2002b).

6.5.3 Expression of recombinant proteins in *E. coli*

BL21 *E.coli* cells carrying the respective plasmid were grown at 25˚C in 2xYT until they reached an OD_{600} of approximately 1. Following induction by 0.5 mM IPTG incubation was continued for 6 hours. The proteins were purified at 4˚C on glutathione-Sepharose 4B as recommended (Pharmacia) and the protein was eluted with GST-elution buffer (75 mM KCl, 50 mM Tris-HCl (pH 7.9), 10% glycerol, 10 mM glutathione (reduced), 0.01% NP-40, 1 mM DTT).

6.5.4 Protein-protein interactions

In vitro translations were performed with the TNT-coupled transcription-translation system (Promega). 100 ng GST fusion protein was incubated with *in vitro* translated [^{35}S]methionine labeled proteins for 1 hour. The mixture was bound in a total volume of 860 µl to 20 µl glutathione sepharose (Pharmacia), which was previously equilibrated in 1 ml PBS, 0.01% NP40 and 100 µg BSA. The matrix was washed 3 times with IPP150 (150 mM KCl, 20 mM Tris-Cl (pH 8.0), 0.01% NP-40) and the proteins were eluted by addition of protein loading buffer and incubation at 95˚C. Bound proteins were separated by SDS-PAGE and visualized by autoradiography. Phosphorylation of GST-CTD was performed as described previously (Hirose and Manley, 1998) and the assay of the CTD interaction was carried out as described (Dichtl et al., 2002b). The two hybrid tests were carried out as described (Fromont-Racine et al., 1997).

6.5.5 Extract preparation and *in vitro* cleavage and polyadenylation assays

Extracts competent for *in vitro* processing were prepared following the procedure previously described (Ohnacker et al., 2000). The cleavage and polyadenylation assays were carried out as described by (Minvielle-Sebastia et al., 1994). To restrict the assay to cleavage only, EDTA replaced MgAc and CTP replaced ATP. For each reaction 30-40 μg total protein were used and, in the case of the reactions carried out at 34°C, pre-incubated at this temperature for 5 min. The RNA substrates were prepared by run-off transcription following the previously described procedure (Dichtl and Keller, 2001).

6.5.6 RNA analyses

Northern analyses and RNase H experiments were carried out as described (Dichtl et al., 2002b). In addition we employed the oligonucleotides anti-U24 (5' TCAGAGATCTTGGTGATAAT 3') and anti–snR13 (5' GGCAAAAGCCAAACAGCAACTCGAGCCAAATGCACTCATA TTCATCATAT 3'), which were labeled with γ-[^{32}P]-ATP by T4 polynucleotide kinase.

The reverse transcription analysis was performed as described (Dichtl et al., 2002b) with primers downstream of snoRNA genes as previously described (Steinmetz et al., 2001).

6.5.7 Protein extraction for Western blotting

The cells were grown at 25°C and shifted to 37°C, during incubation their growth was kept in the log phase. At each time point 40 ml of the culture (OD_{600} = 0,4) were harvested. The following procedure was carried out on ice. The cells were resuspended in IPP150 (150 mM KCl, 20 mM Tris-Cl; pH 8.0, 0.01% NP40) and an equal volume of glass beads was added. The cells were opened by rigorous vortexing. 5 ml IPP150 and protease inhibitors were added and the mixture was centrifuged for 1 hour at 10'000 rpm. 4 ml of the supernatant were thereafter centrifuged for 2 hours at

41’000 rpm and the protein was concentrated by centrifugation in a centricon YM10 (Millipore). The protein concentration was determined by Bradford analysis and equal amounts of total proteins were separated by SDS-PAGE.

Final Discussion

For many years 3' end processing, capping, splicing and transcription have been studied separately from each other. It was and is very useful to be able to carry out the different reactions independently, in order to solve questions about the individual steps. Nevertheless, *in vivo* the pre-mRNA processing steps and transcription are coupled and it is important to understand their influence on each other. For example, in mammals a big number of genes can be alternatively spliced leading to a much broader number of diverse proteins than what could be obtained without alternative splicing, and transcription was shown to be involved in this process.

Even though it is known that all these reactions are coupled *in vivo* there is quite little information about how exactly this coupling is mediated. A big player seams to be the CTD of RNAP II given that proteins of all pre-mRNA processing steps have been found to interact with it.

In this work we focused on the coupling between splicing and 3' end processing. We found that direct interactions between the U2 snRNP and CPSF are involved in coupling the two processes. At present we do not know the molecular mechanism of this coupling. However, CPSF is an important factor for poly(A) site selection and the U2 snRNP is known to be involved in the binding to the branch point sequence. Therefore, it is an intriguing idea that the two factors help recruit and stabilize each other onto the pre-mRNA.

Proteins of the spliceosome and 3' end processing factors travel with the elongating RNA polymerase. Therefore, it is possible that 3' end formation subunits influence the splicing of not only the last but also of internal introns. Recently coupled splicing and transcription assays have been published and it would be fascinating to establish a coupled transcription, splicing and 3' end formation assay. With such a system one could for example address the question if the coupling between 3' end processing and splicing comes into play only at the last intron if the reactions take place cotranscriptionally.

3' end formation and splicing are highly conserved from mammals to yeast and the model system *S. cerevisiae* allows easier *in vivo* analysis and genetic manipulations than mammalian cells. Furthermore, there is almost nothing known about the coupling of splicing and 3' end formation in this organism. Given that only

very few yeast genes carry introns and that these introns are most often located close to the 5' end of the pre-mRNA did not prompt too much research in this field. However, our results suggest that there is indeed coupling between 3' end processing and splicing taking place in yeast and we think that it would be very promising to continue with the research in this area.

Even though many RNA processing reactions are coupled this is not the case for all of them. It was proposed that tRNA splicing and pre-mRNA 3' end processing are interlinked also. However, we showed here that this hypothesis is most likely not true.

Before one can look at coupling mechanisms between different processes it is important to understand at least the basics of the individual reactions. Most of the proteins involved in 3' end processing in yeast have been identified. In this work we characterized the function of the CPF subunit Ydh1p in 3' end formation. However, the role of other subunits of the 3' end processing machinery are still not known. And more questions need to be answered before we have a better understanding of how exactly 3' end processing is taking place. For example it is still not known which enzyme catalyzes the endonucleolytic cleavage and Ysh1p, the homologue of CPSF73 is a hot candidate at present. Interestingly, like Ysh1p, Ydh1p belongs to the ß-lactamase families which have been implicated in nucleotide cleavage. Unlike Ysh1p, Ydh1p contains mutations in the sequences believed to be important for activity. These enzymes were shown to act as dimers. It is possible that Ysh1p could form a homodimer but it could also be that Ydh1p and Ysh1p have to interact together in order to obtain enzymatic activity. We showed here that the two proteins indeed directly interact with each other. Furthermore, Ydh1p binds RNA, whereas Ysh1p has not been shown to directly interact with the substrate. It is therefore an interesting model that Ydh1p could interact with Ysh1p and thereby make the enzyme active and tether it to the RNA.

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.

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing. J. Biol. Chem. *273*, 12685-12688.

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.

Alen, C., Kent, N. A., Jones, H. S., O'Sullivan, J., Aranda, A., and Proudfoot, N. J. (2002). A role for chromatin remodeling in transcriptional termination by RNA polymerase II. Mol. Cell *10*, 1441-1452.

Amrani, N., Minet, M., Le Gouar, M., Lacroute, F., and Wyers, F. (1997a). Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Mol. Cell. Biol. *17*, 3694-3701.

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

Aranda, A., Perez-Ortin, J. E., Moore, C., and del Olmo, M. (1998). The yeast FBP1 poly(A) signal functions in both orientations and overlaps with a gene promoter. Nucleic Acids Res. *26*, 4588-4596.

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

Aravind, L. (1999). An evolutionary classification of the metallo-beta-lactamase fold proteins. In Silico Biol. *1*, 69-91.

Awasthi, S., and Alwine, J. C. (2003). Association of polyadenylation cleavage factor I with U1 snRNP. RNA *9*, 1400-1409.

Bai, C., and Tolias, P. P. (1996). Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger protein. Mol. Cell. Biol. *16*, 6661-6667.

Baldi, M. I., Mattoccia, E., Bufardeci, E., Fabbri, S., and Tocchini-Valentini, G. P. (1992). Participation of the intron in the reaction catalyzed by the Xenopus tRNA splicing endonuclease. Science *255*, 1404-1408.

Barabino, S. L. 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.

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

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.

Bardwell, V. J., Wickens, M., Bienroth, S., Keller, W., Sproat, B. S., and Lamond, A. I. (1991). Site-directed ribose methylation identifies 2'-OH groups in polyadenylation substrates critical for AAUAAA recognition and poly(A) addition. Cell *65*, 125-133.

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

Behrens, S. E., Tyc, K., Kastner, B., Reichelt, J., and Lührmann, R. (1993). Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. Mol. Cell. Biol. *13*, 307-319.

Bentley, D. (2002). The mRNA assembly line: transcription and processing machines in the same factory. Curr. Opin. Cell. Biol. *14*, 336-342.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. J. Biol. Chem. *270*, 2411-2414.

Beyer, A. L., and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev. *2*, 754-765.

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

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.

Birchmeier, C., Schümperli, D., Sconzo, G., and Birnstiel, M. L. (1984). 3' editing of mRNAs: sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. U S A *81*, 1057-1061.

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.

Black, D. L., Chabot, B., and Steitz, J. A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell *42*, 737-750.

Blencowe, B. J., Issner, R., Nickerson, J. A., and Sharp, P. A. (1998). A coactivator of pre-mRNA splicing. Genes Dev. *12*, 996-1009.

Brackenridge, S., and Proudfoot, N. J. (2000). Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. Mol. Cell. Biol. *20*, 2660-2669.

Brosi, R., Groning, K., Behrens, S. E., Lührmann, R., and Kramer, A. (1993). Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science *262*, 102-105.

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

Calado, A., Kutay, U., Kuhn, U., Wahle, E., and Carmo-Fonseca, M. (2000). Deciphering the cellular pathway for transport of poly(A)-binding protein II. RNA *6*, 245-256.

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.

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.

Caspary, F., Shevchenko, A., Wilm, M., and Seraphin, B. (1999). Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. EMBO J. *18*, 3463-3474.

Champion-Arnaud, P., and Reed, R. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. Genes Dev. *8*, 1974-1983.

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, E. J., Frand, A. R., Chitouras, E., and Kaiser, C. A. (1998). A link between secretion and pre-mRNA processing defects in Saccharomyces cerevisiae and the identification of a novel splicing gene, RSE1. Mol. Cell. Biol. *18*, 7139-7146.

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

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, 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.

Chiara, M. D., Champion-Arnaud, P., Buvoli, M., Nadal-Ginard, B., and Reed, R. (1994). Specific protein-protein interactions between the essential mammalian

spliceosome-associated proteins SAP 61 and SAP 114. Proc. Natl. Acad. Sci. U S A *91*, 6403-6407.

Chiou, H. C., Dabrowski, C., and Alwine, J. C. (1991). Simian virus 40 late mRNA leader sequences involved in augmenting mRNA accumulation via multiple mechanisms, including increased polyadenylation efficiency. J. Virol. *65*, 6677-6685.

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.

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.

Conway, L., and Wickens, M. (1987). Analysis of mRNA 3' end formation by modification interference: the only modifications which prevent processing lie in AAUAAA and the poly(A) site. EMBO J. *6*, 4177-4184.

Cooke, C., and Alwine, J. C. (1996). The cap and the 3' splice site similarly affect polyadenylation efficiency. Mol. Cell. Biol. *16*, 2579-2584.

Cooke, C., Hans, H., and Alwine, J. C. (1999). Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. Mol. Cell. Biol. *19*, 4971-4979.

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

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

Cui, D. Y., Brown, C. R., and Chiang, H. L. (2004). The type 1 phosphatase Reg1p-Glc7p is required for the glucose-induced degradation of fructose-1,6-bisphosphatase in the vacuole. J. Biol. Chem. *279*, 9713-9724.

Culver, G. M., McCraith, S. M., Consaul, S. A., Stanford, D. R., and Phizicky, E. M. (1997). A 2'-phosphotransferase implicated in tRNA splicing is essential in Saccharomyces cerevisiae. J. Biol. Chem. *272*, 13203-13210.

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

Dantonel, J. C., Murthy, K. G., Manley, J. L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. Nature *389*, 399-402.

Das, B. K., Xia, L., Palandjian, L., Gozani, O., Chyung, Y., and Reed, R. (1999). Characterization of a protein complex containing spliceosomal proteins SAPs 49, 130, 145, and 155. Mol. Cell. Biol. *19*, 6796-6802.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. Genes Dev. *20*, 1100-1109.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol. Cell *5*, 779-787.

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525-532.

de Vries, H., Rüegsegger, U., Hübner, 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.

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., 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.

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

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

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

Dominski, Z., Erkmann, J. A., Yang, X., Sanchez, R., and Marzluff, W. F. (2002). A novel zinc finger protein is associated with U7 snRNP and interacts with the stem-loop binding protein in the histone pre-mRNP to stimulate 3'-end processing. Genes Dev. *16*, 58-71.

Dominski, Z., Yang, X. C., and Marzluff, W. F. (2005a). The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing. Cell *123*, 37-48.

Dominski, Z., Yang, X. C., Purdy, M., Wagner, E. J., and Marzluff, W. F. (2005b). 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.

Dominski, Z., Zheng, L. X., Sanchez, R., and Marzluff, W. F. (1999). Stem-loop binding protein facilitates 3'-end formation by stabilizing U7 snRNP binding to histone pre-mRNA. Mol. Cell. Biol. *19*, 3561-3570.

Duvel, K., Egli, C. M., and Braus, G. H. (1999). A single point mutation in the yeast TRP4 gene affects efficiency of mRNA 3' end processing and alters selection of the poly(A) site. Nucleic Acids Res. *27*, 1289-1295.

Dye, M. J., Gromak, N., and Proudfoot, N. J. (2006). Exon tethering in transcription by RNA polymerase II. Mol. Cell *21*, 849-859.

Edmonds, M. (2002). A history of poly A sequences: from formation to factors to function. Prog. Nucleic Acid Res. Mol. Biol. *71*, 285-389.

Edwalds-Gilbert, G., and Milcarek, C. (1995). The binding of a subunit of the general polyadenylation factor cleavage-polyadenylation specificity factor (CPSF) to polyadenylation sites changes during B cell development. Nucleic Acids Symp. Ser., 229-233.

Egli, C. M., Springer, C., and Braus, G. H. (1995). A complex unidirectional signal element mediates GCN4 mRNA 3' end formation in Saccharomyces cerevisiae. Mol. Cell. Biol. *15*, 2466-2473.

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.

Fitzgerald, M., and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms parts of the recognition site for polyadenylation of late SV40 mRNAs. Cell *24*, 251-260.

Flaherty, S. M., Fortes, P., Izaurralde, E., Mattaj, I. W., and Gilmartin, G. M. (1997). Participation of the nuclear cap binding complex in pre-mRNA 3' processing. Proc. Natl. Acad. Sci. U S A *94*, 11893-11898.

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.

Forne, T., Labourier, E., Antoine, E., Rossi, F., Gallouzi, I., Cathala, G., Tazi, J., and Brunel, C. (1996). Structural features of U6 snRNA and dynamic interactions with other spliceosomal components leading to pre-mRNA splicing. Biochimie *78*, 436-442.

Frank, D. N., and Pace, N. R. (1998). Ribonuclease P: unity and diversity in a tRNA processing ribozyme. Annu. Rev. Biochem. *67*, 153-180.

Frendewey, D., Krämer, A., and Keller, W. (1987). Different small nuclear ribonucleoprotein particles are involved in different steps of splicing complex formation. Cold Spring Harb. Symp. Quant. Biol. *52*, 287-298.

Fromont-Racine, M., Rain, J. C., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat. Genet. *16*, 277-282.

Gall, J. G., Bellini, M., Wu, Z., and Murphy, C. (1999). Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. Mol. Biol. Cell *10*, 4385-4402.

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., *et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141-147.

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.

Gilmartin, G. M., and Nevins, J. R. (1991). Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. Mol. Cell. Biol. *11*, 2432-2438.

Goldstrohm, A. C., Greenleaf, A. L., and Garcia-Blanco, M. A. (2001). Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. Gene *277*, 31-47.

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.

Gornemann, J., Kotovic, K. M., Hujer, K., and Neugebauer, K. M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol. Cell *19*, 53-63.

Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev. *10*, 233-243.

Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol. Cell. Biol. *18*, 4752-4760.

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

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

Graveley, B. R., Fleming, E. S., and Gilmartin, G. M. (1996). RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor. Mol. Cell. Biol. *16*, 4942-4951.

Gromak, N., West, S., and Proudfoot, N. J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol. Cell. Biol. *26*, 3986-3996.

Gross, S., and Moore, C. (2001a). 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.

Gross, S., and Moore, C. L. (2001b). Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. Mol. Cell. Biol. *21*, 8045-8055.

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

Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., and Mattaj, L. W. (1994). The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. Cell *76*, 531-541.

Gunderson, S. I., Polycarpou-Schwarz, M., and Mattaj, I. W. (1998). U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. Mol. Cell *1*, 255-264.

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). Signals sufficient for 3'-end formation of yeast mRNA. Mol. Cell. Biol. *16*, 2772-2776.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science *253*, 157-163.

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.

Hastings, M. L., and Krainer, A. R. (2001). Pre-mRNA splicing in the new millennium. Curr. Opin. Cell Biol. *13*, 302-309.

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.

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

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

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

Ho, C. K., Lehman, K., and Shuman, S. (1999). An essential surface motif (WAQKW) of yeast RNA triphosphatase mediates formation of the mRNA capping enzyme complex with RNA guanylyltransferase. Nucleic Acids Res. *27*, 4671-4678.

Ho, C. K., Rauhut, R., Vijayraghavan, U., and Abelson, J. (1990). Accumulation of pre-tRNA splicing '2/3' intermediates in a Saccharomyces cerevisiae mutant. EMBO J. *9*, 1245-1252.

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.

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.

Hollingworth, D., Noble, C. G., Taylor, I. A., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. RNA *12*, 555-560.

Hopper, A. K., and Phizicky, E. M. (2003). tRNA transfers to the limelight. Genes Dev. *17*, 162-180.

Howe, K. (2002). RNA polymerase II conducts a symphony of pre-mRNA processing activities. Biochim. Biophys. Acta. *1577*, 308.

Howe, K. J., Kane, C. M., and Ares, M., Jr. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA *9*, 993-1006.

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.

Irniger, S., and Braus, G. H. (1994). Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U S A *91*, 257-261.

Irniger, S., Sanfacon, H., Egli, C. M., and Braus, G. H. (1992). Different sequence elements are required for function of the cauliflower mosaic virus polyadenylation site in Saccharomyces cerevisiae compared with in plants. Mol. Cell. Biol. *12*, 2322-2330.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell *78*, 657-668.

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., and Keller, W. (1995). Cloning of cDNAs encoding the 160 kDa subunit of the bovine cleavage and polyadenylation specificity factor. Nucleic Acids Res. *23*, 2629-2635.

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.

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.

Kambach, C., Walke, S., and Nagai, K. (1999). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. Curr. Opin. Struct. Biol. *9*, 222-230.

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, R. W., Kuhn, U., Aragon, M., Bornikova, L., Wahle, E., and Bear, D. G. (2000). The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. J. Mol. Biol. *297*, 569-583.

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.

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.

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

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. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. EMBO J. *23*, 354-364.

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

Kolev, N. G., and Steitz, J. A. (2005). Symplekin and multiple other polyadenylation factors participate in 3'-end maturation of histone mRNAs. Genes Dev. *19*, 2583-2592.

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.

Krämer, A., Grüter, P., Groning, K., and Kastner, B. (1999). Combined biochemical and electron microscopic analyses reveal the architecture of the mammalian U2 snRNP. J. Cell. Biol. *145*, 1355-1368.

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.

Kuhn, A. N., and Kaufer, N. F. (2003). Pre-mRNA splicing in Schizosaccharomyces pombe: regulatory role of a kinase conserved from fission yeast to mammals. Curr. Genet. *42*, 241-251.

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.

Lenssen, E., James, N., Pedruzzi, I., Dubouloz, F., Cameroni, E., Bisig, R., Maillet, L., Werner, M., Roosen, J., Petrovic, K., *et al.* (2005). The Ccr4-Not complex independently controls both Msn2-dependent transcriptional activation-via a newly identified Glc7/Bud14 type I protein phosphatase module-and TFIID promoter distribution. Mol. Cell. Biol. *25*, 488-498.

Lewis, J. D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. Eur. J. Biochem. *247*, 461-469.

Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C., and Mattaj, I. W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. Genes Dev. *10*, 1683-1698.

Li, H., Trotta, C. R., and Abelson, J. (1998). Crystal structure and evolution of a transfer RNA splicing enzyme. Science *280*, 279-284.

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.

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.

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.

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955-964.

Luna, R., Jimeno, S., Marin, M., Huertas, P., Garcia-Rubio, M., and Aguilera, A. (2005). Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. Mol. Cell *18*, 711-722.

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.

Lutz, C. S., Murthy, K. G., Schek, N., O'Connor, J. P., Manley, J. L., and Alwine, J. C. (1996). Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes Dev. *10*, 325-337.

Lykke-Andersen, J., and Garrett, R. A. (1997). RNA-protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history. EMBO J. *16*, 6290-6300.

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.

Madhani, H. D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell *71*, 803-817.

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.

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.

Maroney, P. A., Romfo, C. M., and Nilsen, T. W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. Mol. Cell *6*, 317-328.

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 Doublié, S. (2000). Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP. EMBO J. *19*, 4193-4203.

Marzluff, W. F., and Duronio, R. J. (2002). Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. Curr. Opin. Cell. Biol. *14*, 692-699.

Mason, P. J., Elkington, J. A., Lloyd, M. M., Jones, M. B., and Williams, J. G. (1986). Mutations downstream of the polyadenylation site of a Xenopus beta-globin mRNA affect the position but not the efficiency of 3' processing. Cell *46*, 263-270.

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

McCraith, S. M., and Phizicky, E. M. (1990). A highly specific phosphatase from Saccharomyces cerevisiae implicated in tRNA splicing. Mol. Cell. Biol. *10*, 1049-1055.

McCraith, S. M., and Phizicky, E. M. (1991). An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. J. Biol. Chem. *266*, 11986-11992.

McDevitt, M. A., Hart, R. P., Wong, W. W., and Nevins, J. R. (1986). Sequences capable of restoring poly(A) site function define two distinct downstream elements. EMBO J. *5*, 2907-2913.

McLauchlan, J., Gaffney, D., Whitton, J. L., and Clements, J. B. (1985). The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini. Nucleic Acids Res. *13*, 1347-1368.

McNeil, J. B., Agah, H., and Bentley, D. (1998). Activated transcription independent of the RNA polymerase II holoenzyme in budding yeast. Genes Dev. *12*, 2510-2521.

Millevoi, S., Geraghty, F., Idowu, B., Tam, J. L., Antoniou, M., and Vagner, S. (2002). A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. EMBO Rep. *3*, 869-874.

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.

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., 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.

Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J. Cell Biol. *143*, 297-307.

Misteli, T., and Spector, D. L. (1998). The cellular organization of gene expression. Curr. Opin. Cell Biol. *10*, 323-331.

Moore, C. L., Chen, J., and Whoriskey, J. (1988). Two proteins crosslinked to RNA containing the adenovirus L3 poly(A) site require the AAUAAA sequence for binding. EMBO J. *7*, 3159-3169.

Moore, M. J., and Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-368.

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.

Murthy, K. G., and Manley, J. L. (1992). Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. J. Biol. Chem. *267*, 14804-14811.

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.

Nakano, A., and Muramatsu, M. (1989). A novel GTP-binding protein, Sar1p, is involved in transport from the endoplasmic reticulum to the Golgi apparatus. J. Cell Biol. *109*, 2677-2691.

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.

Nesic, D., Cheng, J., and Maquat, L. E. (1993). Sequences within the last intron function in RNA 3'-end formation in cultured cells. Mol. Cell. Biol. *13*, 3359-3369.

Nesic, D., and Maquat, L. E. (1994). Upstream introns influence the efficiency of final intron removal and RNA 3'-end formation. Genes Dev. *8*, 363-375.

Niwa, M., and Berget, S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. Genes Dev. *5*, 2086-2095.

Niwa, M., Rose, S. D., and Berget, S. M. (1990). In vitro polyadenylation is stimulated by the presence of an upstream intron. Genes Dev. *4*, 1552-1559.

Noble, S. M., and Guthrie, C. (1996). Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. Genetics *143*, 67-80.

O'Mullane, L., and Eperon, I. C. (1998). The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. Mol. Cell. Biol. *18*, 7510-7520.

Ogden, R. C., Lee, M. C., and Knapp, G. (1984). Transfer RNA splicing in Saccharomyces cerevisiae: defining the substrates. Nucleic Acids Res. *12*, 9367-9382.

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.

Orozco, I. J., Kim, S. J., and Martinson, H. G. (2002). The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. J. Biol. Chem. *277*, 42899-42911.

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.

Paushkin, S. V., Patel, M., Furia, B. S., Peltz, S. W., and Trotta, C. R. (2004). Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation. Cell *117*, 311-321.

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986). Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene. J. Biol. Chem. *261*, 2978-2986.

Pillai, R. S., Grimmler, M., Meister, G., Will, C. L., Lührmann, R., Fischer, U., and Schumperli, D. (2003). Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. Genes Dev. *17*, 2321-2333.

Preker, P. J., and Keller, W. (1998). The HAT helix, a repetitive motif implicated in RNA processing. Trends Biochem. Sci. *23*, 15-16.

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.

Price, D. H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. Mol. Cell. Biol. *20*, 2629-2634.

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

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

Proudfoot, N. J., and Brownlee, G. G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211-214.

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

Query, C. C., Strobel, S. A., and Sharp, P. A. (1996). Three recognition events at the branch-site adenine. EMBO J. *15*, 1392-1402.

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.

Rauhut, R., Green, P. R., and Abelson, J. (1990). Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme. J. Biol. Chem. *265*, 18180-18184.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr. Opin. Cell. Biol. *12*, 340-345.

Reines, D., Conaway, J. W., and Conaway, R. C. (1996). The RNA polymerase II general elongation factors. Trends Biochem. Sci. *21*, 351-355.

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.

Ruby, S. W., and Abelson, J. (1991). Pre-mRNA splicing in yeast. Trends Genet. *7*, 79-85.

Rüegsegger, 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.

Rüegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor I_m is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. Mol. Cell *1*, 243-253.

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., Hampsey, D. M., Zaret, K. S., and Sherman, F. (1991). Distinct cis-acting signals enhance 3' endpoint formation of CYC1 mRNA in the yeast Saccharomyces cerevisiae. EMBO J. *10*, 563-571.

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.

Ryan, K., Murthy, K. G., Kaneko, S., and Manley, J. L. (2002). Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. Mol. Cell. Biol. *22*, 1684-1692.

Sadowski, M., Dichtl, B., Hübner, 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.

Saraste, M., Sibbald, P. R., and Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. *15*, 430-434.

Schang, L. M., Hwang, G. J., Dynlacht, B. D., Speicher, D. W., Bantly, A., Schaffer, P. A., Shilatifard, A., Ge, H., and Shiekhattar, R. (2000). Human PC4 is a substrate-specific inhibitor of RNA polymerase II phosphorylation. J. Biol. Chem. *275*, 6071-6074.

Scharl, E. C., and Steitz, J. A. (1994). The site of 3' end formation of histone messenger RNA is a fixed distance from the downstream element recognized by the U7 snRNP. EMBO J. *13*, 2432-2440.

Scharl, E. C., and Steitz, J. A. (1996). Length suppression in histone messenger RNA 3'-end maturation: processing defects of insertion mutant premessenger RNAs can be compensated by insertions into the U7 small nuclear RNA. Proc. Natl. Acad. Sci. U S A *93*, 14659-14664.

Schaufele, F., Gilmartin, G. M., Bannwarth, W., and Birnstiel, M. L. (1986). Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. Nature *323*, 777-781.

Scott, J. M., and Imperiale, M. J. (1996). Reciprocal effects of splicing and polyadenylation on human immunodeficiency virus type 1 pre-mRNA processing. Virology *224*, 498-509.

Seghezzi, W., Chua, K., Shanahan, F., Gozani, O., Reed, R., and Lees, E. (1998). Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. Mol. Cell. Biol. *18*, 4526-4536.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol. Cell *11*, 965-976.

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.

Sharp, P. A., and Burge, C. B. (1997). Classification of introns: U2-type or U12-type. Cell *91*, 875-879.

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.

Sheets, M. D., and Wickens, M. (1989). Two phases in the addition of a poly(A) tail. Genes Dev. *3*, 1401-1412.

Smith, C. W., and Nadal-Ginard, B. (1989). Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. Cell *56*, 749-758.

Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999). Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. RNA *5*, 221-234.

Staley, J. P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

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.

Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327-331.

Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K., Lee, T. D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol. Cell *9*, 31-44.

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*, 1517-1520.

Sun, H., and Chasin, L. A. (2000). Multiple splicing defects in an intronic false exon. Mol. Cell. Biol. *20*, 6414-6425.

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.

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.

Takagaki, Y., and Manley, J. L. (1997). RNA recognition by the human polyadenylation factor CstF. Mol. Cell. Biol. *17*, 3907-3914.

Takagaki, Y., and Manley, J. L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol. Cell. Biol. *20*, 1515-1525.

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.

Takaku, H., Minagawa, A., Takagi, M., and Nashimoto, M. (2003). A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. Nucleic Acids Res. *31*, 2272-2278.

Tardiff, D. F., and Rosbash, M. (2006) Arrested yeast splicing complexes indicate a stepwise snRNP recruitment during in vivo spliceosome assembly. RNA *6*, 968-979.

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.

Tavtigian, S. V., Simard, J., Teng, D. H., Abtin, V., Baumgard, M., Beck, A., Camp, N. J., Carillo, A. R., Chen, Y., Dayananth, P., *et al.* (2001). A candidate prostate cancer susceptibility gene at chromosome 17p. Nat. Genet. *27*, 172-180.

Tran, D. P., Kim, S. J., Park, N. J., Jew, T. M., and Martinson, H. G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. Mol. Cell. Biol. *21*, 7495-7508.

Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme

with two active site subunits homologous to the archaeal tRNA endonucleases. Cell *89*, 849-858.

Tsai, T. F., Wu, M. J., and Su, T. S. (1998). Usage of cryptic splice sites in citrullinemia fibroblasts suggests role of polyadenylation in splice-site selection during terminal exon definition. DNA Cell. Biol. *17*, 717-725.

Vagner, S., Vagner, C., and Mattaj, I. W. (2000). The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. Genes Dev. *14*, 403-413.

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*, 272-280.

van Helden, J., del Olmo, M., and Perez-Ortin, J. E. (2000). Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res. *28*, 1000-1010.

Venkataraman, K., Brown, K. M., and Gilmartin, G. M. (2005). Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. *19*, 1315-1327.

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.

Wahle, E. (1991a). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66*, 759-768.

Wahle, E. (1991b). 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. (1995). Poly(A) tail length control is caused by termination of processive synthesis. J. Biol. Chem. *270*, 2800-2808.

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., Lustig, A., Jenö, P., and Maurer, P. (1993). Mammalian poly(A)-binding protein II. Physical properties and binding to polynucleotides. J. Biol. Chem. *268*, 2937-2945.

Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. *1*, 945-951.

Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. Genes Dev. *12*, 1409-1414.

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.

Weiss, E. A., Gilmartin, G. M., and Nevins, J. R. (1991). Poly(A) site efficiency reflects the stability of complex formation involving the downstream element. EMBO J. *10*, 215-219.

Westaway, S. K., Phizicky, E. M., and Abelson, J. (1988). Structure and function of the yeast tRNA ligase gene. J. Biol. Chem. *263*, 3171-3176.

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 Lührmann, R. (1997). Protein functions in pre-mRNA splicing. Curr. Opin. Cell. Biol. *9*, 320-328.

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

Will, C. L., Schneider, C., MacMillan, A. M., Katopodis, N. F., Neubauer, G., Wilm, M., Lührmann, R., and Query, C. C. (2001). A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. EMBO J. *20*, 4536-4546.

Will, C. L., Urlaub, H., Achsel, T., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. EMBO J. *21*, 4978-4988.

Wilusz, J., and Shenk, T. (1988). A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadenylation motif. Cell *52*, 221-228.

Wilusz, J., and Shenk, T. (1990). A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. Mol. Cell. Biol. *10*, 6397-6407.

Wilusz, J., Shenk, T., Takagaki, Y., and Manley, J. L. (1990). A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. Mol. Cell. Biol. *10*, 1244-1248.

Wittmann, T., and Wahle, E. (1997). Purification and characterization of full-length mammalian poly(A) polymerase. Biochim. Biophys. Acta. *1350*, 293-305.

Woychik, N. A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. Cell *108*, 453-463.

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.

Xiao, S., Scott, F., Fierke, C. A., and Engelke, D. R. (2002). Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. Annu. Rev. Biochem. *71*, 165-189.

Yoshihisa, T., Yunoki-Esaki, K., Ohshima, C., Tanaka, N., and Endo, T. (2003). Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria. Mol. Biol. Cell *14*, 3266-3279.

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.

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

Zarkower, D., Stephenson, P., Sheets, M., and Wickens, M. (1986). The AAUAAA sequence is required both for cleavage and for polyadenylation of simian virus 40 pre-mRNA in vitro. Mol. Cell. Biol. *6*, 2317-2323.

Zarkower, D., and Wickens, M. (1988). A functionally redundant downstream sequence in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex in vitro. J. Biol. Chem. *263*, 5780-5788.

Zeevi, M., Nevins, J. R., and Darnell, J. E., Jr. (1982). Newly formed mRNA lacking polyadenylic acid enters the cytoplasm and the polyribosomes but has a shorter half-life in the absence of polyadenylic acid. Mol. Cell. Biol. *2*, 517-525.

Zeng, C., and Berget, S. M. (2000). Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol. Cell. Biol. *20*, 8290-8301.

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., 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.

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.

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.

Zhelkovsky, A. M., Kessler, M. M., and Moore, C. L. (1995). Structure-function relationships in the Saccharomyces cerevisiae poly(A) polymerase. Identification of a

novel RNA binding site and a domain that interacts with specificity factor(s). J. Biol. Chem. *270*, 26715-26720.

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..

Acknowledgements

I would like to thank

Walter Keller for giving me the opportunity to do my Ph.D. in his lab. I am most grateful for his continuous support not only in scientific but also in personal matters.

Bernhard Dichtl for helpful suggestions and discussions and for being on my thesis committee.

Angela Krämer for many helpful discussions and provisions of reagents.

Mihaela Zavolan for being on my thesis committee.

Verena Widmer for her good nature and patience in doing the same experiments again and again.

Georges Martin (the column master) for teaching me how to run the different columns and systems.

All the former and present members of the lab for discussions and support.

The “Mädels” (including Anja) for being great friends and talking all night long.

Thomas for many enjoyable coffee breaks.

My sister and brother for spending many relaxing breakfasts together.

Special thanks to my parents for their unconditional love and support.

Gary Kooznetsoff for helping me through my highs and lows and being a never-ending source of energy, support and motivation.

Curriculum vitae

Name	Andrea Martina Kyburz Kooznetsoff
Date of birth	May 26, 1976
Place of birth	Brugg

Education and Academic Training

1983 – 88	Primary school in Lupfig
1988 – 92	Bezirksschule Windisch
1992 – 94	Diplommittelschule Wettingen
1994 – 97	Kantonsschule Wettingen (Maturität Typus PSG and C)
1997 – 02	Undergraduate studies in Biology II at the Universität Basel; Diploma thesis at the Biozentrum Basel of the University of Basel in the laboratory of Prof. W. Keller ("Molecular characterization of the yeast cleavage and polyadenylation factor subunit Ydh1/Cft2")
2002	6-month internship in the laboratory of Prof. W. Keller, Biozentrum Basel
2003-2006	Ph.D. student at the Biozentrum of the University of Basel in the laboratory of Prof. W. Keller ("Characterization of factors involved in the coupling of 3' end processing and splicing and in the 3' end formation of mRNA precursors")

Publications

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, in press.

Oral Presentations

Kyburz, A., and Keller, W. (2004). A possible role for CPSF and SF3b in the coupling of pre-mRNA 3' end processing and splicing. Swiss RNA Workshop in Bern.

Kyburz, A., and Keller, W. (2005). CPSF and U2 snRNP mediate coupling of pre-mRNA 3' end processing and splicing. Eukaryotic mRNA Processing Meeting in Cold Spring Harbor.

Ich erkläre hiermit, dass ich die Dissertation

“Characterization of factors involved in the coupling of 3’ end processing and splicing and in the 3’ end formation of mRNA precursors”

nur mit der darin angegebenen Hilfe verfasst und bei keiner anderen Fakultät eingereicht habe.

Andrea Kyburz